Evolution, 56(12), 2002, pp. 2383–2394

PHYLOGEOGRAPHIC BREAKS WITHOUT GEOGRAPHIC BARRIERS TO GENE FLOW

DARREN E. IRWIN[1]
Section for Animal Ecology, Department of Ecology, Lund University, S-223 62 Lund, Sweden
E-mail: darreni@ucla.edu

Abstract.—The spatial distribution of genetic markers can be useful both in estimating patterns of gene flow and in reconstructing biogeographic history, particularly when gene genealogies can be estimated. Genealogies based on nonrecombining genetic units such as mitochondrial and chloroplast DNA often consist of geographically separated clades that come into contact in narrow regions. Such phylogeographic breaks are usually assumed to be the result of long-term barriers to gene flow. Here I show that deep phylogeographic breaks can form within a continuously distributed species even when there are no barriers to gene flow. The likelihood of observing phylogeographic breaks increases as the average individual dispersal distance and population size decrease. Those molecular markers that are most likely to show evidence of real geographic barriers are also most likely to show phylogeographic breaks that formed without any barrier to gene flow. These results might provide an explanation as to why some species, such as the greenish warblers (*Phylloscopus trochiloides*), have phylogeographic breaks in mitochondrial or chloroplast DNA that do not coincide with sudden changes in other traits.

Key words.—Chloroplast DNA, coalescent theory, dispersal, genealogy, mitochondrial DNA, *Phylloscopus trochiloides*, phylogeography.

Received April 3, 2002. Accepted August 30, 2002.

Variation in genetic markers is often used to infer current or historical patterns of gene flow in a species or group of species (Bossart and Prowell 1998; Avise 2000). Historical geographic processes such as population division, range expansion, and long-distance colonization are expected to produce distinct patterns in the distribution of alleles and relationships between them (Templeton et al. 1995), and therefore it is reasonable to think that those processes can be inferred from patterns of genetic variation. Mitochondrial DNA (Avise 2000) and chloroplast DNA (Soltis et al. 1997; Petit et al. 2002) are widely used for this purpose because they are usually only maternally inherited and do not undergo recombination. These characteristics potentially allow the reconstruction of matrilineal genealogies, which are useful because they are hierarchical and show clear relationships among individuals. Recent years have seen explosive growth in the use of mitochondrial and chloroplast genealogies to make inferences regarding historical processes and current patterns of gene flow (Soltis et al. 1997; Avise 2000; Hare 2001).

The inference of patterns of current or historical gene flow from gene genealogies seems particularly straightforward when there is a sharp geographic boundary between two widely distributed clades (Upton and Murphy 1997; Patton and da Silva 1998; Riddle et al. 2000). Usually, researchers assume that such breaks are the result of geographic barriers to dispersal, cryptic species boundaries, or recent contacts between historically allopatric populations. Theoretical analyses of gene trees have generally focused on how phylogeographic structure can be caused by barriers to gene flow (e.g., Neigel and Avise 1986; Nei and Takahata 1993; Wakeley and Hey 1997) or metapopulation structure (Takahata and Slatkin 1990; Marjoram and Donnelly 1994; Hudson 1998; Wakeley and Aliacar 2001). In some cases, however, phylogeographic breaks do not coincide with sudden changes in other traits (e.g., Soltis et al. 1997; Gibbs et al. 2000; Bond et al. 2001; Puorto et al. 2001) or known biogeographic boundaries (e.g., Burton 1998). For example, in the greenish warblers (*Phylloscopus trochiloides*), a widely distributed and geographically variable ring species in Asia, there are two deep mitochondrial clades that come into contact in two narrow regions (Fig. 1; Irwin et al. 2001a, b). In central Siberia, the mitochondrial break is coincident with sudden changes in plumage, songs, song recognition, and migratory behavior. In Kashmir, in the western Himalayas, birds belonging to either mitochondrial clade are similar in these other traits. Prior to the mitochondrial DNA sequencing, no researchers had proposed that Kashmir was the site of a species boundary, whereas Ticehurst (1938) had proposed that central Siberia was a species boundary between two forms of greenish warbler. Irwin et al. (2001a), considering variation in all of the traits, argued that there is no species boundary in Kashmir and proposed two possible explanations for the phylogeographic break in mitochondrial DNA there. In the first, an ancestral species was split by a geographic barrier into western and eastern populations and then these populations recently expanded and now meet in Kashmir. In the second, the species was always continuously distributed across Kashmir, but phylogeographic structure in mitochondrial DNA developed as a result of low individual dispersal distances.

This second explanation is an unusual one that, according to Wake (2001, p. 299), ''runs counter to a central tenet of phylogeography, which sees history as having been recorded in the phylogeny of sequence lineages, and so is likely to be controversial.'' Avise et al. (1987) and Avise (1994) proposed several hypotheses for mitochondrial gene trees, one of which was that intraspecific monophyletic groups distinguished by large genealogical gaps usually arise from long-term extrinsic (biogeographic) barriers to gene flow. Although Neigel and Avise (1993) and Avise (2000) have cautioned that this hypothesis is not a rule and that gaps might arise without a geographic barrier to gene flow, many re-

[1] Present address: Center for Tropical Research, Institute of the Environment, 1609 Hershey Hall, University of California, Los Angeles, California 90095-1496.

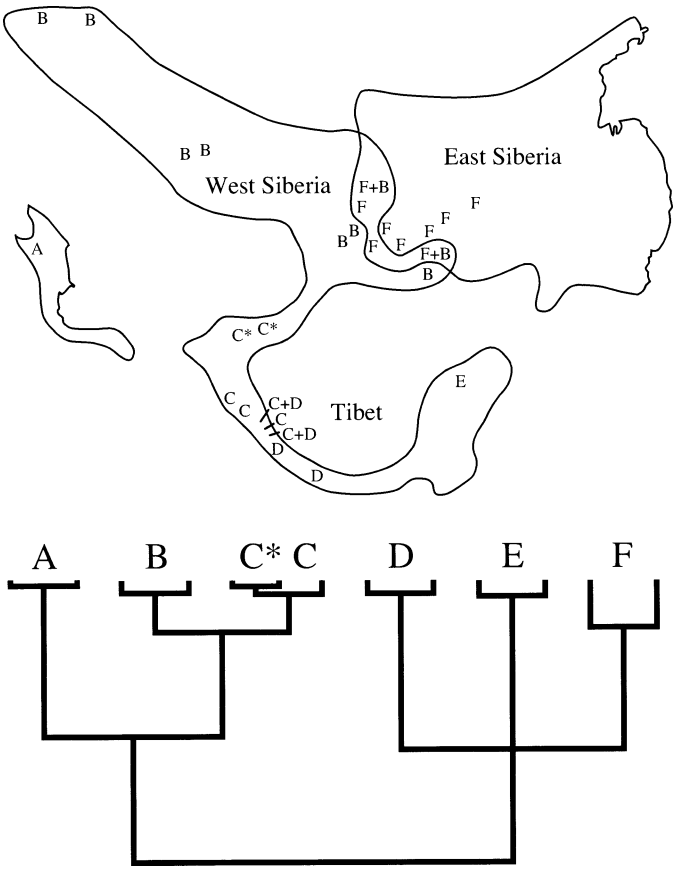

FIG. 1. The geographic range of greenish warbler (*Phylloscopus trochiloides*) populations in Asia and a simplified mitochondrial gene tree showing estimated matrilineal relationships between individuals from various sampling locations (modified from Irwin et al. 2001a). Letters on the map correspond to the mitochondrial clade to which individuals at that site belonged.

searchers assume that genealogical breaks are indicative of major breaks in gene flow in the past or present. For example, Riddle et al. (2000, p. 14438), introducing a phylogeographic study based on mitochondrial DNA of 12 species groups in the Baja California peninsular desert, stated that ''To the extent that divergent phylogroups are cryptically embedded within a single widespread species, the role of vicariance in structuring biotas will *necessarily* be underestimated.'' Upton and Murphy (1997, p. 104), referring to a phylogeographic break in mitochondrial DNA of side-blotched lizards of Baja California, wrote that ''the discontinuity *requires* a long-lasting isolation'' (italics added in both quotes). The direct interpretation that a genealogical break is the result of a long-term geographic barrier to gene flow relies on the assumption that genealogical breaks cannot arise when there are not geographic barriers to gene flow. However, because genealogies are influenced by chance, in the form of genetic drift, the possibility should be considered that phylogeographic breaks might develop in the absence of geographic barriers.

Here I show that phylogeographic breaks, often quite deep, readily arise in continuously distributed species if the average individual dispersal distances and/or population size of the species are low. This is a natural outcome of low dispersal distances and the way that mitochondrial and chloroplast DNA are inherited. As theoretical studies of isolation-by-distance models (Wright 1943; Slatkin and Maddison 1990; Neigel et al. 1991; Slatkin 1991, 1993; Neigel and Avise 1993; Barton and Wilson 1995) have shown, low dispersal distances in a continuously distributed species often lead to a positive correlation between the ages of lineages and their spatial extents. Two individuals are more likely to be closely related if they are geographically close to each other than if they are widely separated. Because mitochondrial and chloroplast DNA are usually inherited from only one parent (i.e., there is no recombination), they evolve through bifurcating genealogies. Individuals can belong to either of the two deepest genealogical clades, but they cannot be genealogically intermediate. These conditions lead unavoidably to the formation of phylogeographic structure. Under extreme conditions of low dispersal, one pair of adjacent sampling locations is likely to have individuals that belong entirely to different deep genealogical clades, while other pairs of adjacent locations have much higher genealogical relatedness.

Although this paper focuses on the formation of phylogeographic breaks when there are no geographic barriers, I also investigate how long and under what conditions the imprint of a historical barrier to gene flow will persist. The strength of a phylogeographic break that is caused by a barrier to gene flow tends to rapidly diminish with time after the barrier disappears. Only under conditions of very low individual dispersal distances does the phylogeographic break persist for long. Under these same conditions of low dispersal, phylogeographic breaks that are not the result of a geographic barrier are likely to arise.

METHODS

Simulation Models

Most theoretical studies of spatial genetic structure have assumed a small number of subpopulations that exchange genes through dispersal of individuals between subpopulations. Typically, such models take the form of either an island model, in which a number of subpopulations exchange a fixed proportion of genes with each other and usually have no spatial relationships (e.g., Slatkin and Maddison 1989; Slatkin 1991; Marjoram and Donnelly 1994; Hudson 1998; but see Wakeley 2001), or a stepping-stone model, in which subpopulations are arranged spatially and only exchange genes with neighbors (e.g., Slatkin and Maddison 1990; Slatkin 1991). For each type of model, the probabilities of reproducing are identical for all individuals within a single subpopulation. These models have been used for reasons of mathematical tractability rather than biological reality (Wakeley 2001). Many species do not consist of semi-isolated panmictic populations. Rather, some species are distributed over a relatively continuous range, with each individual having a unique location within that range. In such a system, the probability of reproducing is different for each individual, depending on local factors such as local intraspecific competition for resources (Felsenstein 1975). The structure of gene genealogies within such a system is difficult to model analytically (Felsenstein 1975; Barton and Wilson 1995). I therefore designed computer simulations, building on those

used by Neigel and Avise (1986, 1993), Slatkin and Maddison (1990), and Neigel et al. (1991). This is the first paper to specifically examine the formation of phylogeographic breaks in continuously distributed species using computer simulations; using similar simulations, Neigel and Avise (1986) and Neigel et al. (1991) studied the relationship between the geographic distribution and age of lineages, and Slatkin and Maddison (1990) examined the relationship between the distance between sampling sites and the effective migration rate between them.

To simulate matrilineal genealogies in a continuously distributed species, I considered a species with nonoverlapping generations that is distributed along a one-dimensional range. I chose to limit the simulations to linear ranges because simulations in two dimensions would take much more computing time, and because results in a single dimension are more easily analyzed and presented than those in two. I assumed that there was no geographic variation in habitat quality and that there were no geographic barriers to gene flow. Because the goal was to simulate matrilineal genealogies, males and mating behavior were ignored. The simulations modeled two fundamental stochastic processes, the reproduction of each female (i.e., the number of daughters) and the dispersal distance and direction of each daughter. The simulations recorded the relationships between individuals in the population as these two processes were repeated for many generations. I took two approaches to these simulations, a forward-in-time approach and a backward-in-time, coalescent approach. All simulations were done using MATLAB (The Mathworks, Natick, MA) on Macintosh computers.

Note that this paper concerns the simulation of true maternal genealogies. In an empirical study, such genealogies usually must be estimated from genetic data such as mitochondrial and chloroplast DNA sequences. The inference of genealogies from genetic data is a complex topic that has been discussed in depth elsewhere (e.g., Avise 1994; Mindell 1997; Page and Holmes 1998). This paper will focus on genealogical patterns that would be recovered if DNA data allowed the perfect recovery of a matrilineal genealogy.

Forward-in-Time Model

I started with a linear range of length = 1 and carrying capacity N for the entire species. On this range I placed n individuals (initially, $n = N$), each of which had a unique location that was randomly determined from a uniform distribution. Next, the reproductive probability of each individual was adjusted to reflect local density dependence. This was necessary to ensure that the species stayed relatively evenly spread across the entire range (Felsenstein 1975; Barton and Wilson 1995) and to ensure that the number of individuals in the species (n) stayed close to the carrying capacity (N). Presumably because of competition for resources, an individual in a dense part of the range had a lower reproductive output than an individual in a sparse part. Local density was measured by weighing the locations of all other individuals using a normal curve centered on the focal individual's location, x_0, with standard deviation w. To calculate μ_{rep}, the expected reproductive output, for each individual, I compared the local density calculated for that individual with the local density that would be calculated if N individuals were distributed evenly throughout the range:

$$\mu_{rep} = \frac{\sum \exp\left[-\frac{1}{2}(x_{even} - x_0)^2\right]}{\sum \exp\left[-\frac{1}{2}(x_{real} - x_0)^2\right]}, \quad (1)$$

where x_{even} represents the locations of N individuals spread evenly throughout the range and x_{real} represents the real locations of all individuals in the simulation. If an individual was in a dense part of the range, the expected reproduction, μ_{rep}, calculated according to the above equation was less than one, whereas an individual in a sparsely populated part of the range had a μ_{rep} greater than one. The actual reproduction of each individual (i.e., the number of daughters produced by each mother) was then determined by drawing from a Poisson distribution with mean μ_{rep}. Note that this method of determining reproduction automatically took into account the fact that individuals near the edge of the range had fewer individuals near them than those in the center of the range; on average, individuals at the edge had the same reproduction as those in the center. After the number of offspring for each individual was determined, I dispersed each offspring from the location of the parent by drawing from a normal distribution with standard deviation σ_{disp}, expressed relative to a range of length = 1. If the new location was outside of the range of the species, that result was rejected and more draws were done until the offspring had a location within the range.

The processes of reproduction and dispersal described above were repeated for many generations, while keeping track of the relationships between individuals and their locations. After many generations, all individuals in the population were descended from one individual in the original population, allowing a matrilineal genealogy to be constructed.

The main strength of the forward-in-time model is its intuitive relationship with reality. All females in the species are included in the model, there are easily understood phases of reproduction and dispersal, population size and local densities fluctuate as a natural consequence of the stochastic nature of reproduction and dispersal, and the simulation moves forward through time.

Coalescent Model

The second model, which was based on a backward-in-time approach using coalescent theory (Kingman 1982; reviewed by Harding 1996), allowed the simulation of much larger population sizes for longer periods of time. The geographic range consisted of a one-dimensional array of N points, each point representing a single female. I began by choosing a sample of individuals whose genealogy would be constructed. Then the location of the mother of each individual was determined by moving a number of points determined by a random draw from a normal distribution with standard deviation σ_{disp}. Whenever two or more individuals had the same mother (i.e., the same point was occupied by their mothers), there was a coalescent event in the genealogy. By moving backward through the generations until there was

only a single ancestral individual, the genealogy of the entire sample could be constructed. This approach was inspired by the lattice model of Slatkin and Maddison (1990).

The backward-in-time approach is computationally efficient because individuals that are not ancestral to the sampled individuals can be ignored. It allows the simulation of genealogies spanning millions of generations in species consisting of hundreds of thousands of individuals. Unlike the forward-in-time model, it does not require any a priori estimate of how many generations the simulation needs to run; each simulation runs until all lineages have coalesced.

Note that both models have two important parameters, both of which have similar roles in each model. These are N, the carrying capacity of the species, and σ_{disp}, the standard deviation of the dispersal curve. The forward-in-time model has an additional parameter, w, the standard deviation of the curve that determines the strength of local density dependence. In the coalescent model, density dependence is not explicitly defined, but it is inherent in the model because the species is assumed to have a constant population size with individuals that are evenly distributed across the range.

Analysis of the Simulations

For each simulation, I sampled the 10 closest individuals from each of six locations that were evenly distributed across the linear range, with the two outer locations at the extreme edges of the range (i.e., because the range spans locations from zero to one, sampling locations were at 0.0, 0.2, 0.4. 0.6, 0.8, and 1.0; hereafter, these locations will be referred to as 1 through 6). This sampling scheme roughly resembles those that are often used in real phylogeographic studies (e.g., Milot et al. 2000). In the forward-in-time model, the sampling was done after the simulation. In the coalescent model, the sampled individuals were picked first and then the simulation was used to construct their genealogy. In each case, the output of each simulation was a matrilineal genealogy of 60 individuals.

Nine simulations were run using the forward-in-time model, each lasting 100,000 generations. Each combination of the parameters N = (400, 800, 1600), σ_{disp} = (0.005, 0.01, 0.02), and w = 0.005 was used. In each simulation, genealogies were recorded every 10,000 generations, providing 10 genealogies under each of combination of parameters.

Using the coalescent model, simulations were run using every combination of the parameters N = (400; 800; 1600; 3200; 6400; 12,800; 25,600; 51,200; 102,400) and σ_{disp} = (0.00125, 0.0025, 0.005, 0.01, 0.02, 0.04). Ten simulations were done under each of the 54 combinations of parameters.

Using each of the resulting genealogies, I quantified the strength of major phylogeographic breaks. First, I determined which of the two deepest genealogical clades each individual belonged to. Then, for each pair of adjacent sampling sites (comparing locations 1 and 2, 2 and 3, and so on), I calculated the coefficient of association, ϕ (Sokal and Rohlf 1995), for the association between clade membership and location. Values of ϕ ranged from zero, when individuals from both sites belonged to the same deep clade, to one, when the two sites had individuals belonging entirely to different deep clades. The maximum of the five values of ϕ calculated between adjacent locations, ϕ_{max}, was used to represent the strength of the deepest phylogeographic break in the genealogy.

While the above method only considers the deepest split in the genealogy, I also measured phylogeographic structure using the entire structure of the genealogy. First, I calculated the average coalescent time between the individuals of adjacent locations, D, which is analogous to average genetic distance. I then calculated the ratio of the maximum to the minimum of the five average coalescent times between adjacent sampling sites, D_{max}/D_{min}. If this ratio was high, it meant that one pair of adjacent sampling sites was much more distantly related than another pair.

An important question is whether the patterns observed in the simulated genealogies depend on the number of individuals sampled per population. To investigate this possibility, I used the coalescent model to generate simulated genealogies for samples of 40 individuals, rather than the 10 individuals used in the other simulations, from each of the six populations across the range. Ten simulations were run using each of the four combinations of the parameters N = (1600; 25,600) and σ_{disp} = (0.0025, 0.02). The characteristics of the resulting genealogies, as measured using ϕ_{max}, D_{max}/D_{min}, and coalescent time, were compared with the genealogies based on samples of 10 individuals per population using Wilcoxon rank-sum tests.

Simulations of Historical Geographic Barriers

The simulations described above were used to model genealogies in continuously distributed species with no geographic barriers. I also used simulations to investigate how long a phylogeographic break that is caused by a geographic barrier will persist. I considered a situation in which there was a long-term barrier to gene flow, such that populations on either side of the barrier became reciprocally monophyletic with respect to the gene of interest. This barrier then disappeared, allowing dispersal of individuals between the two previously isolated regions. After various periods of time had passed, I then asked whether a phylogeographic break was present in the general location where the barrier had been.

The coalescent model, as described above, was used for these simulations, with a few modifications. I envisioned a long-term, complete barrier to dispersal in the center of the linear range (i.e., located at position 0.5 within a species range spanning from 0.0 to 1.0) that was present until T generations ago, when it suddenly disappeared. The model started with a sample of 10 individuals from each of six populations spread evenly across the range, and then the simulation moved backward in time while constructing the genealogy for T generations, when the simulation stopped. At this point the genealogy revealed the location of the maternal ancestor of each of the 60 individuals in the sample. Each ancestor was labeled ''A'' if it was on the left side of the geographic barrier (i.e., if its location was less than 0.5) and ''B'' if it was on the right side (i.e., location greater than 0.5). The same labels were then assigned to the descendents of each ancestor. In this way each of the 60 sampled individuals could be labeled as a matrilineal descendent of individuals on either the left or right side of the historical

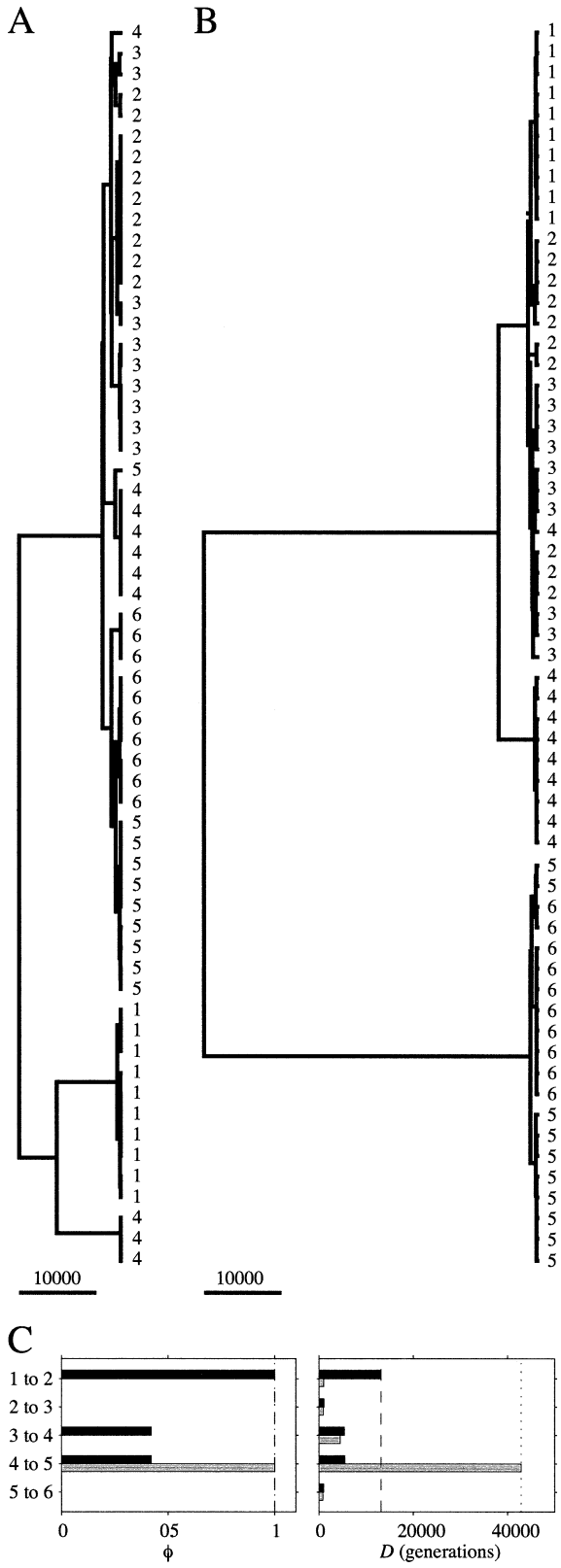

FIG. 2. Two maternal genealogies produced by the forward-in-time model, with a population size (N) of 1600 and standard deviation of dispersal distance (σ_{disp}) of 0.005. Each genealogy represents the relationships among 60 individuals, 10 drawn from each of six locations that are evenly distributed across a linear range. The numbers at the tips represent the location that the individual was sampled from, and scale bars indicate time in generations. (A) The genealogy after the simulation was run for 50,000 generations. (B) The genealogy after the same simulation was run for 100,000 generations. (C) Graphs of two measures of genetic divergence, ϕ (left) and D (right), between pairs of adjacent sampling locations. Dark bars show data from tree A, and gray bars show data from tree B. Dashed lines (for genealogy A) and dotted lines (for genealogy B) show ϕ and D between populations 1 and 6. There is strong phylogeographic structure in each genealogy, but the pattern of relationships is different in the two genealogies.

geographic barrier. Because I assumed that populations on either side of the barrier were reciprocally monophyletic when the barrier disappeared, the labels ''A'' and ''B'' referred to two divergent genealogical clades. To measure the strength of the phylogeographic break due to the historical barrier and in its general location, I measured the coefficient of association, ϕ (Sokal and Rohlf 1995), for the association between clade membership and location between sites 3 and 4, on either side of the historical barrier.

These simulations were run with $N = 6400$, $\sigma_{disp} =$ (0.00125, 0.0025, 0.005, 0.01, 0.02, 0.04), and $T =$ (25, 50, 100, 200, 400, 800, 1600, 3200, 6400). To check whether the results are sensitive to population size, I also ran simulations with $N = 25{,}600$, $\sigma_{disp} =$ (0.00125, 0.0025, 0.005, 0.01, 0.02, 0.04), and $T =$ (50, 200, 800, 3200). Ten simulations were done under each of the 78 combinations of parameters.

RESULTS

Phylogeography without Geographic Barriers

Both the forward-in-time and coalescent models show that phylogeographic breaks can develop in a continuously distributed species, and that phylogeographic structure can vary dramatically between simulations run under identical conditions. For example, in Figure 2 I show two simulated genealogies produced by the forward-in-time model, with $N =$ 1600 and $\sigma_{disp} = 0.005$. These two genealogies both have much geographic structure, but differ greatly in the apparent relationships between the populations. In the first (Fig. 2A), there is a deep phylogeographic break between locations 1 and 2, and there is relatively little divergence between other pairs of adjacent locations (see Fig. 2C for measures of genealogical divergence). In the second (Fig. 2B), there are two geographically separated clades with a deep genealogical break between them; individuals from locations 1–4 are highly related and those from locations 5 and 6 are highly related, compared to the genealogical distance between those groups. The two genealogies also differ dramatically in the coalescent time of the entire sample. All samples in Figure 2A descend from a single individual 13,152 generations back, whereas those in Figure 2B have a coalescence time of 42,913 generations.

Two examples of genealogies produced by the coalescent model (Fig. 3), with $N = 102{,}400$ and $\sigma_{disp} = 0.00125$, also differ dramatically while each showing phylogeographic structure. The first (Fig. 3A) has relatively similar divergence

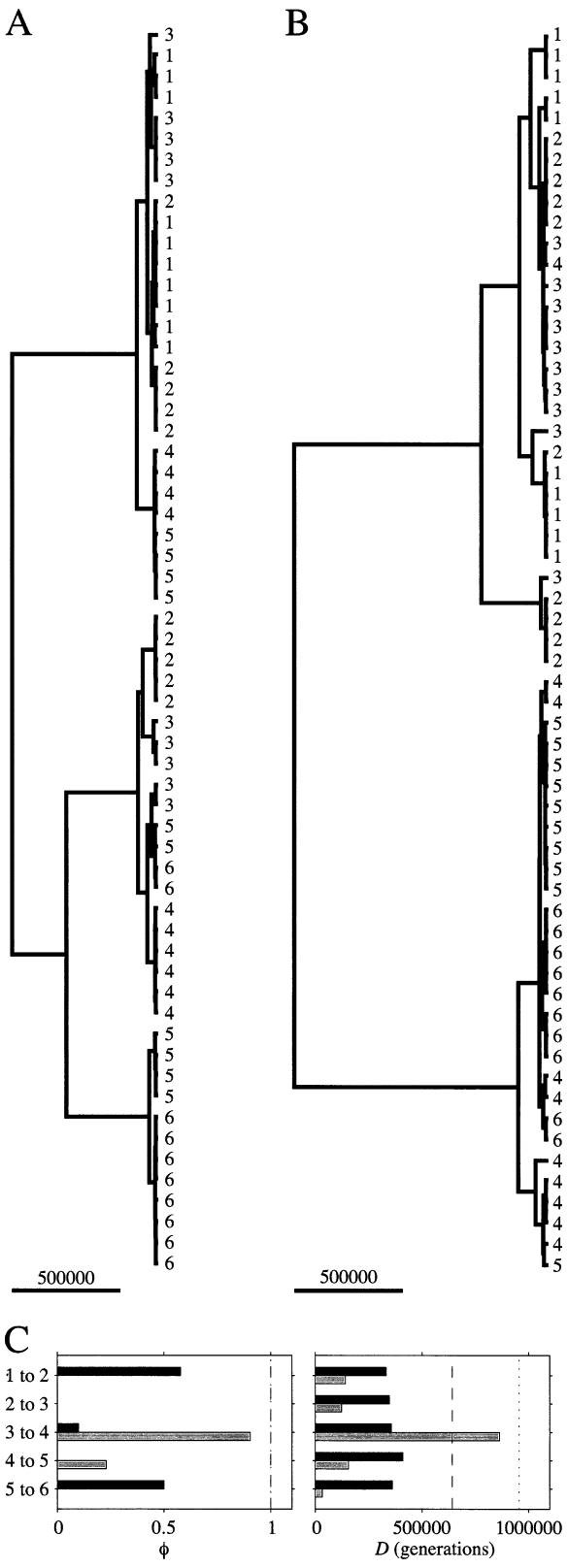

FIG. 3. Two maternal genealogies produced by independent runs of the coalescent model, with a population size (N) of 102,400 and standard deviation of dispersal distance (σ_{disp}) of 0.00125. Each genealogy represents the relationships among 60 individuals, 10 drawn from each of six locations that are evenly distributed across a linear range. The numbers at the tips represent the location that the individual was sampled from, and scale bars indicate time in generations. The two trees were chosen for illustration because of 10 simulated genealogies under these conditions, genealogy A had the lowest D_{max}/D_{min} ratio, and genealogy B had the highest. (C) Graphs of two measures of genetic divergence, ϕ (left) and D (right), between pairs of adjacent sampling locations. Dark bars show data from tree A, and gray bars show data from tree B. Dashed lines (for genealogy A) and dotted lines (for genealogy B) show ϕ and D between populations 1 and 6.

between all pairs of adjacent populations, whereas the second (Fig. 3B) has two geographically separated clades with a deep genealogical split between them, between populations 3 and 4 (see Fig. 3C for measures of genealogical divergence). The two genealogies also differ in their coalescence times (641,183 generations for Fig. 3A; 953,945 for Fig. 3B).

The amount of phylogeographic structure that is observed depends both on the average distance that individuals disperse each generation and on the number of individuals in the species. In Figure 4A, I show how ϕ_{max}, the strength of the phylogeographic break between the two deepest clades in the phylogeny, depends on σ_{disp} and N. Generally, genealogies produced by simulations with a combination of large population size and large dispersal distance have low values of ϕ_{max}, indicating that they do not have strong phylogeographic breaks. As dispersal distance and population size decrease, the strength of the deepest phylogeographic break increases. In Figure 4B, I show that the amount of phylogeographic structure in the entire genealogy (D_{max}/D_{min}) also increases with decreasing dispersal and population size.

The coalescent time of the simulated genealogies also depends on dispersal and population size (Fig. 4C). According to coalescent theory, the expected matrilineal coalescent time of a panmictic species of N females is $2N$ (Nei 1987; Avise 2000). The simulated genealogies, each representing relationships of 60 individuals sampled across the range, should almost always contain representatives of both of the deepest clades of the species, allowing a comparison between the coalescent time of the 60 individuals and the theoretical coalescent time of the species. At large population sizes and large dispersal distances, the average coalescent time of the simulated genealogies is close to $2N$, although there is a large amount of variation around this expected value. At lower population sizes and dispersal distances, coalescent times can be greatly increased, reaching more than $100N$ in some cases (Fig. 4C). There appears to be a close association between the appearance of phylogeographic structure (Fig. 4A,B) and an increase in the coalescent time (Fig. 4C); generally, values of N and σ_{disp} that lead to phylogeographic structure also lead to coalescent times above $2N$.

In Figure 4, I show average results from the coalescent model, but the results of the forward-in-time model are similar. Nine combinations of parameters were used for the forward-in-time model, and the resulting genealogies have ϕ_{max} values, D_{max}/D_{min} ratios, and coalescent times that do not differ significantly from the genealogies produced by the coalescent model (27 Wilcoxon rank sum tests, median P =

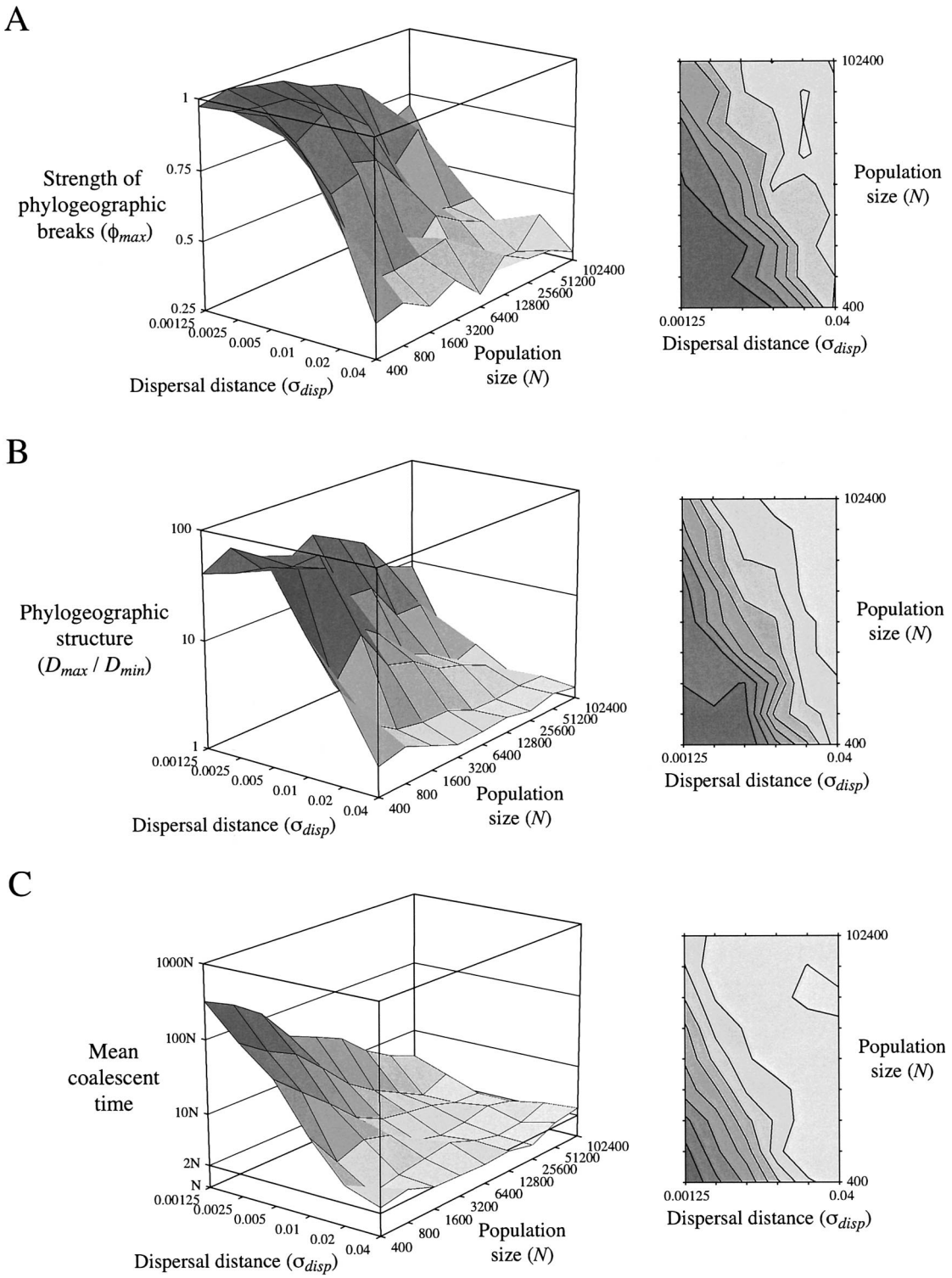

FIG. 4. The characteristics of gene genealogies depend on both the dispersal distance and population size of the species. On the left are three-dimensional graphs showing mean values from 10 simulations using the coalescent model at each of 54 combinations of σ_{disp} and N, and on the right are contour plots generated from the same data. As σ_{disp} and N decrease, (A) the strength of phylogeographic breaks (ϕ), (B) phylogeographic structure (D_{max}/D_{min}), and (C) coalescent time increase.

TABLE 1. Analysis of whether the presence of phylogeographic breaks (ϕ_{max}), the amount of phylogeographic structure (D_{max}/D_{min}), or coalescent time depend on sample size per sampling site. Under four combinations of σ_{disp} and N, 10 simulations were run using 10 samples per sampling site, and another 10 were run using 40 samples per sampling site. For each set of simulated genealogies, the table shows the means and, in parentheses, ranges of ϕ_{max}, D_{max}/D_{min}, and coalescent time. Also shown are P-values from Wilcoxon rank-sum tests comparing genealogies with 10 and 40 samples per sampling site. Only two tests (shown in bold) have P-values less than 0.05.

σ_{disp}	N	Samples per site	ϕ_{max}	P	D_{max}/D_{min}	P	Coalescent time ($\times 10^3$)	P
0.0025	1600	10	0.96 (0.58–1.00)		36.1 (5.0–131.5)		67.5 (16.0–157.1)	
		40	0.92 (0.73–1.00)	0.36	38.8 (4.1–174.5)	0.65	49.2 (15.2–129.0)	0.15
0.0025	25,600	10	0.57 (0.23–0.82)		5.2 (1.9–9.6)		77.5 (30.7–155.3)	
		40	0.60 (0.23–0.93)	0.76	5.1 (1.4–10.0)	0.76	110.3 (48.0–237.0)	0.11
0.02	1600	10	0.44 (0.23–0.73)		2.0 (1.2–3.5)		3.5 (1.1–5.6)	
		40	0.41 (0.26–0.60)	0.88	1.9 (1.2–3.1)	0.94	4.0 (1.0–8.2)	0.76
0.02	25,600	10	0.30 (0.20–0.44)		1.3 (1.1–2.0)		42.0 (25.9–89.7)	
		40	0.18 (0.08–0.29)	**0.004**	1.3 (1.0–1.7)	0.94	68.8 (37.0–164.3)	**0.010**

0.496; only one P-value [0.019] is less than $\alpha = 0.05$, but still greater than the Bonferroni-corrected $\alpha = 0.002$).

In Table 1, I show that the strength of phylogeographic breaks, the amount of phylogeographic structure, and coalescent times of simulated genealogies generally differ little between simulations using 10 individuals per sampling site and those using 40 individuals per sampling site. Of 12 comparisons, only two Wilcoxon rank-sum tests, one referring to ϕ_{max} and the other referring to coalescent time, resulted in P-values less than $\alpha = 0.05$, but still approximately equal or greater than the Bonferroni-corrected $\alpha = 0.0042$ (Table 1). Both of these apply to simulations run with a relatively high combination of population size and dispersal ($N = 25600$ and $\sigma_{disp} = 0.02$), conditions that do not lead to strong phylogeographic breaks, strong phylogeographic structure, or to high coalescent times (Fig. 4), and are therefore not of primary concern in this paper. Under these conditions, the dependence of ϕ_{max} on sampling size is a result of the fact that small samples tend to deviate more from population means than large samples. When there is no real association between genealogy and geography, the coefficient of association, ϕ, is lower for larger samples than for smaller samples. The possible dependence of coalescent time on sample size under these conditions might arise because larger samples are more likely to include representatives of both of the deepest clades in the species. The results in Table 1 indicate that, under conditions that lead to phylogeographic structure, results are relatively insensitive to sample size.

Phylogeography Due to Historical Geographic Barriers

In Figure 5, I show how the strength of the phylogeographic break due to a real geographic barrier tends to decline rapidly after disappearance of the barrier. The rate at which

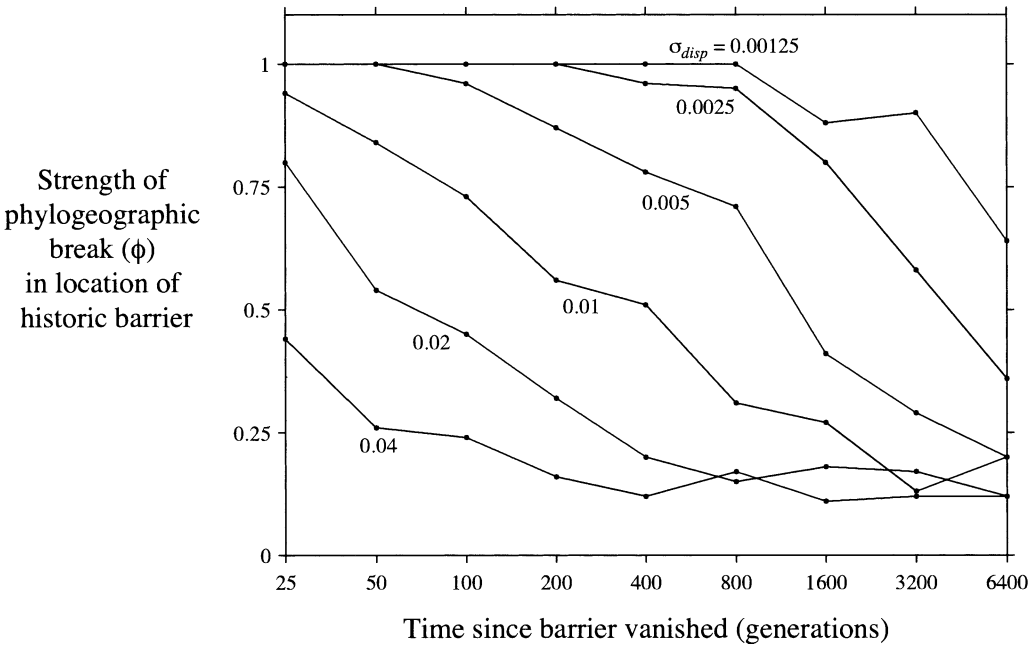

FIG. 5. Following the removal of a real barrier to dispersal, the strength of the phylogeographic break due to that barrier tends to decline rapidly. The figure shows mean values of ϕ from 10 simulations at each of 54 combinations of T (the time since removal of the barrier, on the horizontal axis) and σ_{disp} (shown by separate lines), with $N = 6400$.

this occurs increases with increasing individual dispersal distances (Fig. 5). Only when σ_{disp} is very low does the phylogeographic break persist for long. After 6400 generations, there is almost no association between clade membership and location when σ_{disp} is 0.005 or higher (for comparison, if one site has six clade A samples and four clade B, while the other site has four clade A and six clade B, $\phi = 0.2$).

The data presented in Figure 5 are based on simulations in which the total population size of the species, *N*, was 6400, but results from simulations in which $N = 25600$ were similar. Using $N = 25600$, 24 combinations of parameters were run (see Methods), and resulting ϕ-values did not differ significantly from ϕ-values in Figure 5 (24 Wilcoxon rank sum tests, median $P = 0.65$, range 0.13 to 1.00). These results indicate that the rate of disappearance of phylogeographic breaks that are caused by historical geographic barriers does not depend on the population size of the species.

Discussion

Many authors of phylogeographic studies argue, correctly, that species with small dispersal distances are more likely than those with large dispersal distances to show phylogeographic breaks caused by geographic barriers (e.g., Avise 1994; Bond et al. 2001). However, as the simulations presented here show, species that have small dispersal distances are also likely to show genealogical breaks that are not a result of geographic barriers. It is well known from coalescent theory that intraspecific genealogies are typically divided into two clades that are relatively distantly related (Barton and Wilson 1995). Less well known is that when dispersal distances are low, these two clades can be geographically separated, resulting in the appearance of phylogeographic breaks. These breaks have usually been attributed to long-term geographic barriers, but clearly this need not be the case. Such genealogical structure readily arises in a stable, continuously distributed species if dispersal distances and population sizes are sufficiently low.

Although this is the first paper to explore in detail the conditions under which phylogeographic breaks may arise by chance, the phenomenon has been suggested before. Neigel and Avise (1993, p. 1219), who conducted simulations with a goal of understanding the relationship between the age and the geographic distribution of a mitochondrial DNA lineage, stated that ''nested phylogeographic distributions can arise in the absence of barriers to gene flow'' and that ''specific explanations for every feature of an empirical [mitochondrial] DNA distribution may be unwarranted.'' These cautions have seldom been heeded; the great majority of studies that reveal phylogeographic breaks fail to consider the possibility that the break formed as a result of stochastic causes in a continuously distributed species rather than as a result of specific geographic causes (exceptions include Reeb and Avise 1990; Soltis et al. 1997; James and Moritz 2000; and Irwin et al. 2001a).

Under conditions of low dispersal and low population size, matrilineal relationships between groups of individuals sampled from different locations across a geographic range can vary tremendously. These matrilineal relationships may deviate widely from the true relationships when considering the entire genome. Most parts of the genome are inherited from both parents, and sexual reproduction results in the creation of new combinations of genes that have different genealogical histories. Although all copies of a nonrecombining marker such as mitochondrial or chloroplast DNA descended from a single ancestral individual, different sections of the recombining part of the genome were inherited from different ancestral individuals. Thus, under the conditions of the simulations, relationships based on the entire genome would be much less variable than matrilineal relationships.

An interesting outcome of the simulations is that two factors—dispersal distance and population size—interact to determine the amount of phylogeographic structure. The role of small dispersal distance in causing phylogeographic structure is obvious, but the role of population size is less intuitively clear. The link can be understood by considering the role of coalescent times. In a panmictic population (i.e., one in which all individuals have an equal chance of reproducing) of *N* haploid individuals, the expected coalescence time is $2N$ generations (Avise 2000). However, the simulations show that when a species is distributed across a geographic range and each individual disperses a short distance compared to the size of that range, expected coalescence times can be much greater (Fig. 4C). This is a result of the fact that it can take more than $2N$ generations for the descendants of any one individual to spread across the range. Using a random-walk diffusion model, Neigel et al. (1991) showed that $\sigma_G^2 = 2G\sigma_{disp}^2$, where σ_G^2 is the expected variance in distances between pairs of descendants after *G* generations. Assume, for the moment, that the descendants of an individual at the very center of the range spread by simple diffusion all the way across the range. The expected time in generations that this process would take is given by $G = (\sigma_G^2/2\sigma_{disp}^2) = (0.1617/2\sigma_{disp}^2)$, because $\sigma_G^2 = 0.1617$ if the descendants are spread evenly over a range of length = 1. Descendants of individuals near the edge of the range would take even longer to spread across the range. If the time necessary for this diffusion process is similar or greater than $2N$, then the actual coalescent times will tend to be greater than $2N$ and there will be phylogeographic structure. For example, when $N = 1600$ and $\sigma_{disp} = 0.005$, $2N = 3200$ and $G = 3234$. Under these parameters, strong phylogeographic structure forms in the simulations (Fig. 4A,B), and the coalescence times are greater than $2N$ (Fig. 4C). If we hold *N* at 1600 and increase dispersal to $\sigma_{disp} = 0.02$, then $G = 202$, so that time needed for dispersal is much less than $2N$ and there is less population structure. Similarly, if we keep $\sigma_{disp} = 0.005$ and increase *N* to 12,800, then $G = 3234$ is much less than $2N$ and there is less population structure. Note that local density dependence plays an essential role in these interactions, because it is what keeps the species evenly spread across the range (Felsenstein 1975; Barton and Wilson 1995). If there were no local density dependence, all individuals would have an equal chance of reproducing and the expected coalescence time would be $2N$.

Researchers often argue that mitochondrial and chloroplast DNA are more likely than nuclear markers to show evidence of real barriers to gene flow (Moore 1995; Wiens and Penkrot 2002), for two main reasons. First, maternally inherited markers have effective population sizes that are generally one-fourth those of nuclear genes. As a result, when a single

species is divided into two populations, mitochondrial and chloroplast DNA tend to become reciprocally monophyletic more quickly than nuclear markers. Second, mitochondrial and chloroplast DNA do not undergo recombination, and hence clear genealogical patterns can be reconstructed. For these same reasons, maternally inherited markers are more likely than nuclear markers to show phylogeographic gaps that are not caused by geographic barriers. As the simulations here have shown, the likelihood of observing phylogeographic breaks increases with decreasing population size; the fact that maternally inherited markers have one-fourth the effective population size of nuclear genes makes them more likely to show phylogeographic structure in a continuously distributed species. The lack of recombination causes mitochondrial and chloroplast haplotypes to evolve through bifurcating genealogies; an individual must belong to either of the two deepest genealogical clades, but cannot be genealogically intermediate. It is ironic that the markers that are most likely to show evidence of real barriers are also most likely to show evidence of barriers that never existed.

The purpose of the simulations was to investigate the conditions under which phylogeographic breaks can form in an idealized species that is continuously distributed across a range that has no variation in habitat quality over space or time. Of course, real species are almost never evenly distributed, and their ranges almost always vary in quality (Barton and Wilson 1995). If phylogeographic breaks can develop in the simulations presented here, they could even more easily develop in species subject to more realistic conditions. Factors such as small local extinctions, spatial variation in habitat quality, and variation in dispersal distances across the range could increase phylogeographic structure. Another factor that is rarely considered but potentially important is selection; if different mitochondrial haplotypes are favored in different locations, phylogeographic breaks can persist despite much gene flow. The hypothesis that phylogeographic breaks are caused primarily by long-term vicariant events should be reconsidered in light of the variety of scenarios that could lead to such phylogeographic structure.

The possibility that phylogeographic breaks might arise in locations where there is no specific geographic cause presents a problem to researchers who would like to use genetic variation to uncover current and historical patterns of gene flow. There are several possible solutions to this problem. First, researchers should study variation in multiple independent genetic markers in any phylogeographic study (Hare 2001). If phylogeographic structure is simply a result of small dispersal distances, different units of inheritance should show different phylogeographic patterns (Barton and Wilson 1995). Only when multiple independent genetic markers show genealogical breaks in the same geographic location should it be concluded that there is a specific geographic cause for the breaks (Avise 2000; Hare 2001). Second, researchers can study several species to see if they have the same patterns of phylogeographic breaks, in which case those breaks likely have specific geographic causes (Patton and da Silva 1998; Avise 2000; Riddle et al. 2000). However, even in these situations there may not have been complete isolation between populations; phylogeographic breaks in multiple genes or in multiple species may tend to coexist in regions of poor habitat quality, ecological gradients, or reduced dispersal. Third, hypothesis-testing approaches should be employed as much as possible in phylogeographic studies (Puorto et al. 2001). Before doing a survey of genetic markers, researchers should propose specific locations as species boundaries, barriers to gene flow, or areas of recent contact between historically allopatric populations. Then, a phylogeographic break in one of those areas would be supportive of the hypothesized geographic factor. If a phylogeographic break is found in an area where none was hypothesized to be, more evidence should be obtained to investigate whether the break is the result of a specific geographic factor at that location. Variation in morphological, behavioral, ecological, or physiological traits should be used together with the molecular variation to test hypotheses regarding biogeographic history and species boundaries (Puorto et al. 2001; Wiens and Penkrot 2002).

Although phylogeographic breaks can form without geographic barriers when dispersal distances are small, simulations also show that phylogeographic breaks that are caused by geographic barriers can disappear rapidly if dispersal distances are not very small (Fig. 5). However, these simulations did not take into account other factors that may stabilize the phylogeographic break after the geographic barrier is removed. Such factors include reproductive isolation between the genealogical groups (i.e., a species boundary) or ecological and morphological divergence of the two groups (Reeb and Avise 1990), perhaps with selection against hybrids. In these cases, in which the phylogeographic break is no longer maintained by a geographic barrier to dispersal, examination of other traits should reveal differences between the groups.

In the case of the ring of greenish warbler populations, a mitochondrial DNA genealogy showed two major phylogeographic breaks (Fig. 1). The northern break, in central Siberia, was predicted in advance, based on morphological and behavioral evidence that there is a species boundary there (Ticehurst 1938; Irwin 2000). The southern break, in Kashmir, was surprising because it was not predicted and is not coincident with sharp changes in morphological and behavioral traits (Irwin et al. 2001a). The traditional interpretation would be that the southern break also represents a species boundary or a recent contact between long-separated allopatric populations, but the simulations in this paper illustrate the possibility that the mitochondrial break arose without any barrier to gene flow in the past or present.

It is difficult to compare the simulations with real species numerically for several reasons. First, the simulations were based on a one-dimensional species range, whereas real species ranges are two- (or three-) dimensional. Second, variation in reproduction and dispersal may differ between the simulations and real species. Third, real species experience fluctuations in population size greater than those in the simulations. With these caveats in mind, we can compare the greenish warbler mitochondrial DNA phylogeny (Fig. 1) with the simulated genealogy in Figure 3B. The two trees are roughly comparable, both having a deep phylogeographic break and high relatedness between some neighboring sampling sites. The depths of the two trees are also similar (2–3 million years for the greenish warbler mitochondrial DNA phylogeny [Irwin 2001b], 953,945 generations for the sim-

ulated genealogy, which corresponds to 1.9 million years if the generation time is 2 years). The simulated tree was produced using parameter values of $\sigma_{disp} = 0.00125$ and $N = 102{,}400$. In the greenish warblers, given the distance around the ring of populations encircling Tibet of about 8500 km, these values would correspond to a standard deviation of dispersal of 10.6 km (i.e., a mean dispersal distance of 8.5 km) and a density of about 12 breeding females/km around the ring. I have little data on dispersal distances in greenish warblers, but comparisons with data from other passerine species show that a σ_{disp} of 10.6 km is reasonable. Accurate data on single-generation dispersal is difficult to obtain; in detailed studies of breeding populations, birds that leave the study area are not included in the dispersal estimates, whereas in broader banding-recovery studies, individuals recovered after dispersing far may not be successfully breeding (Moore and Dolbeer 1989). Barrowclough (1980) attempted to correct for the first problem in studies of dispersal in seven passerine species, and the resulting estimates of σ_{disp} (often referred to as RMS dispersal distances) ranged from 0.34 km in the song sparrow (*Melospiza melodia*) to 1.68 km in the Bewick's wren (*Thryomanes bewickii*), with an average of 1.00 km. Moore and Dolbeer (1989), using data on banding recoveries over broad areas, estimated a σ_{disp} of 94.6 km in the red-winged blackbird (*Agelaius phoeniceus*) and 111.4 km in the common grackle (*Quiscalus quiscula*). Hansson et al. (2002) estimated the σ_{disp} of great reed warblers (*Acrocephalus arundinaceus*), which are in the same family (Silviidae) as greenish warblers, to be 33 km. It is likely that greenish warblers in the southern part of their range would have small single-generation dispersal distances compared to most passerine birds, for two reasons. First, they inhabit a narrow altitudinal range in treeline habitat (Price 1991), perhaps making it difficult for long-distance dispersers to find appropriate habitat. Second, geographic variation in songs and morphological characteristics (Irwin et al. 2001a) suggest that populations are locally adapted, perhaps causing long-distance dispersers to have difficulty finding a mate and successfully breeding.

The density of breeding females around the ring is also difficult to estimate. Population densities of males on seven study sites range from 20 to 180 per km^2 (mean 72.3; Irwin 2000), suggesting that the 12 females/km used in Figure 3B is too low. However, the average density of females around the entire ring may be far less than the density at sites that were chosen for research on greenish warblers, and population fluctuations may cause the number of females that should be used in the model to be far less than the census number. Furthermore, the treeline habitat of greenish warblers through the Himalayas is very narrow, in many places only several hundred meters wide because of the steepness of the mountains and the narrow altitudinal range. Overall, these considerations suggest that the values used to generate the tree in Figure 3B are within the realm of possibility, suggesting that it may have been possible for the phylogeographic break in Kashmir to have formed in the absence of barriers to gene flow. Ongoing work using other molecular markers will help determine whether the pattern is a result of a long-term barrier to dispersal or a result of low dispersal in a more-or-less continuously distributed species.

In reality most species, including the greenish warblers, have probably experienced a very complex biogeographic history, with local extinctions, range expansions, and spatial and temporal variation in habitat quality. Presently, theoretical models of gene trees are simple, usually only examining the idealized situations of island models, stepping stone models, or the continuous distribution modeled here. More complex and realistic models are needed to help us more confidently reconstruct biogeographic histories using genetic data. Meanwhile, phylogeographers should consider the possibility that phylogeographic breaks may not be the result of specific geographic factors such as barriers, cryptic species boundaries, or recent contact between historically allopatric populations. Multiple sources of evidence must be considered before concluding that a phylogeographic break is the result of a long-term geographic barrier to dispersal.

Acknowledgments

I am especially grateful to T. Case for enthusiastic encouragement and discussion during the early and final stages of this project and to T. Price for his careful readings of the manuscript and for suggesting the use of the coefficient of association (ϕ) to measure the strength of phylogeographic breaks. I also thank S. Bensch, M. Dantzker, H. L. Gibbs, J. Irwin, M. Kirkpatrick, K. Marchetti, K. Martien, J. Neigel, A. Suarez, E. Svensson, and one anonymous reviewer for helpful comments and/or discussion and K. Marchetti for computer use. Support was provided by the National Science Foundation (grants DEB-9806692 to T. Price and INT-0076212); T. Price and the Division of Biological Sciences at the University of California, San Diego; and S. Bensch and the Department of Animal Ecology at Lund University.

Literature Cited

Avise, J. C. 1994. Molecular markers, natural history and evolution. Chapman and Hall, New York.

———. 2000. Phylogeography: the history and formation of species. Harvard Univ. Press, Cambridge, MA.

Avise, J. C., J. Arnold, R. M. Ball Jr., E. Bermingham, T. Lamb, J. E. Neigel, C. A. Reeb, and N. C. Saunders. 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. Annu. Rev. Ecol. Syst. 18: 489–522.

Barrowclough, G. F. 1980. Gene flow, population sizes, and genetic variance components in birds. Evolution 34:789–798.

Barton, N. H., and I. Wilson. 1995. Genealogies and geography. Philos. Trans. R. Soc. Lond. B 349:49–59.

Bond, J. E., M. C. Hedin, M. G. Ramirez, and B. D. Opell. 2001. Deep molecular divergence in the absence of morphological and ecological change in the Californian coastal dune endemic trapdoor spider *Aptostichus simus*. Mol. Ecol. 10:899–910.

Bossart, J. L., and D. P. Prowell. 1998. Genetic estimates of population structure and gene flow: limitations, lessons and new directions. Trends Ecol. Evol. 13:202–206.

Burton, R. S. 1998. Intraspecific phylogeography across the Point Conception biogeographic boundary. Evolution 52:734–745.

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

Gibbs, H. L., R. J. G. Dawson, and K. A. Hobson. 2000. Limited differentiation in microsatellite DNA variation among northern populations of the yellow warbler: evidence for male-biased gene flow? Mol. Ecol. 9:2137–2147.

Hansson, B., S. Bensch, D. Hasselquist, and B. Nielsen. 2002. Re-

stricted dispersal in a long-distance migrant bird with patchy distribution, the great reed warbler. Oecologia 130:536–542.
Harding, R. M. 1996. New phylogenies: an introductory look at the coalescent. Pp. 15–22 *in* P. H. Harvey, A. J. Leigh Brown, J. Maynard Smith, and S. Nee, eds. New uses for new phylogenies. Oxford Univ. Press, Oxford, U.K.
Hare, M. P. 2001. Prospects for nuclear gene phylogeography. Trends Ecol. Evol. 16:700–706.
Hudson, R. R. 1998. Island models and the coalescent process. Mol. Ecol. 7:413–418.
Irwin, D. E. 2000. Song variation in an avian ring species. Evolution 54:998–1010.
Irwin, D. E., S. Bensch, and T. D. Price. 2001a. Speciation in a ring. Nature 409:333–337.
Irwin, D. E., J. H. Irwin, and T. D. Price. 2001b. Ring species as bridges between microevolution and speciation. Genetica 112/113:223–243.
James, C. H., and C. Moritz. 2000. Intraspecific phylogeography in the sedge frog *Litoria fallax* (Hylidae) indicates pre-Pleistocene vicariance of an open forest species from eastern Australia. Mol. Ecol. 9:349–358.
Kingman, J. F. C. 1982. The coalescent. Stochastic Process. Appl. 13:235–248.
Marjoram, P., and P. Donnelly. 1994. Pairwise comparisons of mitochondrial DNA sequences in subdivided populations and implications for early human evolution. Genetics 136:673–683.
Milot, E., H. L. Gibbs, and K. A. Hobson. 2000. Phylogeography and genetic structure of northern populations of the yellow warbler (*Dendroica petechia*). Mol. Ecol. 9:667–681.
Mindell, D. P., ed. 1997. Avian molecular evolution and systematics. Academic Press, San Diego, CA.
Moore, W. S. 1995. Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees. Evolution 49:718–726.
Moore, W. S., and R. A. Dolbeer. 1989. The use of banding recovery data to estimate dispersal rates and gene flow in avian species: case studies in the red-winged blackbird and common grackle. Condor 91:242–253.
Nei, M. 1987. Molecular evolutionary genetics. Columbia Univ. Press, New York.
Nei, M., and N. Takahata. 1993. Effective population size, genetic diversity, and coalescence time in subdivided populations. J. Mol. Evol. 37:240–244.
Neigel, J. E., and J. C. Avise. 1986. Phylogenetic relationships of mitochondrial DNA under various demographic models of speciation. Pp. 515–534 *in* E. Nevo and S. Karlin, eds. Evolutionary processes and theory. Academic Press, New York.
———. 1993. Application of a random walk model to geographic distributions of animal mitochondrial DNA variation. Genetics 135:1209–1220.
Neigel, J. E., R. M. Ball Jr.,and J. C. Avise. 1991. Estimation of single generation migration distances from geographic variation in animal mitochondrial DNA. Evolution 45:423–432.
Page, R. D. M., and E. C. Holmes, eds. 1998. Molecular evolution. Blackwell Science, Oxford, U.K.
Patton, J. L., and M. N. F. da Silva. 1998. Rivers, refuges, and ridges: the geography of speciation of Amazonian mammals. Pp. 202–213 *in* D. J. Howard and S. H. Berlocher, eds. Endless forms: species and speciation. Oxford Univ. Press, Oxford, U.K.
Petit, R. J., U. M. Csaikl, S. Bordács, and 26 other authors. 2002. Chloroplast DNA variation in European white oaks: phylogeography and patterns of diversity based on data from over 2600 populations. For. Ecol. Manage. 156:5–26.
Price, T. 1991. Morphology and ecology of breeding warblers along an altitudinal gradient in Kashmir, India. J. Anim. Ecol. 60: 643–664.
Puorto, G., M. da Graça Salomão, R. D. G. Theakston, R. S. Thorpe, D. A. Warrell, and W. Wüster. 2001. Combining mitochondrial DNA sequences and morphological data to infer species boundaries: phylogeography of lanceheaded pitvipers in the Brazilian Atlantic forest, and the status of *Bothrops pradoi* (Squamata: Serpentes: Viperidae). J. Evol. Biol. 14:527–538.
Reeb, C. A., and J. C. Avise. 1990. A genetic discontinuity in a continuously distributed species: mitochondrial DNA in the American oyster, *Crassostrea virginica*. Genetics 124:397–406.
Riddle, B. R., D. J. Hafner, L. F. Alexander, and J. R. Jaeger. 2000. Cryptic vicariance in the historical assembly of a Baja California Peninsular Desert biota. Proc. Natl. Acad. Sci. USA 97: 14438–14443.
Slatkin, M. 1991. Inbreeding coefficients and coalescence times. Genet. Res. 58:167–175.
———. 1993. Isolation by distance in equilibrium and non-equilibrium populations. Evolution 47:264–279.
Slatkin, M., and W. P. Maddison. 1989. A cladistic measure of gene flow inferred from the phylogenies of alleles. Genetics 123: 603–613.
———. 1990. Detecting isolation by distance using phylogenies of genes. Genetics 126:249–260.
Sokal, R. R., and F. J. Rohlf. 1995. Biometry: the principles and practice of statistics in biological research. W. H. Freeman, New York.
Soltis, D. E., M. A. Gitzendanner, D. D. Strenge, and P. S. Soltis. 1997. Chloroplast DNA intraspecific phylogeography of plants from the Pacific Northwest of North America. Plant Syst. Evol. 206:353–373.
Takahata, N., and M. Slatkin. 1990. Genealogy of neutral genes in two partially isolated populations. Theor. Popul. Biol. 38: 331–350.
Templeton, A. R., E. Routman, and C. A. Phillips. 1995. Separating population structure from population history: a cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in the tiger salamander, *Ambystoma tigrinum*. Genetics 140: 767–782.
Ticehurst, C. B. 1938. A systematic review of the genus *Phylloscopus*. Trustees of the British Museum, London.
Upton, D. E., and R. W. Murphy. 1997. Phylogeny of the side-blotched lizards (Phrynosomatidae: *Uta*) based on mtDNA sequences: support for a midpeninsular seaway in Baja California. Mol. Phylogenet. Evol. 8:104–113.
Wake, D. B. 2001. Speciation in the round. Nature 409:299–300.
Wakeley, J. 2001. The coalescent in an island model of population subdivision with variation among demes. Theor. Popul. Biol. 59: 133–144.
Wakeley, J., and N. Aliacar. 2001. Gene genealogies in a metapopulation. Genetics 159:893–905.
Wakeley, J., and J. Hey. 1997. Estimating ancestral population parameters. Genetics 145:847–855.
Wiens, J. J., and T. A. Penkrot. 2002. Delimiting species using DNA and morphological variation and discordant species limits in spiny lizards (*Sceloporous*). Syst. Biol. 51:69–91.
Wright, S. 1943. Isolation by distance. Genetics 28:114–138.

Corresponding Editor: H. L. Gibbs

AUTHOR'S NOTE ADDED IN PROOF: Recently, G. Hoelzer has also developed models that show how population structure can arise in continuously distributed species. See: Hoelzer, G. 2001. The self-organization of population substructure in biological systems. InterJournal of Genetics, no. 345, available at http://www.interjournal.org/cgi-bin/manuscript_abstract.cgi?59390.

ERRATUM

Please note that equation 1 on page 2385 should read as follows:

$$\omega_{rep} = \frac{\exp\left[-\frac{(x_{even} - x_0)^2}{2w^2}\right]}{\exp\left[-\frac{(x_{real} - x_0)^2}{2w^2}\right]}.$$

This is the correct equation that was used to calculate expected reproductive output in the forward-in-time simulations.