UC Irvine
UC Irvine Previously Published Works

Title

Estimation of Single Generation Migration Distances from Geographic Variation in Animal Mitochondrial DNA

Permalink

https://escholarship.org/uc/item/1bd2g193

Journal

Evolution, 45(2)

ISSN

0014-3820

Authors

Neigel, Joseph E
Ball, R Martin
Avise, John C

Publication Date

1991-03-01

DOI

10.2307/2409675

Copyright Information

Peer reviewed

ESTIMATION OF SINGLE GENERATION MIGRATION DISTANCES FROM GEOGRAPHIC VARIATION IN ANIMAL MITOCHONDRIAL DNA

Joseph E. Neigel,[1] R. Martin Ball, Jr.,[2] and John C. Avise[2]

[1]*Department of Biology, University of Southwestern Louisiana, Lafayette, LA 70504 USA*
[2]*Department of Genetics, University of Georgia, Athens, GA 30602 USA*

Abstract.—A new approach is introduced for the analysis of dispersal from the geographic distributions of mtDNA lineages. The method is based on the expected spatial distributions of lineages arising under a multigeneration random walk process. Unlike previous methods based on the predicted equilibria between genetic drift and gene flow, this approach is appropriate for non-equilibrium conditions, and yields an estimate of dispersal distance rather than dispersal rate. The theoretical basis for this method is examined, and an analysis of mtDNA restriction site data for *Peromyscus maniculatus* is presented as an example of how this approach can be applied to empirical data.

A primary goal of geographic surveys of genetic variation in natural populations has been to determine rates and patterns of gene flow. The inferences derived from these surveys are generally based on genetic drift theory, which specifies a relationship between gene flow and the similarity of allele frequency distributions among populations (Slatkin, 1987). Estimation of Wright's F_{st} and related parameters that represent the correlation of alleles in subpopulations (Wright, 1951, 1965; Weir and Cockerham, 1984), and the application of Slatkin's rare allele methods (Slatkin, 1981, 1985), lead to inferences about the rate at which genes are exchanged among subpopulations. These population genetic methods avoid the practical difficulties of direct observations of dispersal, and the spatial and temporal limitations associated with mark and recapture studies (discussed in Southwood, 1978).

In this paper we utilize geographic variation in the animal mitochondrial genome to estimate gene flow by an approach that is in some ways analogous to mark and recapture data gathered over a large number of generations. This approach is motivated by the observation that for many animal species, mitochondrial DNA (mtDNA) lineages appear not to have reached equilibrium distributions across the species' range (see examples in Avise et al., 1987). Non-equilibrium distributions, while complicating the application of genetic drift theory, provide the basis for the analysis employed here. The geographic distribution of a lineage (or "clade") is interpreted as the result of a multigeneration process of dispersion from a single site of origin.

The objective of this analysis will be to estimate an "average" single-generation dispersal distance from a geographic survey of mtDNA variation. We used a published data set on the deer mouse, *Peromyscus maniculatus* (Lansman et al., 1983), to provide an example of how this approach could be applied. There are three components to the analysis: (1) development of a model of the dispersal process, (2) estimation of ages of maternal lineages from empirical data on mtDNA sequence divergence, and (3) estimation of single-generation dispersal distances from observed variances in the geographic distributions of mtDNA lineages of defined ages. For the purposes of this exploratory investigation, we will use a simplified approach to each component of the analysis. However, it may often be appropriate to tailor some or all parts of the analysis to specific organisms or sampling designs.

MODEL OF THE DISPERSAL PROCESS

Description of the Model

Animal mitochondrial DNA (mtDNA) genomes normally appear to be transmitted from maternal parent to progeny (Avise, 1986). This matrilineal transmission parallels the process of dispersal from the maternal birth site of one generation to the birth sites of the next, and provides the structure of our basic model. The dispersal

of mtDNA lineages was modeled as a random walk process, much the same way that dispersal of individuals over a single generation has been modeled (Richardson, 1970).

For dispersal in one dimension, the model is as follows: in any generation, each progeny disperses from its site of birth to a new location, d, according to some probability distribution, $F(d)$. The exact form of F need not be specified. If it is assumed that F is symmetric around the birth site, the expected value of d is zero. The standard deviation of F, σ_F, is a measure of the single-generation dispersal distance. We will refer to this parameter as the "standard dispersal distance." Other measures, such as the arithmetic mean dispersal distance, can be estimated from σ_F if the form of the distribution, F, is known. However, the primary goal of this analysis is to estimate the parameter, σ_F. If F is constant between generations and σ^2_F is the variance of F, after n generations the variance of the probability distribution for the location of a descendant relative to an ancestor will be $n\sigma^2_F$. Since in general the past location of the ancestor will not be known, we will instead make use of the probability distribution, $g(d)$, for the distance between a pair of codescendants related by a common maternal ancestor n generations removed. This follows essentially the same random walk model as the ancestor–descendant distribution, but with twice the number of steps. Thus for a pair of codescendants, the variance of the probability distribution for the distance between them will be $\sigma^2_g = 2n\sigma^2_F$. For estimation purposes, this distribution is equivalent to the joint probability distribution of the relative geographic positions of a set of independent codescendant pairs, $G(d)$. It follows that an estimate of the standard dispersal distance, σ_F, is $(\sigma^2_G/2n)^{1/2}$.

By the central limit theorem, the probability distribution of the distance between descendants, $G(d)$, will approach a normal distribution as n increases. The fit of observed distributions to normal distributions could be one test for the validity of the random dispersal model. Deviations from normality may also suggest alternative models. For a two-dimensional random walk, a bivariate normal distribution is expected.

However, for the analysis of two-dimensional distributions, it may be more informative to analyze north/south and east/west migration separately. Information is lost by combining two spatial coordinates into a single distance measure, and dispersal may be direction dependent. For the examples presented here, we have analyzed the longitudinal and latitudinal components of geographic distributions separately.

As will be shown below, there may be practical difficulties in the estimation of the parameter σ^2_G. This has led us to consider the alternative use of the geographic distribution of a family of codescendants related by a common ancestor n generations removed, which we will designate $H(d)$. An important distinction between $H(d)$ and $G(d)$ is that some members of a family of codescendants may share common ancestors more recent than the ancestor of n generations in the past, and therefore the paths of descent from ancestor to descendants would not be independent of one another.

ESTIMATION OF THE DISPERSAL PARAMETER

If independent pairs of related individuals could be sampled randomly from a population (or species), the estimation of σ_F from σ^2_G would be relatively straightforward, as shown above. However, leaving the problem of determining relatedness (the parameter n) aside, there are two principal obstacles to obtaining suitable samples from natural populations. First, because many randomly selected pairs of individuals will share portions of their paths of descent through a phylogenetic tree, it may be difficult to obtain a large sample of independent codescendant pairs. This problem becomes especially severe if pairs separated by large numbers of generations are required.

The second problem is in obtaining a sample of individuals that is random with respect to geographic location. Ideally, every individual in the population or species would be identified, and then a random subset would be selected for the analysis. However, in many cases, an individual is sampled precisely because it occurs at a specific geographic location (a collecting site), and thus is not selected at random. The problem

is especially acute if multiple individuals are sampled from the same site, since this would represent an extreme bias toward small geographic distances within the sample. This problem can be stated in terms of conditional probabilities. We need to estimate the probability of observing a pair of individuals separated by a geographic distance $d = x$, conditional on their having a common ancestor $n = r$ generations ago, while our data provide instead an estimate of the probability of observing $n = r$ conditional on $d = x$.

For those cases in which the above two sampling problems can be avoided, standard dispersal distance can be estimated directly as $(\sigma^2_G/2n)^{1/2}$. However, in the interest of making this approach more generally applicable, we have attempted to develop an alternative method for the more general case in which individuals are collected from a number of geographic locations, and are not independent with respect to ancestry. The goal of this method is to estimate, or "reconstruct" the geographic distribution of a lineage from its distribution among a set of collecting sites. The variance of this estimated distribution represents an estimate of σ^2_H, the variance of the positions of a family of descendants relative to a single ancestor. This will then be used as a surrogate for σ^2_G to estimate the standard dispersal distance. We will refer to this estimator of the standard dispersal distance as δ.

A computer program "PHYFORM" was written in Microsoft C v 5.1 and run on a Zenith Data Systems Model 386 microcomputer to reconstruct geographic distributions and estimate standard dispersal distances. Copies of this program may be obtained from J. Neigel. Input data consisted of (1) a list of sampled individuals that are grouped by geographic locale, (2) the x and y coordinates of the locales, and (3) a phylogenetic tree for these individuals, in the form of a list of the number of generations separating adjacent termini of the tree (see Neigel and Avise, 1986 for a description of this representation).

For the *Peromyscus* data, the angular longitude and latitude coordinates of the collecting sites were transformed into planar x and y coordinates as follows. The minimum latitude and longitude among collecting sites

was used as the origin of the coordinate axes, and conversion factors were calculated by assuming the earth is a sphere with a radius of 6,376 km. The latitude conversion factor was a constant 111.3 km per degree of latitude from the origin, while the longitude conversion factor was a function of the latitude: $6{,}376 \times 2\pi \times \cos(\text{latitude})/360$.

The molecular data used for our analysis came from an mtDNA survey of *Peromyscus maniculatus* conducted by Lansman et al. (1983). Using 8 restriction endonucleases to assay 136 individuals collected from 40 locales across the species' range in North America, a total of 61 different mtDNA haplotypes was observed. A phylogenetic tree was estimated from this data by the following approach. First, the site method of Nei and Li (1979) was used to calculate a matrix of sequence divergence values from the restriction site data. Then a tree matrix was generated by the SAHN program in NTSYS version 1.50 with the UPGMA clustering option (Sneath and Sokal, 1973; Rohlf, 1988). For estimation of times of divergence, we used a conventional mtDNA "clock" calibration for mammalian lineages: 1% sequence change per million years in each lineage (Moritz et al., 1987).

Our calculations of standard dispersal distance for *Peromyscus* assume a generation time of 0.2 years. This is based on observations of gestation periods of 24 days and a mean age of first estrus of 49 days (Layne, 1968).

The PHYFORM program performed a nested series of operations on the input data. The distributions of lineages among discrete locales were used to "reconstruct" geographic distributions from which variances were estimated. The parameter δ was then estimated for lineages of different ages, and weighted averages of δ were calculated for each of 10 lineage "age classes."

Geographic distributions were reconstructed by the following procedure. For each locale, i, a given lineage, k, was found at a certain frequency, f_{ik}, relative to other lineages also present. Each such frequency was transformed into a relative "abundance" value:

$$a_{ik} = f_{ik}/\sum f_{ik}$$

such that $\sum a_{ik} = 1.0$. The center of the kth

lineage (X_k, Y_k) was then estimated as the abundance-weighted average of the coordinates (x_k, y_k) of each locale:

$$X_k = \sum a_{ik} x_i$$
$$Y_k = \sum a_{ik} y_i$$

A composite distribution of abundance as a function of distance from the lineage center was then built from all lineages of a specified age. Twenty distance classes were obtained by dividing the maximum observed distance into 20 equal intervals. (From visual inspection, this number of classes consistently yielded histograms resembling the expected normal distributions.) An average abundance value was then calculated for each class, and these values were scaled so that they summed to one. The variance of this distribution was calculated and used to estimate the standard dispersal distance (δ) for all lineages of a specified age, by taking the square root of the quotient of the variance divided by the lineage age:

$$\delta = (\sigma^2/n)^{1/2}$$

An example of a geographic distribution histogram reconstructed from the *Peromyscus* data is shown in Figure 1.

SIMULATION OF DISPERSAL MODEL

There are several potential sources of error in our estimates of the standard dispersal distance parameter, σ_F. These include sources that can be associated with the estimation of divergence times between lineages, with our model of dispersal and demographic processes, and with our method of reconstructing the geographic distributions of lineages. The accuracy of "molecular clock" based estimates and the applicability of the underlying model must be judged for specific cases, and will not be treated further here. Bias and variance in δ, the estimator of standard dispersal distance, are, however, general statistical properties of our method that warrant examination.

To explore the relationship between the distributions G and H and to assess the accuracy of our statistical methods, we employed computer simulations of populations exhibiting random walk dispersal. The
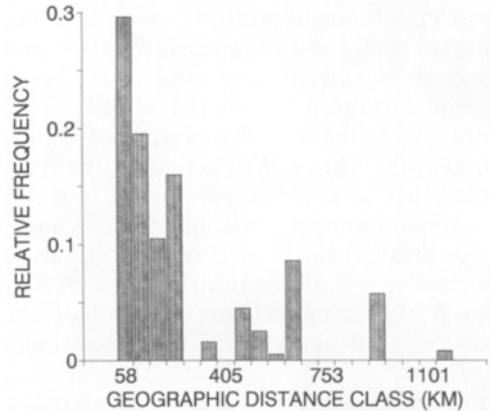

FIG. 1. An example of a composite distribution for *P. maniculatus* mitochondrial DNA lineages. Geographic distributions of seven lineages, each bound by a limit of 0.2% sequence divergence, were combined. A variance of 116,683 km² was estimated from this distribution.

simulations were an extension of those developed earlier for examining the development of mtDNA phylogenies under various demographic conditions (Avise et al., 1984; Neigel and Avise, 1986). For each discrete, nonoverlapping generation, a variable number of progeny was assigned to every female in the population. The number of progeny followed a Poisson distribution with the parameter, λ, adjusted to provide density-regulated population growth:

$$\lambda = 2 \exp[(K - N)/K]$$

Replicate simulations were run for populations with a standard dispersal distance of 1 unit along each coordinate axis. Two-dimensional geographic coordinates were assigned to progeny as integer values between 1 and 10,000. These coordinates were generated by either adding or subtracting 1 to the x and y coordinates of the female parent, depending on the value produced by a random number generator. Populations were initiated with 1,000 individuals, regulated at a carrying capacity of 1,000, and allowed to run until the entire population became monophyletic. The model's output consisted of the geographic position of each individual and an exact maternal phylogeny for the population. The programs were written in Microsoft C v 5.1 and run on a Zenith Data Systems Model 386 microcomputer. A more detailed description of these simulations is in preparation.
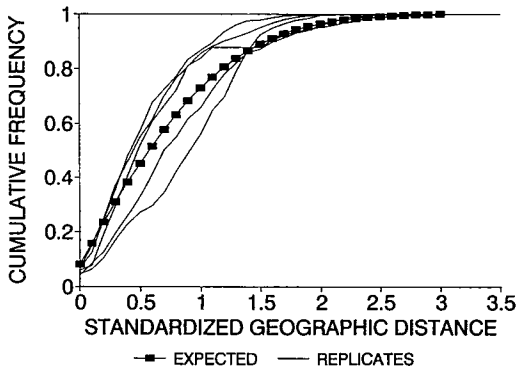
FIG. 2. Cumulative frequency distributions of geographic distances in the simulations of random walk dispersal of maternal lineages. Results are shown for five replicate simulations, compared with a standard normal distribution.

TABLE 1. Comparison of simulations of random walk dispersal of maternal lineages with standard normal distribution. Number of pairs sampled, maximum deviation of cumulative distribution function from normal distribution, Kolmogorov–Smirnov test P value, and the standard deviations of the distribution are shown for each replicate and a composite of all replicates.

Replicate	N	D_{max}	P	σ
1	505	0.23	10^{-23}	1.001
2	499	0.12	10^{-7}	0.990
3	477	0.13	10^{-7}	0.710
4	543	0.16	10^{-12}	0.645
5	490	0.13	10^{-7}	0.876
Composite	2,514	0.06	0.10	0.999

To examine the use of δ as an estimator of σ_H, discrete locale samples were generated from the simulations by a method that was intended to produce data sets comparable in structure to the *Peromyscus* data. A list of all the individuals in the population was sorted by x-coordinate position, and then partitioned into eight equally sized groups. Within each of these groups, a list of individuals was sorted by y-coordinate position. From each of these lists, five sets of three consecutively listed individuals were selected, with the maximum possible spacing between them. Each set of three individuals was treated as a sample from a single locale, and the means of the geographic coordinates of individuals within a locale were used as the locale's coordinates. This procedure created samples of 120 individuals divided equally among 40 locales.

RESULTS OF SIMULATION
ANALYSIS

In the simulations, the standard dispersal distance, σ_F, and the maternal phylogeny for the entire population were known precisely. This allowed us to examine the effects of the nonindependence of related lineages within a population on the distribution of multigeneration dispersion distances. For five replicate simulations we compared the distribution of geographic distances between pairs of randomly selected individuals that are separated by n generations from a common ancestor with normal distributions. Pairs of individuals were sampled at random without replacement from each population and combined into a standard distribution by dividing each observed geographic distance by the expected random walk geographic distance for n generations of separation. These combined distributions were then compared with a standard normal distribution, which represents the distribution $G(d)$ for independent pairs.

As shown in Figure 2 there is a fair correspondence between the distributions of geographic distances in the simulations and the standard normal distribution that assumes independence in the paths of descent. A statistical analysis of these distributions is shown in Table 1. Although for each replicate the deviation from a normal distribution was highly significant (Kolmogorov–Smirnov test), the deviation of the composite distribution was only weakly significant. Furthermore, standard deviations of individual replicates were close to 1.0 (the value expected for independent pairs). Thus it appears that the effect of the nonindependence of lineages is to create departures from normality that are not strongly biased. This is in agreement with analytical results (Neigel and Davis, in preparation).

The simulations were also used to examine the relationship between the geographic distribution of a family of maternal codescendants, $H(d)$, and the age of the lineage that it represents. Thus rather than building a distribution from pairs of randomly sampled individuals, the members of each lineage were sampled as groups. For each of 100 replicates, the variance of every lineage within the population was calculat-
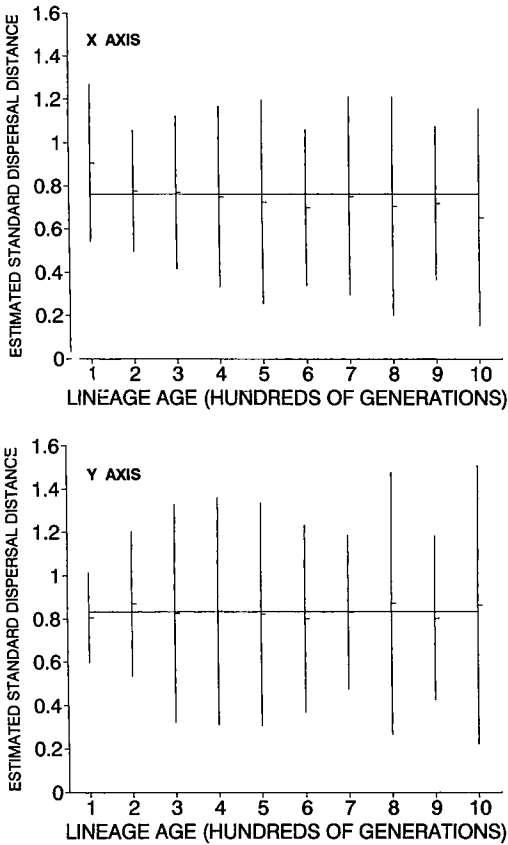
FIG. 3. Means and standard deviations of the standard dispersal distance estimator δ versus lineage age class for 50 replicate simulations. Results for x-coordinate and y-coordinate values from the same set of replicates are shown separately. The actual value of the standard dispersal distance is 1.0.

TABLE 2. Estimates of distance variances from the geographic distributions of mtDNA lineages in *Peromyscus maniculatus*.[1] D is the upper limit of sequence divergence that defines an age class of lineages, N is the number of lineages in that age class, n is the median age of lineages in that age class (in millions of generations), and σ^2_X and σ^2_Y are the estimates of longitudinal and latitudinal variances, respectively, in km^2.

D	N	n	σ^2_X	σ^2_Y
0.004	23	1	26,935	19,147
0.008	19	3	127,249	88,920
0.012	7	5	183,306	189,519
0.016	4	7	207,481	156,739
0.020	2	9	483,147	292,274
0.024	1	11	290,501	235,824
0.028	1	13	323,227	266,585
0.032	1	15	816,468	452,951
0.036	1	17	743,786	532,533
0.040	1	19	1,438,328	338,861

ed, and used to calculate a standard deviation, σ_H, as an estimate of σ_F.

Across all replicates and lineage ages, the average value of σ_H was 0.831 (the true value of the dispersal parameter was 1.0). The minimum value obtained for any one replicate was 0.763, the maximum was 0.926, and the standard deviation among replicates was 0.0336. These results indicate that the geographic range of randomly dispersing individuals within a branching phylogenetic lineage tends to be about 17% smaller than an analogous random walk distribution.

The statistical properties of the δ estimator were assessed by analysis of simulations in which the geographic distributions of lineages were reconstructed from discrete locale samples. Fifty replicate simulations were run, and values of the esti-

mator δ were calculated for each of 10 lineage age classes in each replicate. Figure 3 shows the means and standard deviations of these replicates. Combined estimates were also calculated for each replicate, in which lineages from all age classes and distributions along both x and y axes were weighted equally. The mean combined estimate was 0.848, with a minimum of 0.663, a maximum of 1.591, and a standard deviation of 0.136 among replicates. This mean δ value does not differ significantly from the mean value of σ_H estimated from the simulations (t test, $\alpha = 0.05$). These results suggest that δ provides an estimate of σ_H with little or no bias.

RESULTS OF ANALYSIS OF *PEROMYSCUS* DATA

Table 2 shows the geographic variance estimates for *Peromyscus maniculatus* mtDNA lineages in 10 different age classes. As expected for nonequilibrium distributions, higher variances were associated with older lineages. Spearman rank correlation coefficients indicated that these relationships were significant: $r_s = 0.952$ ($P = 0.004$) and $r_s = 0.915$ ($P = 0.006$) for the longitudinal and latitudinal components, respectively. Under a random walk model in which the variance of the distributions would increase at a constant rate, these relationships are expected to be linear. Linear correlation coefficients were $r^2 = 0.76$ ($P = 0.001$) for
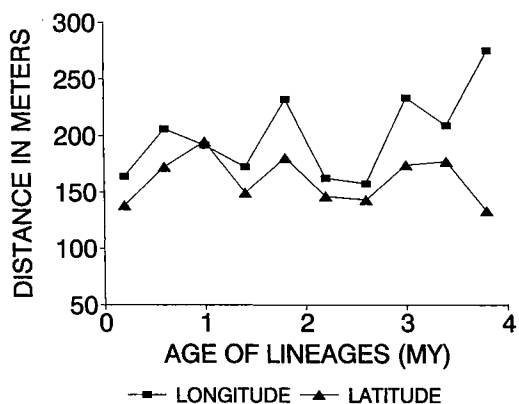
FIG. 4. Estimated standard dispersal distances versus age of lineage for the mtDNA data in *P. maniculatus*. Longitudinal and latitudinal components were estimated separately. Calculations were based on a rate of 1% sequence change per lineage per million years, and a generation time of 0.2 years.

longitude and $r^2 = 0.78$ ($P = 0.001$) for latitude.

Estimates of the dispersal parameter, δ, can be calculated from the values in Table 2 as follows: First, divide the distance variance (shown in kilometers) for either longitude or latitude by the median age for the age class (shown in millions of generations) to obtain the single generation variance. The square root of the single generation variance yields the single generation standard deviation, δ. For example, the value of δ for the longitude axis and lineages defined by a limit of 0.4% sequence divergence was calculated from the first line of Table 2 as follows:

$$\delta = \sqrt{\frac{26,935 \text{ km}^2}{1,000,000}} = 0.164 \text{ km}$$

The overall estimate of the parameter δ for all lineage age classes was 187 m for longitude and 159 m for latitude. Estimates of the parameter δ for specific age classes are shown in Figure 4. These values of δ ranged from 133 to 275 m along a single directional axis, with a mean of 200 m in longitude and 161 m in latitude. After correction for an assumed bias of 0.85, the overall estimates of the standard dispersal distance are 220 and 187 m for longitude and latitude, respectively. Corrected age class-specific estimates range from 156 to 322 m. These estimates agree well with previously reported dispersal distances based on direct mark-recapture data (see beyond).

Although we are reluctant to suggest that confidence limits can be accurately determined for our estimates of σ_F without consideration of additional sources of variance, the simulation results can be used to assess the accuracy of the estimation procedure itself. For combined estimates of the parameter δ, 95% of the replicates were within 32.7% of the true value of the standard dispersal distance, σ, and 99% were within 34.0% of the actual value.

DISCUSSION

Assumptions and Caveats of the Dispersal Model

The interpretation of nonequilibrium geographic distributions of genetic variation requires an approach that differs from the genetic drift paradigm of classical population genetics. If the rate at which new genetic variants arise is high relative to the rate at which they are dispersed geographically, then lineage distributions will not reflect an equilibrium between genetic drift and gene flow. The *P. maniculatus* data set analyzed here appears to reflect a nonequilibrium distribution, because there was a strong positive correlation between the ages of lineage separations and the variances of their geographic distances (Table 2). If these lineages had achieved equilibrium distributions, no correlation with age would be expected. The results of the simulations are also consistent with this interpretation. They demonstrate that the observed restriction of lineages to portions of the species' total geographic range and the correlation between lineage age and geographic variance is an expected consequence of a high rate of sequence divergence relative to the rate of dispersal. Nonetheless, other factors, such as genetic drift and historical biogeographic effects, are also likely to influence the distributions of lineages, and are not excluded by this interpretation. Future efforts will examine the combined effects of these structuring processes.

The assumptions required for interpretation of nonequilibrium distributions are different from those required for genetic drift interpretations. The most basic assumptions are those that constitute the underlying model that specifies how genetic vari-

ants are dispersed over time. The simplest model is a random walk, which assumes that the direction and distance of each step is independent of previous steps. There are several ways in which this model could be inappropriate. For example, geographic variation in dispersal distances, including that arising from barriers to dispersal, would create correlations between successive steps in the dispersal process. Furthermore, there are generally geographic limits to the range of a species. These limits would truncate the distributions of lineages relative to the expectations for a random walk. If additional information on dispersal patterns were available, it could be incorporated into a more detailed model. The goal we have set here, the estimation of a single average dispersal distance for the entire species, should not be considered the only object of the nonequilibrium approach. In many instances, it may be appropriate and informative to consider how the observed patterns of dispersal depart from a simple random walk process. In this respect, the random walk model could be viewed as a null hypothesis against which nonrandom components of dispersal could be detected.

A second set of assumptions concerns the molecular clock model that specifies the relationship between sequence divergence and the time of lineage separation. Because the accumulation of mutations is likely to be a stochastic rather than metronomic process, the accuracy of the estimate of divergence time will depend on the number of mutations on which the estimate is based, which in the case of *P. maniculatus* (or similar data sets for other species) is not large.

The clock calibration itself is another potential source of error. Direct calibrations of mtDNA sequence divergence rates based on paleontological data are available for only a few species, and appear to vary considerably among taxa (though probably not by more than an order of magnitude—Moritz et al., 1987). Furthermore, it is unclear whether molecular clock calibrations should be based on absolute time or generation time. For purposes of the current illustration, we employed a conventional clock calibration based on absolute time, and assumed a constant relationship between absolute time and generation time for *P.*

maniculatus. In general, this latter assumption could be eliminated if an estimate of dispersal distance per year were used in place of dispersal distance per generation. Also, because the final estimates of standard dispersal distance are based on square roots of divergence times and variances in geographic distributions, the effects of errors in these estimates may be reduced.

The statistical assumptions used in this analysis are perhaps the most difficult to evaluate. We have not considered all sources of variance to establish confidence limits for dispersal distances, and we have not developed statistical tests for differences in dispersal distance parameters. Errors could arise from anomalies in the actual geographic distributions of lineages, estimates of these geographic distributions, and inferences of divergence time from molecular data. The nonindependence of lineages within a species makes it difficult to determine the degrees of freedom for the estimate of the variance of $G(d)$, the geographic distance between a set of independent ancestor–descendant pairs. Although this problem needs further exploration, results of the computer simulations presented here suggest that the distributions of a family of descendants relative to a single ancestor provide a reasonably good estimate of $G(d)$.

Application to Peromyscus maniculatus

The deer mouse is found in a variety of habitats in North America, from southern Mexico to the Northwest Territories of Canada, and from the Pacific to Atlantic coasts. We chose *P. maniculatus* for this initial application of the nonequilibrium dispersal models for two major reasons. First, the combination of limited mobility, extensive geographic distribution, and high values of estimated mtDNA sequence divergence (Lansman et al., 1983) makes the species a suitable candidate for the method of analysis proposed here. Second, results of several mark and recapture studies, which provide direct estimates of dispersal distances, are available for comparison.

In *P. maniculatus,* dispersal away from the home ranges of the parents generally occurs as young mice approach maturity. Mature mice may also shift the positions of their home ranges. Stickel (1968) tabulated

distributions of dispersal distances for juveniles from three field studies. The average dispersal distance in each case was less than 150 m. In a study of *P. m. bairdi,* adult females shifted the location of their home ranges by an average of 230 m (Blair, 1940). A study of populations in southern Michigan (Dice and Howard, 1951) yielded a distribution of dispersal distances from birth site to breeding site with a standard deviation of 264 m. From these studies, it appears that the standard dispersal distance for *P. maniculatus* is probably between about 100 and 300 m.

The overall bias-corrected mtDNA-based estimates of standard dispersal distance for *P. maniculatus* were 220 m for longitude and 187 m for latitude. Bias corrected estimates for the youngest age class of lineages, which are most likely to have dispersed under climatic and biogeographic conditions equivalent to those experienced by contemporary populations, were approximately 180 m for both latitude and longitude. If it is assumed that the two-dimensional (2-D) distributions of single generation dispersal distances are bivariate normal, then the overall estimates correspond to a mean 2-D dispersal distance of roughly 250 m (approximately 1.25 times the 1-D standard deviation), and the estimates for the youngest lineages correspond to a 2-D dispersal distance of about 225 m. These estimates are thus consistent with the results of the mark and recapture studies.

Estimates of standard dispersal distances varied with the ages of the lineages on which they were based (Fig. 4). If the pattern is indeed significant, it could reflect actual variation in standard dispersal distance during the species' history. Alternatively, various assumptions of the random walk model, such as uniform dispersal over the species range, may not always hold.

We suspect that the times and spatial scales of historical barriers to gene flow could differentially affect estimates of dispersal based on the dispersion of lineages of different ages. For example, a long-standing geographic barrier to movement might reduce the dispersion of older lineages more so than recent ones, resulting in a relative decrease in dispersal distances for older lineages. Conversely, very recent geographic or ecologic barriers to gene flow might result in an apparent underdispersion of younger lineages compared to older lineages that dispersed before the barriers arose (perhaps the situation observed in *P. maniculatus*).

Clearly, many questions remain to be explored in the application and interpretation of nonequilibrium models to issues of dispersal and population structure. Since many species (such as *P. maniculatus*) consist of populations with a strong phylogeographic orientation, further attempts to add a historical dimension to the study of demographic processes seem desirable.

LITERATURE CITED

AVISE, J. C. 1986. Mitochondrial DNA and the evolutionary genetics of higher animals. Phil. Trans. R. Soc. London Ser. B 312:325–342.

AVISE, J. C., J. ARNOLD, R. M. BALL, 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.

AVISE, J. C., J. E. NEIGEL, AND J. E. ARNOLD. 1984. Demographic influences on mitochondrial DNA lineage survivorship in animal populations. J. Mol. Evol. 20:99–105.

BLAIR, W. F. 1940. A study of prairie deer-mouse populations in southern Michigan. Am. Midland Nat. 24:273–305.

DICE, L. R., AND W. E. HOWARD. 1951. Distance of dispersal by prairie deer mice from birthplaces to breeding sites. Cont. Lab. Vert. Biol. Univ. Mich. 50:1–15.

LANSMAN, R. A., J. C. AVISE, C. F. AQUADRO, J. F. SHAPIRA, AND S. W. DANIEL. 1983. Extensive genetic variation in mitochondrial DNA's among geographic populations of the deer mouse, *Peromyscus maniculatus.* Evolution 37:1–16.

LAYNE, J. N. 1968. Ontogeny, pp. 148–253. *In* J. A. King (ed.), Biology of *Peromyscus* (Rodentia). Amer. Soc. Mamm. Spec. Pub. No. 2.

MORITZ, C., T. E. DOWLING, AND W. M. BROWN. 1987. Evolution of animal mitochondrial DNA: Relevance for population biology and systematics. Annu. Rev. Ecol. Syst. 18:269–292.

NEI, M., AND W.-H. LI. 1979. Mathematical model

for studying genetic variation in terms of restriction endonucleases. Proc. Natl. Acad. Sci. U.S.A. 76: 5269–5273.

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, N.Y.

RICHARDSON, R. H. 1970. Models and analyses of dispersal patterns, pp. 79–103. *In* K. Kojima (ed.), Mathematical Topics in Population Genetics. Springer-Verlag, N.Y.

ROHLF, F. J. 1988. NTSYS-pc Numerical Taxonomy and Multivariate Analysis System, Version 1.50. Exeter Publishing, Ltd., N.Y.

SLATKIN, M. 1981. Estimating levels of gene flow in natural populations. Genetics 99:323–335.

———. 1985. Rare alleles as indicators of gene flow. Evolution 39:53–65.

———. 1987. Gene flow and the geographic structure of natural populations. Science 236:787–792.

SNEATH, P. H. A., AND R. R. SOKAL. 1973. Numerical Taxonomy. Freeman, San Francisco.

SOUTHWOOD, R. 1978. Ecological Methods: With Particular Reference to the Study of Insect Populations. Chapman and Hall, London.

STICKEL, L. F. 1968. Home range and travels, pp. 373–411. *In* J. A. King (ed.), Biology of *Peromyscus* (Rodentia). Am. Soc. Mamm. Spec. Publ. No. 2.

WEIR, B. S., AND C. C. COCKERHAM. 1984. Estimating F-statistics for the analysis of population structure. Evolution 38:1358–1370.

WRIGHT, S. 1951. The genetical structure of populations. Ann. Eugen. 15:323–354.

———. 1965. The interpretations of population structure by F-statistics with special regard to systems of mating. Evolution 19:395–420.

Corresponding Editor: A. G. Clark