A COMPARISON OF ALTERNATIVE STRATEGIES FOR ESTIMATING GENE FLOW FROM GENETIC MARKERS[1]

Joseph E. Neigel

Department of Biology, University of Southwestern Louisiana, Lafayette, Louisiana 70504; e-mail: jneigel@usl.edu

ABSTRACT

The estimation of gene flow from the distribution of genetic markers in populations requires an indirect approach. Gene flow parameters are defined by demographic models, and population genetic models provide the link between these parameters and the distributions of genetic markers. Following the introduction of allozyme methods in the 1960s, a standard approach to the estimation of gene flow was developed. Wright's island model of population structure was used to relate the distribution of allozyme alleles in populations to $N_e m$, the product of the effective population size and the rate of migration.

Alternative strategies for the estimation of gene flow have been developed using different genetic markers, different models of demography and population genetics, and different methods of parameter estimation. No alternative strategy now available is clearly superior to the standard approach based on Wright's model and allozyme markers. However, this may soon change as methods are developed that fully utilize the genealogical relationships of DNA sequences. At present, alternative strategies do fill important needs. They can provide independent estimates of gene flow, measure different components of gene flow, and detect historical changes in population structure.

---

[1]The US government has the right to retain a nonexclusive, royalty-free license in and to any copyright covering this paper.

INTRODUCTION

In the late 1960s, surveys of allozyme variation in natural populations set a new direction for studies of gene flow (61–65). Wright's analysis of nonrandom mating provided a quantitative relationship between his measure of population subdivision, F_{ST}, and the two parameters: population size, N, and rate of migration, m (127, 130–133). Because F_{ST} can be interpreted as the standardized variance in the frequency of an allele among populations (132), it could be readily estimated from allozyme data. A standard approach to quantifying gene flow among populations was put into practice, using allozyme markers to estimate F_{ST} and Wright's island model (127, 129) to transform estimates of F_{ST} into estimates of Nm (95, 97). This approach is considered an indirect method (95) for the estimation of gene flow, because population genetic models are required to infer the magnitude of gene flow from its effects on the distributions of genetic markers. Throughout this review, I refer to F_{ST} analysis of allozyme data as the F_{ST}-allozyme approach.

The F_{ST}-allozyme approach has been widely applied to natural populations of plants, animals, and microorganisms. Its many advantages explain its continued use. Allozyme surveys require relatively little expense, time, or specialized laboratory equipment, and they can be successfully carried out for most species (63). The theoretical basis of this approach is well established, which has led to progressive refinements in how F_{ST} is estimated (21, 102, 124). However, there are compelling reasons for investigating alternative strategies. First, it is difficult to directly test the reliability of Nm estimates. Our confidence in them rests largely on their robustness under diverse conditions in theoretical models (70, 95, 97, 102). Empirical tests such as comparisons with mark-and-recapture estimates of dispersal are at best ambiguous. Indirect methods estimate the cumulative effects of gene flow, acting over all temporal and spatial scales. In contrast, direct estimates of gene flow apply only to the interval of time and space over which observations are made (95). Alternative strategies are also needed to better utilize new types of data now available to population geneticists. A shift in emphasis from classical Mendelian alleles to DNA sequences has changed the way population genetic variation is described. DNA sequences represent a more diverse set of genetic markers than allozymes, and we now have the tools to survey them in natural populations (68). It is also possible to infer genealogical relationships among DNA sequences (4, 62). Interpretation of these relationships has led to the development of coalescent models (43, 53, 54, 115) and to more powerful methods for the estimation of population genetic parameters (34, 37). We can now develop new strategies to estimate gene flow from alternative DNA-based genetic marker systems, population genetic models, and statistical methods. It is not yet clear which, if any, single

new approach will replace the F_{ST}-allozyme approach as the standard method for gene flow estimation.

Although the efficiency of methods for DNA analysis has greatly improved, allozyme methods are still more accessible, and methods for the analysis of DNA sequence data are still under development. However, alternative strategies have progressed sufficiently to yield some interesting results. They have allowed us to examine aspects of gene flow distinct from those that can be examined with allozyme data alone. And in a few cases, results with DNA-based markers appear to contradict those obtained with allozyme data and thus have led us to reevaluate F_{ST}-allozyme based estimates of gene flow.

In this review, I first identify the basic methodological, theoretical, and statistical components necessary for the indirect estimation of gene flow from genetic markers. I examine alternatives for each of these components and how they may be used in gene flow estimates. Finally, I consider the assumptions made about the scale of relevant parameters in the estimation of gene flow, and how scale should influence our interpretations of these estimates.

COMPONENTS OF AN INDIRECT APPROACH

Indirect methods of gene flow estimation characterize the spatial distribution of genotypes by some parameter and then apply a population genetic model to ask what level of gene flow would produce a distribution with the same parameter value (95, 97). The logic of indirect methods is thus more complex than estimation from direct measurements, a point first carefully articulated by Weir & Cockerham (124). Organisms are sampled from populations, and a statistic is calculated from the distribution of a genetic marker in the samples. This statistic is then used to estimate a parameter of the genetic marker's distribution in the populations. Estimates of this distribution parameter from one or more genetic markers are combined and used with a population genetic model to estimate a gene flow parameter. Four components are combined by this logic: a demographic model, a genetic marker system, a population genetic model, and a parameter estimator. Alternative methods for the estimation of gene flow have introduced changes in each of these components, and the significance of these modifications should be considered individually.

TERMINOLOGY

Before gene flow can be measured, it must be defined, and any quantitative definition of gene flow must be based upon a model of population structure. The terminology applied to gene flow in population models is potentially confusing because the more general meanings of the terms *migration* and *dispersal* do not

match their meanings when applied to gene flow. In a general sense, migration can be defined to include any movements, including cyclical movements that may regularly return an organism to its original location (7). In contrast, the term dispersal is more precisely restricted to movements that increase the distances between organisms, gametes, or propagules (7). Gene flow is defined as the movement of genes in populations, and thus it includes all movements of gametes, propagules, and individuals that are effective in changing the spatial distributions of genes (95). The distinction between movements of organisms and movements of genes is clearly important and should not be forgotten when interpreting estimates of gene flow. However, the terms migration and dispersal have been given secondary meanings in the population genetic literature. When we consider population genetic models, we usually refer to gene flow between discrete subpopulations as migration, and to gene flow within a continuous population as dispersal. I employ this useful division of the concept of gene flow for this review.

The terms deme, subpopulation, population, total population, and metapopulation refer to different scales of population structure, but these scales are not defined precisely or consistently. For this review, I use only two terms: subpopulation and total population. I consider a subpopulation to be a unit of population structure that exchanges migrants with other subpopulations. I consider a total population to be the set of all subpopulations that exchange migrants.

F_{ST} -ALLOZYME APPROACH

Demographic Models

Most indirect estimates of gene flow have been based on models in which migration occurs between discrete subpopulations. Each subpopulation is conceived of as an "island," with no internal spatial structure, and with gene flow via migration of individuals. These models represent a restricted subset of diverse demographic models that have been developed without a population genetic component (see for example 79, 85). The population genetic consequences of demographic models and their use in estimates of gene flow have been the subject of several reviews (32, 95, 97, 102).

The simplest migration model is the infinite island model (95), which is equivalent to the continent-island model introduced by Wright (127, 129). It features an infinitely large source population, and one or more finite subpopulations ("islands") that receive migrants from the source at a constant rate. The source population can either be a single infinitely large population (a "continent"), or an infinite number of finite subpopulations (the total population) that

both produce and receive migrants ("infinite islands"). In either case the source population is infinite and therefore not subject to genetic drift. Subpopulations are filled by N zygotes each generation, of which the fraction m are migrants from the source population. The number and relative locations of the island subpopulations are unspecified because each is dependent only on the source population.

The infinite island model leaves out much of what we expect to be true of migration in nature. However, the simplicity of this model has led to its widespread use for estimates of gene flow. Its only parameters are those of concern here, N and m. A more general model would specify a matrix of migration rates between all subpopulations, but such a model would be too complex to be useful for the indirect estimation of gene flow (32). However, some special cases with simple migration matrices have been considered.

In the finite island model or n-Island model (95), migration occurs between n finite subpopulations of size N (60). The migration rate, m, represents the fraction of each subpopulation derived from migration, with all subpopulations equally likely to serve as the source of a migrant. In *stepping stone* models (52), migration occurs only between neighboring subpopulations. The subpopulations may be arrayed in one or more dimensions. Predictions of the model are made with reference to the number of steps between pairs of subpopulations. Although direct gene flow occurs only between neighboring subpopulations, the "stepping stone" effect allows gene flow to occur between nonneighboring subpopulations via the subpopulations that connect them.

Allozyme Markers

The literature on allozymes as genetic markers is extensive. Lewontin's book (61), written less than a decade after the first applications of allozyme methods to population genetics, clearly lays out the rationale for studying allozyme variation and summarizes the early findings. Lewontin has also provided thoughtful retrospectives on how allozyme studies have influenced the development of population genetics (62, 63). A general overview of allozymes and other molecular markers in studies of natural history and evolution is provided by Avise (2).

Allozymes have several desirable properties as genetic markers for indirect estimates of gene flow. They can be assayed without formal genetic analysis and therefore can be surveyed in nearly any species. The nature of allozyme variation and the electrophoretic methods used for its detection provide markers that fit a simple Mendelian model. In higher plants and animals, most allozymes are encoded by autosomal loci that exhibit standard biparental Mendelian inheritance. Multiple polymorphic loci can be surveyed, and typically two or three

alleles are detected at each. The rates of mutations that produce detectable differences in allozyme alleles are on the order of 10^{-6} (119).

Genetic Models

Wright's work on nonrandom mating provided the theoretical foundation for estimates of gene flow from allozyme data (127, 130–133). He defined a set of correlation coefficients, which he called F-statistics, to partition departures from random mating into components due to nonrandom mating within populations and to population subdivision (132). For the continent island model of population structure with discrete, nonoverlapping generations of diploid organisms, Wright (127) found the following often-cited relationship between F_{ST}, and $N_e m$, the product of effective population size and migration rate: $F_{ST} \simeq 1/(4N_e m - 1)$. Following Slatkin (99), the quantity $N_e m$ can be considered "the amount of gene flow," and be represented by the single parameter M.

Parameter Estimation

There has been much confusion over what F-statistics actually represent, and how they should be estimated (21, 95, 124, 133). It is useful to consider three meanings. First, F-statistics can be defined as demographic parameters, without reference to any specific form of genetic variation. This corresponds to Wright's definition of an inbreeding coefficient (126). In this sense, F-statistics are parameters that define how gametes are sampled to form zygotes. The quantity relevant to gene flow is F_{ST} (132).

F-statistics may also be defined as parameters of an allele's frequency distribution among individuals and subpopulations. In the absence of selection, the expectations for these distributions correspond to those of the parameters that define how gametes are sampled. In this sense, F_{ST} is the standardized variance of p, the frequency of an allele, among subpopulations: $F_{ST} \simeq \sigma_p^2 / p(1 - p)$ (132). However, because allele frequencies are subject to stochastic variation, the value of F_{ST} is expected to vary among alleles at different loci, among alleles at a multiallelic locus, and over time for any single allele. The stochastic variance imposes a lower limit on the variance of any estimate of F_{ST} as a demographic parameter if that estimate is based on a single locus. In addition to this stochastic variance, realized values of F_{ST} may reflect the effects of selection and mutation.

F-statistics have also been defined as true statistics, quantities calculated from data. These statistics are used to estimate parametric values of F_{ST}. Weir & Cockerham (124) have recommended that parameters be given precise definitions and that they be clearly distinguished from statistics. This is a critical distinction, because the expectation of F_{ST} as a statistic is not equal to F_{ST} as a parameter. F_{ST} as a statistic adds components of variance due to sampling

individuals and subpopulation to the actual variance in allele frequencies among subpopulations (74, 124).

F_{ST} can also be related to the process that generates genealogical relationships of genes, known as the coalescent (53). The coalescence time of two genes is the number of generations in the past that separates them from a common ancestral gene. Slatkin (98) derived the following relationship between Wright's F_{ST} and coalescence times, as a limit for mutation rates that are too small to affect F_{ST}: $F_{ST} = (\bar{t} - \bar{t}_0)/\bar{t}$ where $\bar{t}_0$ is the average coalescence time of two genes drawn from the same subpopulation, and $\bar{t}$ is the average coalescence time of two genes drawn from the same total population. This relationship is important in comparisons of alternative methods of estimating gene flow, because it separates the demographic processes of migration and genetic drift from the process of mutation.

In addition to F_{ST}, some related quantities have been introduced that correspond to the different meanings that have been attached to F_{ST}. G_{ST} was originally introduced by Nei (72) to provide an explicit procedure for combining information from multiple loci and alleles in a single measure of population subdivision. Because G_{ST} was defined for samples of populations and individuals, it has been considered a statistic that can be used as an estimator of F_{ST} (124). However, the meaning of G_{ST} has broadened with subsequent usage. Crow & Aoki (23) defined it as a parameter rather than a statistic, and it has been used as a measure of population subdivision for animal mtDNA (112). N_{ST} was first introduced as a statistic based on average differences between DNA sequences in populations (64), and can be considered an estimator of F_{ST} (98) In general, the relationships among these quantities have been determined, and it has been useful to consider them as alternative estimators of the same underlying parameter (21, 98, 102, 124).

Several approaches have been taken to the problem of estimating F_{ST} and related quantities from allozyme data. As discussed above, the problem is not trivial because the statistics calculated from allozyme data are removed by several logical steps from the parameters defined in population genetic models. Weir & Cockerham (124) emphasized that F_{ST} as a parameter cannot be estimated unless it is adequately defined. They developed such a definition for a specific model in which a finite number of subpopulations act as independent realizations of the same process. With this parametric definition of F_{ST}, it is possible to define estimators that, under the conditions of the model, are unbiased. However, if the underlying model is not valid, it is unclear what is being estimated or whether the estimator's properties are preserved. For this reason, Cockerham & Weir (21) did not consider estimation of F_{ST} to be equivalent to an estimation of gene flow. Others (72, 74, 102) have focused on the theoretical link between F_{ST} and M, and the development of workable methods for the

estimation of M from population genetic data. Estimators are sought that are robust under a range of conditions, rather than well defined under restricted conditions. The standard approach for estimation of M is based on Wright's island model (127), which suggests the following estimator:

$$\hat{M} = (1 - F_{ST})/4F_{ST}.$$

A few methods have been explored that bypass the estimation of F_{ST} by using other parameters of allele frequency distributions that can be linked to M in population genetic models. Maximum likelihood estimates of M based on Wright's island model (128) assume a beta distribution for allele frequencies among subpopulations (8, 123). The parameters of the beta distribution can be expressed in terms of M and average allele frequencies among subpopulations. Slatkin introduced the "private alleles" method for the estimation of M from the frequencies of rare alleles that are found in only one or a few subpopulations (96). Although this method does not involve explicit estimation of F_{ST}, it is based on the same population genetic models and the properties of allele frequency distributions that correspond to F_{ST} (9). Spatial autocorrelation methods can be used to describe patterns of allele frequency distributions (106), detect population structure (107), and estimate gene flow parameters (105). However, migration parameters are expected to be only weakly correlated with the quantities estimated by spatial autocorrelation analysis (29, 95, 101), which are also subject to mutation, stochastic variance, and nonequilibrium conditions (101).

The process of gene flow in natural populations may be far more complex than its representation in population models. However, if a model is to be used for indirect estimation of gene flow, it may be of little value to add additional parameters if their values cannot be determined. There are differing opinions on the robustness of the methods used to estimate population genetic parameters. Lewontin (62) expressed the opinion that parameter estimates would be too sensitive to such unaccounted influences as natural selection and population history to be reliable. Slatkin (95) acknowledged Lewontin's concerns but argued that, in the case of migration parameter estimates, a more optimistic view was warranted.

Models too complex to be used for estimation of gene flow may still be useful for testing the robustness of conclusions drawn from relatively simple models. For example, Slatkin & Barton (102) used a combination of analytical models and numerical simulations to show that gene flow estimates may be robust with respect to unknown details of population structure and low levels of selection or mutation. They concluded that the F_{ST}-allozyme approach can be generalized to more complex models of population structure.

ALTERNATIVE APPROACHES

Demographic Models

METAPOPULATION MODELS   There has been much interest in the ecological and evolutionary consequences of metapopulation dynamics (see, for example, 40 and others in that volume). Recolonization can maintain an assemblage of subpopulations in which individual subpopulations are subject to frequent extinction. The term metapopulation has been applied to such assemblages, although the term is sometimes used more broadly to include almost any process that involves multiple populations (including conventional migration). The parameters of a metapopulation model define rates of population replacement and the number and source of propagules or individuals that recolonize populations.

To my knowledge, no general method has been proposed for indirect estimation of gene flow parameters in metapopulations. There may be little basis for such a method, because the effect of gene flow on the distribution of genetic variation in metapopulations is expected to depend on the details of the recolonization process, including the founding number, probability of common origin, and kin structure (94, 120), and the outcome may also be critically affected by nonequilibrium dynamics (125). If any general conclusion can be drawn from these theoretical studies, it may be that naive application of a standard method for the estimation of gene flow to populations that are subject to frequent extinctions could yield misleading results. However, it is conceivable that population genetic data could be used in conjugation with other information to estimate parameters within the framework of a metapopulation model.

SPATIAL MODELS   For organisms of relatively limited vagility, gene flow may be restricted by distance alone within a single population. Wright called this effect isolation by distance (130, 131). Although consideration of gene flow in spatial models is not new, these models have not been developed as fully as subpopulation models and are less often used for indirect estimates of gene flow. Two basic types of spatial models have been considered for estimates of gene flow. In continuum models, the locations of individuals are not specified by an array but are generated by the distributions of dispersal movements, births, and deaths within a continuous space. In the special case of every individual leaving exactly one offspring, a random distribution of individuals will develop, such as expected for particles subject to Brownian movement (92). However, if individuals leave multiple offspring, siblings will form clusters that reflect the distribution of dispersal distances, and their descendants in turn will form larger clusters that represent patterns of multigeneration dispersal. This process leads to spatial distributions with infinitely large clumps (87), which are biologically

unrealistic, and severely complicates the analysis of the population genetic consequences of these models (31, 32).

In natural populations, it can be assumed that there are mechanisms that distribute individuals more evenly than predicted by simple continuum models (1). Lattice models (102) represent the most rigid form of local density regulation. Individuals are distributed in regular arrays of one or two dimensions, and in the simplest case, exactly one individual occupies each position in the array. This restriction provides a constant population density and size, and greatly simplifies mathematical analysis. Lattice models are formally similar to island and stepping stone models, because each point on the lattice may be treated as a subpopulation. Other forms of density regulation cannot be easily treated analytically and are likely to be explored primarily with numerical simulations.

DNA Markers

ANIMAL MITOCHONDRIAL DNA   The first DNA-based genetic marker system that could be routinely applied to surveys of genetic variation in natural populations was animal mitochondrial DNA (mtDNA). This molecule was amenable to analysis by methods that were widely available in the early 1980s for manipulating plasmid DNA. Animal mtDNA could be easily purified by ultracentrifugation, and sequence differences could be detected as restriction fragment length polymorphisms (RFLPs) (15, 59). More recently, the introduction of the polymerase chain reaction (PCR) has nearly eliminated the need to directly purify specific DNA sequences (86). Improvements in DNA sequencing methods have made it feasible to determine the exact nucleotide sequence of amplified regions of mtDNA for large numbers of individuals (69).

In many important respects, mtDNA polymorphisms are quite different from allozyme alleles. Animal mtDNA is generally inherited maternally as a single linkage unit of about 15 kb (5, 12, 13, 59). Because the mitochondrial genome is transmitted without recombination as a single linkage unit, mtDNA sequence variants are usually referred to as haplotypes, rather than alleles. The inheritance of mtDNA is thus formally similar to a single haploid locus. Mutation rates are often higher for animal mtDNA than for nuclear genes (14, 118), and there are generally many more detectable polymorphisms than there are for single allozyme loci (3, 5). Perhaps most significantly, detailed characterization of sequence differences can be used to infer genealogical relationships and estimate divergence time (reviewed in 4, 13).

CHLOROPLAST DNA   The chloroplast genome is a circular molecule, typically about 160 kb in size (80). In most higher plants, chloroplast DNA (cpDNA) is uniparentally inherited—maternally in angiosperms and paternally in conifers (22, 41, 90). Variation in cpDNA sequences can be detected by either restriction

site analysis or sequencing. In these respects, cpDNA is similar to animal mtDNA. However, variation in cpDNA is more often used to infer relationships among species than as a population genetic marker. This is primarily because in land plants, cpDNA has a relatively slow rate of evolution, and consequently low levels of intraspecific polymorphism (20, 80). However, intraspecific polymorphism has been found in some species (19, 27, 36, 42, 51, 65, 71, 81, 108, 109), and so it has been possible to use cpDNA variation to investigate gene flow (27, 66).

NUCLEAR DNA   Nuclear DNA (nDNA) variation has been used to estimate gene flow in a few studies. Early studies of nDNA polymorphisms were either limited to surveys of RFLPs detected by Southern hybridization (58, 88), or they required laborious molecular cloning methods to isolate genes for nucleotide sequencing (57). The advent of the PCR (polymerase chain reactions), a growing database of sequences that can be used to design PCR primers, and automated sequencing (39, 46) have greatly increased the feasibility of nDNA sequence surveys (93). Two forms of nuclear sequence variation are widely used as genetic markers: base substitutions and variable numbers of tandem repeats (VNTRs). Base substitutions are more difficult to survey in populations but offer a greater possibility of inferring genealogical relationships among sequences. VNTRs are relatively easy to survey in populations. They are often highly polymorphic, and length variation can be detected by simple electrophoretic methods. VNTR sequences are classified by size as microsatellites or minisatellites. Microsatellites consist of up to 50 copies of tandemly repeated sequences that are each 1–10 basepairs (bp) in length. Because their total length is usually less than a few hundred bp, it is possible to analyze microsatellite length variation by direct size measurements of PCR-amplified sequences on electrophoretic gels (113a). Minisatellite sequences contain up to several hundred copies of repeat units that range from 10–200 bp in length, with total lengths up to 50 kb. Because of their large size, minisatellite sequences often cannot be amplified by PCR, and their length polymorphisms are usually detected as RFLPs with Southern hybridization (48a).

Two mechanisms are expected to generate variation in the number of tandem repeats. Recombination is expected to generate a broad distribution of length changes, at rates that may be as high as 5×10^{-2} (16, 47). Thus, in sequence comparisons, little correlation is expected between the magnitude of accumulated length differences and the number of unequal crossing over events that produced them. Replication slippage that occurs during DNA replication appears to favor small stepwise changes in the number of tandem repeats (25, 91, 117). The rates at which variants are generated by this process appears to be between 10^{-4} and 10^{-3} (24, 28, 122). In contrast to variation generated by recombina-

tion, a correlation is expected between the accumulated length differences and the number of generating events, provided the number of these events is not too large (117). Although the mechanisms that generate length variation in VNTRs are not well understood, the distribution of length variation in microsatellite sequences is most consistent with replication slippage (89, 114, 117). Recombination and related gene conversion processes appear to be responsible for variation in minisatellite sequences (47, 48).

Genetic Models

There is an obvious temptation to apply estimators based on familiar population genetic models developed for the F_{ST}-allozyme approach to other forms of gene flow and other genetic marker systems. In some cases, the fit would not be bad, and the models would be expected to yield reasonable gene flow estimates. In other cases, the assumptions of these models would be severely violated and could result in very inaccurate estimates of gene flow. However, even where the standard models are not inappropriate, new models may provide more effective methods for the estimation of gene flow.

COALESCENT MODELS   The introduction of nucleotide sequence data to population genetics has been accompanied by the development of genealogical, or coalescent, models (43, 53, 54, 115), which have led to new approaches for both testing evolutionary hypotheses (49) and estimating population genetic parameters (35, 37). Two kinds of information in a genealogy of DNA sequences can be used for estimates of gene flow. First, the lengths of the branches that connect sequences in the genealogical tree, which correspond to coalescence times, are sometimes referred to as the distances between the sequences. Second, the relative order of branches in the tree, which corresponds to the order of coalescence events, constitute the cladistic relationships of the sequences. Both types of information can be used to estimate gene flow, and an ideal method would make the fullest use of both.

The results of standard population genetic models of gene flow can be recast in genealogical terms because of the close relationship between coalescence time and allelic identity (98). However, because these models consider only average pairwise relationships between sequences, they do not fulfill the potential of genealogical models. Progress in the development of genealogical models that incorporate gene flow (49, 111, 113) is expected to lead to the further development of methods for estimating gene flow and other population genetic parameters from DNA sequence data (35, 45).

RAPIDLY MUTATING LOCI   If the mutation rate at a locus is extremely high, individual alleles generated by mutation will appear and be dispersed only briefly in a population before they are altered by subsequent mutations. Under these

conditions, the spatial distributions of alleles in a population can be interpreted as short-term traces of gene flow. O'Connell & Slatkin (78) used both analytic and simulation models to determine expected spatial distributions of alleles at rapidly mutating loci. The results of their analyses cannot be simply summarized, but they suggest that if the mutation rate is known, it should be possible to estimate neighborhood size from the spatial distributions of alleles at rapidly mutating loci. These models could be applied to microsatellite or minisatellite loci, although the predictions of these models are sensitive to details of the mutation process (78).

ORGANELLAR GENOMES   Specific models have been developed for the population genetics of organellar genomes (mtDNA and cpDNA) (29a, 81a, 112). These models indicate that rates of both gene flow and genetic drift should depend strongly on the mode of transmission of organellar genomes (maternal, paternal or biparental), the effective population sizes of haploid and uniparentally transmitted genomes, and the consequences of multiple modes of dispersal (e.g. pollen vs seed). A common prediction of these models is that organellar genomes should often be subject to greater genetic drift than nuclear genomes, because both haploidy and uniparental transmission may reduce effective population size. A relative reduction in gene flow is also expected for maternally inherited organellar genomes in plants, because both seed and pollen movements result in gene flow for nuclear genes, while only seed dispersal results in gene flow for maternally inherited genes.

Parameter Estimation

CLADISTIC MEASURE OF GENE FLOW   In comparison with allozyme studies, most surveys of DNA sequence variation sample fewer individuals and loci but may provide information that can be used to infer genealogical relationships among sequences. In these respects, DNA sequence data are not well suited for analysis by F_{ST}, which is best applied to large samples and does not make use of cladistic relationships. The cladistic measure of gene flow was developed as an alternative (103). DNA sequences are sampled from subpopulations, and sequence variation is used to infer a phylogenetic tree by any of several standard methods (for example, 110). Each sequence on the tree is assigned a multistate unordered character that indexes the subpopulation from which it was sampled. A parsimony criterion is then used to determine the minimum number of character state transitions required for the phylogenetic distribution of this character to be consistent with the tree. A robust relationship between s and M was found in simulations of island subpopulations over a range of conditions. A table can be used to convert values of s to estimates of M. From additional simulations, a simple relationship was also found between M and

the geographic distance between samples in stepping stone and lattice models (104).

The performance of the cladistic measure of gene flow was compared to F_{ST}-based estimates in simulations of a finite island demographic model, with sequences subject to recombination and an infinite-sites model of mutation (45). Performance criteria were bias, variance, and the accuracy of values for the median and other percentiles of the estimator distributions. For low migration rates, F_{ST} provided the best estimates. At moderate to high rates of migration, the two methods were differentially affected by recombination rate. Low rates of recombination favored the cladistic method, while F_{ST} based estimates improved with higher rates of recombination. Contrary to initial expectations (103), the cladistic method performed nearly as well as F_{ST} under conditions of moderate to high migration and high recombination. At present, the cladistic measure of gene flow is probably best applied to animal mtDNA sequences. Unlike nDNA, mtDNA is not subject to recombination, and unlike cpDNA, there are usually sufficient polymorphic sites in mtDNA to fully resolve genealogies.

NEIGHBORHOOD SIZE   Two parameters can be related to gene flow in spatial models. One corresponds to single generation dispersal distances, the other is the effective neighborhood size, N_b which is analogous to M in subpopulation models. The relationship between these parameters can be understood by considering a symmetrical distribution of single-generation dispersal movements. With respect to a single spatial coordinate, this distribution includes both positive and negative movements, and it has a mean of zero. A typical dispersal distance can be characterized by the standard deviation of this distribution, σ_D. This quantity may be referred to as the standard dispersal distance (77). Together with the population density, D, σ_D determines N_b. If selfing can occur, N_b is the expected number of individuals within a range of $2\sigma_D$ (130).

Underlying similarities between spatial models and stepping-stone island models suggest that F_{ST} analysis can be used to estimate N_b (102). It is encouraging that similar population genetic results have been reached by analyses of different spatial models (32, 78, 102). However, the validity of the approximations used for continuum models has been questioned (32), and these models may not represent the full range of effects that local density regulation could have on gene flow. Simulation models of gene flow, while not on a par with analytic models (21), do allow the investigation of different mechanisms of density regulation on gene flow in continuous populations (76, 77).

DISPERSAL DISTANCE   Striking patterns are often observed when the genealogical relationships of animal mtDNA sequences are overlaid on their geographic locations. These patterns can be interpreted by a combination of population ge-

netic, systematic, and biogeographic principles, and the synthesis of these principles is referred to as phylogeography (4). We investigated phylogeographic patterns that could develop from isolation by distance alone (76, 77) with extensions of models that had been developed for the phylogenetic relationships of mtDNA lineages within species (6) and between species (75). We first considered a continuum model in which mtDNA lineages were dispersed by a simple random-walk process. If two members of the same lineage coalesce n generations in the past and the distribution of single generation dispersal distances of females has variance σ_F^2, the probability distribution for the distance between them will have variance $\sigma_F^2 = 2n\sigma_F^2$. This suggests that it should be possible to estimate the standard dispersal distance, σ_F^2, from an appropriate sample of pairs of mtDNA sequences with estimates of n based on sequence divergence. An attractive feature of this approach is that it is based on a process that is not dependent on genetic drift and provides an estimate of σ_F^2, the standard dispersal distance, which could be directly compared with mark-and-recapture based estimates of dispersal.

There are two problems in applying the random walk model to pairwise data. The first is obtaining pairs of individuals that represent independent realizations of the random walk process. In branching genealogies, many individuals will share portions of their paths of descent and thus also the paths of dispersal that led to their present locations. Data from such pairs of individuals are not independent. There are several possible solutions to this problem, including simply paring down a sample to eliminate nonindependent pairs.

The second and more difficult problem is obtaining distributions of pairwise distances that are conditional on coalescence times (10, 77). The variances of these spatial distributions would provide a direct estimate of σ_G^2. However, in general, we cannot randomly sample individuals with respect to their spatial locations. Rather, we select the locations from which the samples are taken, which imposes the distribution of these locations on the sample.

We developed a workable solution to these problems by using the variance of the spatial distribution of all members of a lineage, σ_H^2, as an approximation of the variance for independent pairs of individuals, σ_G^2 (77). Because members of the same lineage are not independent, σ_H^2 tends to be smaller than σ_G^2, and thus we should anticipate some bias in estimates based on σ_H^2. The second problem now becomes sampling the spatial distributions of entire lineages, rather than pairs of individuals. Ideally, a uniform spatial distribution of sampling locations would be used, but few mtDNA surveys provide such data. However, samples from nonuniform distributions can be used by weighting them appropriately, as we have described. In simulations, this sampling method provided a distribution of estimates for σ_F with the expected bias of σ_H/σ_G and a standard deviation of $0.14\sigma_F$ (77). An advantage of sampling entire lineages

is that separate estimates of σ_F^2 can be based on individual lineages or lineages grouped by location or age. This raises the interesting possibility of analyzing spatial or temporal heterogeneity in the dispersal process.

A simple random walk model of lineage dispersion assumes that the process is uniform and unconstrained. This is unlikely to be true for natural populations. For most species, local population density regulation and barriers to dispersal, including range limits, would violate these assumptions. Furthermore, over long time scales, range limits and population sizes would be expected to change. Therefore, it is important to consider the robustness of the random walk model and to identify conditions under which it may fail.

Range limits will impose an upper limit on σ_H^2. This range saturation effect can be predicted by comparing the expected spatial distribution of a lineage with the size of the available range. As range saturation occurs, it is expected that the rate at which σ_H^2 increases with time will decline, and the correlation between σ_H^2 and σ_F^2 will weaken (76). In additional simulations, with local population density regulation and cyclical range contractions, the relationship $\sigma_G^2 = 2n\sigma_F^2$ was robust for lineages in which range saturation had not occurred (76).

The predictions of the random walk model were tested with published mtDNA surveys (76, 77). As expected, the best fits were found for species with broad geographic ranges and low vagility. In these cases, strong correlations were observed between lineage age and σ_H^2. For the deer mouse (*Peromyscus maniculatus*), our bias-corrected estimate of σ_F^2 was 225 m (77), which corresponds closely to mark-and-recapture estimates of 230 m for females (11) and 264 m for both sexes (26). Also as expected, there was evidence of range saturation for highly vagile species of birds and marine fishes. For these cases, there was no significant correlation between lineage age and σ_H^2, and the observed values of σ_H^2 were not significantly different from those obtained with geographically randomized data (76).

CONSIDERATIONS OF SCALE AND INTERPRETATION

Population Size

Assumptions about population size are often implicit in gene flow models. Estimates of *M* are really measures of the relative strengths of genetic drift and migration. In small populations, the effects of genetic drift on allele frequencies will be strong relative to other forces. Equilibrium between genetic drift and migration will be quickly established (23) and will be mostly insensitive to relatively weaker forces such as selection or mutation (102). However, in large populations, genetic drift is a relatively weak force. If the migration rate is also small, equilibria that are determined primarily by genetic drift will be reached slowly (23), and the distributions of genetic markers will be more susceptible

to mutation and selection. At very low migration rates, the expected time in generations to equilibrium is on the order of the effective population size. If this approaches the ages of populations, historical relationships among populations, rather than gene flow, may be the primary determinant of the distribution of genetic markers (33).

In large populations, selection acting on genetic markers may bias estimates of gene flow. A persistent question about allozymes is to what extent their polymorphisms are subject to natural selection. Some allozyme data is at least consistent with the effects of selection (18, 56, 57a, 83, 121). Balancing selection may be of particular relevance to gene flow estimation. It has long been suggested that either heterosis or frequency dependent selection maintains some allozyme polymorphisms (44, 55, 56). Estimates of F_{ST} based on allozymes subject to balancing selection would lead to overestimates of gene flow (102).

A thought-provoking comparison of allozyme, mtDNA, and nDNA markers has been provided by studies of the American oyster (*Crassostrea virginica*). Its life history is typical of many benthic marine invertebrates, with planktonic larvae that are capable of long-distance dispersal. Buroker (17) surveyed allozymes in American oyster populations from the Atlantic coast of North America and the Gulf of Mexico. Allele frequencies at five polymorphic loci were generally similar among all locales, indicating a high rate of gene flow. In contrast, Reeb & Avise (84) found Atlantic and Gulf populations were distinguished by two very divergent groups of mtDNA. The distributions of these haplotypes implies that Atlantic and Gulf of Mexico populations have long been completely isolated with respect to gene flow. Several alternative explanations can generally be proposed to account for differences in gene flow estimates from allozymes and mtDNA. By conventional wisdom, an unlikely explanation would be that selection has acted on all five allozyme loci to maintain similar gene frequencies in isolated populations. However, based on subsequent evidence, this appears to be the most probable explanation. A survey of four randomly cloned nDNA markers revealed a consistent geographic pattern similar to the distribution of mtDNA haplotypes (50). These nuclear sequences should be subject to the same forces of gene flow and genetic drift as allozyme loci, but, because they are not protein-coding loci, they are not likely subject to the same selective forces. It appears that in the two large regional populations of oysters, balancing selection at allozyme loci may have overcome the relatively weak force of genetic drift and thereby created the appearance of high gene flow.

Mutation Rate

The relationship between F_{ST} and M predicted by standard population genetic models is based on the assumption that mutation rates are much lower than migration rates. If mutation rates are higher than assumed, estimates of F_{ST}

will be downwardly biased. The exact form and magnitude of this bias depends on the mutation process. For an infinite alleles model of mutation and an island model of gene flow: $F_{ST} \simeq 1/(1 + 4N_e m + 4N\mu)$ (23). Thus, mutation has the same effect on F_{ST} as gene flow. For a stepwise mutation model, which may be appropriate for microsatellite loci (89, 114, 117), the bias is greater (100). This dependence of estimates of F_{ST} on mutation rates should not be important for allozyme markers, which have very low mutation rates. However, for DNA sequences with much higher mutation rates, this problem could be quite severe.

If DNA sequence data are used, the problem of mutation saturation can be avoided by using any of several measures that effectively weight pairwise comparisons of sequences by the number of mutations that differentiate them (64, 73, 112). More generalized methods for the use of weighted pairwise data in analysis of population structure have also been developed (30, 82). However, for mtDNA data, a cladistic measure of gene flow (103, 104) may be preferable to an F_{ST}-based estimate (45).

It is less obvious how to correct estimates of F_{ST} for the high mutation rates of microsatellite and minisatellite loci. A weighted pairwise distance measure (38) and a measure of population structure that is analogous to F_{ST} (100) have been proposed for microsatellites that undergo a stepwise mutation process. These measures perform well in simulations, but further testing of the stepwise mutation model is needed before their reliability can be assumed. For minisatellites, an infinite alleles model may provide a reasonable approximation of random unequal crossover events and a large number of potential length states. However, estimators that use corrections based solely on multiple event probabilities would be subject to the high variance of these corrections, and the loss of information that occurs when only a small proportion of alleles remain identical in state.

Time

Barton & Wilson (10) raised a general criticism of any attempt to estimate gene flow from genealogical data, such as mtDNA genealogies. With coalescent models of isolation by distance and neighborhood sizes typical of most species, average coalescence times were sufficiently long for the dispersal of lineages to reach range saturation and thus randomize most information about gene flow. Because in many species mtDNA genealogies are not geographically randomized, they concluded that this long-term expectation of isolation by distance is never realized and that mtDNA distributions are generated primarily by historical processes.

This is a serious criticism not only of gene flow estimates from genealogical data but of any method of gene flow estimation based on a process that may be strongly influenced by population history (33). The problem is not so

much the models as the scale on which they are applied. For example, the time in generations, t, required for F_{ST} to approach equilibrium is approximately $1/[2m + 1/(2N_e)]$, where m is the rate of migration and N_e the effective population size (23). While it is often stated that F_{ST} approaches equilibrium quickly, clearly this is not the case for large N_e and small m. Does this invalidate estimates of gene flow for broad relatively isolated regional populations, which are likely to have been influenced by historical processes as well as gene flow?

The answer to this may be in part a matter of perspective. Large-scale patterns of population structure, whether described by allele frequencies or DNA sequence genealogies have almost certainly been influenced by history. However, throughout the history of a population, the mechanisms behind broad transformations of its structure must necessarily include finer-scale processes of dispersal, migration, and genetic drift, albeit in a more dynamic context than is generally represented in models. The challenge is to develop tools that can resolve these processes and not to assume that simple measurements necessarily require simplistic interpretations.

Gene flow estimates are usually made without reference to a specific timeframe; however, it is possible to adapt them to detect historical changes in patterns of gene flow. For stepping-stone models of population structure, the relationship between estimates of M and the distance between samples can provide evidence for a history that has prevented the attainment of equilibrium (99). Similar analyses can be based on cladistic measures of gene flow (104). Comparisons of population structure measures for nuclear vs. organellar genomes (67, 81a), or for rapidly and slowly mutating loci (100), also have the potential to reveal whether gene flow and genetic drift have reached equilibrium. Estimates of dispersal distance based on the geographical distribution of mtDNA lineages have been viewed as particularly sensitive to historical effects (10). However, this sensitivity could be used to examine the history of dispersal by comparing gene flow estimates among lineages of different ages (76, 77). Finally, it may be possible to identify the processes that generate complex phylogeographic patterns (4) by testing the fit of identifiable elements of these patterns to alternative models that represent both historical and dispersal processes (2, 116). Progress in these areas should benefit from the availability of genetic markers that represent different components of gene flow, from models that are built from different assumptions, and from fresh perspectives on gene flow as a complex process operating on many spatial and temporal scales.

Literature Cited

1. Andrewartha HG, Birch LC. 1954. *The Distribution and Abundance of Animals.* Chicago: Univ. Chicago Press

2. Avise JC. 1994. *Molecular Markers, Natural History and Evolution.* New York: Chapman & Hall

3. Avise JC. 1986. Mitochondrial DNA and the evolutionary genetics of higher animals. *Philos. Trans. R. Soc. London* B312:325–42

4. Avise JC, Arnold J, Ball RM, Bermingham E, Lamb T, et al. 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annu. Rev. Ecol. Syst.* 18:489–522

5. Avise JC, Lansman RA. 1983. Polymorphism of mitchondrial DNA in populations of higher animals. In *Evolution of Genes and Proteins,* ed. M Nei, RK Koehn, pp. 147–64. Sunderland, MA: Sinauer

6. Avise JC, Neigel JE, Arnold J. 1984. Demographic influences on mitochondrial DNA lineage survivorship in animal populations. *J. Mol. Evol.* 20:99–105

7. Baker RR. 1978. *The Evolutionary Ecology of Animal Migration.* New York: Holmes & Meier

8. Barton NH, Halliday RB, Hewitt GM. 1983. Rare electrophoretic variants in a hybrid zone. *Heredity* 50:139–46

9. Barton NH, Slatkin M. 1986. A quasi-equilibrium theory of the distribution of rare alleles in a subdivided population. *Heredity* 56:409–15

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

11. Blair WF. 1940. A study of prairie deer-mouse populations in southern Michigan. *Am. Midland Nat.* 24:273–305

12. Brown WM. 1983. Evolution of animal mitochondrial DNA. In *Evolution of Genes and Proteins,* ed. M Nei, RK Koehn, pp. 62–88. Sunderland, MA: Sinauer

13. Brown WM. 1985. The mitochondrial genome of animals. In *Molecular Evolutionary Genetics,* ed. RJ MacIntyre. New York: Plenum

14. Brown WM, George JM, Wilson AC. 1979. Rapid evolution of animal mitochondrial DNA. *Proc. Natl. Acad. Sci. USA* 76:1967–71

15. Brown WM, Goodman HM. 1979. Quantification of intrapopulation variation by restriction analysis of human mitochondrial DNA. In *Extrachromosomal DNA,* ed. DJ Cummings, P Borst, IB Dawid, SM Weismann, CF Fox. New York: Academic

16. Bruford MW, Hanotte O, Brookfield JFY, Burke T. 1992. Multi- and single-locus DNA fingerprinting. In *Molecular Analysis of Populations: A Practical Approach,* ed. AR Hoelzel, pp. 225–69. Oxford: IRL Press

17. Buroker NE. 1983. Population genetics of the American oyster *Crassostrea virginica* along the Atlantic coast and the Gulf of Mexico. *Mar. Biol.* 75:99–112

18. Burton RS, Feldman MW. 1983. Physiological effects of an allozyme polymorphism: glutamate-pyruvate transaminase and response to hyperosmotic stress in the copepod *Tigriopus californicus. Biochem. Genet.* 21:239–51

19. Byrne M, Moran GF. 1994. Population divergence in the chloroplast genome of *Eucalyptus nitens. Heredity* 73:18–28

20. Clegg MT, Gaut BS, Learn GH, Morton BR. 1994. Rates and patterns of chloroplast DNA evolution. *Proc. Natl. Acad. Sci. USA* 91:6795–801

21. Cockerham CC, Weir BS. 1993. Estimation of gene flow from F-Statistics. *Evolution* 47:855

22. Corriveau JL, Coleman AW. 1988. Rapid screening method to detect potential biparental inheritance of plastid DNA and results for over 200 angiosperm species. *Am. J. Bot.* 75:1443–58

23. Crow JF, Aoki K. 1984. Group selection for a polygenic behavioral trait:

estimating the degree of population subdivision. *Proc. Natl. Acad. Sci. USA* 81:6073–77

24. Dallas JF. 1992. Estimation of microsatellite mutation rates in recombinant inbred strains of mouse. *Mammal. Genome* 5:32–38

25. Di Rienzo A, Peterson AC, Garza JC, Valdes AM, Slatkin M, et al. 1994. Mutational processes of simple-sequence repeat loci in human populations. *Proc. Natl. Acad. Sci. USA* 91:3166–70

26. Dice LR, Howard WE. 1951. Distances of dispersal by prairie deer mice from birthplaces to breeding sites. *Cont. Lab. Vert. Biol. Univ. Mich.* 50:1–15

27. Dong J, Wagner DB. 1994. Paternally inherited chloroplast polymorphism in *Pinus*: estimation of diversity and population subdivision, and tests of disequilibrium with a maternally inherited mitochondrial polymorphism. *Genetics* 136:1187–94

28. Edwards A, Hammond HA, Jin L, Caskey CT, Chakraborty R. 1992. Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. *Genomics* 12:241–53

29. Endler JA. 1977. *Geographic Variation, Speciation and Clines.* Princeton, NJ: Princeton Univ. Press

29a. Ennos RA. 1994. Estimating the relative rates of pollen and seed migration among plant populations. *Heredity* 72:250–59

30. Excoffier L, Smouse PE, Quattro JM. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131:479–91

31. Felsenstein J. 1975. A pain in the torus: some difficulties with models of isolation by distance. *Am. Nat.* 109:359–68

32. Felsenstein J. 1976. The theoretical population genetics of variable selection and migration. *Annu. Rev. Genet.* 10:253–80

33. Felsenstein J. 1982. How can we infer geography and history from gene frequencies? *J. Theor. Biol.* 96:9–20

34. Felsenstein J. 1992. Estimating effective population size from samples of sequences: a bootstrap monte carlo integration method. *Gen. Res.* 60:209–20

35. Felsenstein J. 1992. Estimating effective population size from samples of sequences: inefficiency of pairwise and segregating sites as compared to phylogenetic estimates. *Genet. Res. Camb.* 59:139–47

36. Fenster CB, Ritland K. 1992. Chloroplast DNA an isozyme diversity in two *Mimulus* species (Scrophulariaceae) with contrasting mating systems. *Am. J. Bot.* 79:1440–47

37. Fu YX. 1994. Estimating effective population size or mutation rate using the frequencies of mutations of various classes in a sample of DNA sequences. *Genetics* 138:1375–86

38. Goldstein DB, Ruiz Linares A, Cavalli-Sforza LL, Feldman MW. 1995. An evaluation of genetic distances for use with microsatellite loci. *Genetics* 139:463–71

39. Gyllensten UB, Erlich HA. 1988. Generation of single-stranded DNA by the polymerase chain reaction and its application to direct sequencing of the HLA-DQA locus. *Proc. Natl. Acad. Sci. USA* 85:7652–56

40. Hanski I, Gilpin M. 1991. Metapopulation dynamics—brief history and conceptual domain. *Biol. J. Linn. Soc.* 42:3–16

41. Harris H, Hopkinson DA. 1976. *Handbook of Enzyme Electrophoresis in Human Genetics.* Oxford, UK: North Holland

42. Hong Y, Hipkins VV, Strauss SH. 1993. Chloroplast DNA diversity among trees, populations and species in the California closed-cone pines (*Pinus radiata, Pinus muricata* and *Pinus attenuata*). *Genetics* 135:1187–96

43. Hudson RR. 1990. Gene genealogies and the coalescent process. In *Oxford Surveys in Evolutionary Biology*, ed. DJ Futuyma, J Antonovics

44. Hudson RR, Kreitman M, Aguade M. 1987. A test of neutral molecular evolution based on nucleotide data. *Genetics* 116:153–59

45. Hudson RR, Slatkin M, Madison WP. 1992. Estimation of levels of gene flow from DNA sequence data. *Genetics* 132:583–89

46. Hunkapiller T, Kaiser RJ, Koop BF, Hood L. 1991. Large-scale and automated DNA sequence determination. *Science* 254:59–68

47. Jeffreys AJ, Royle NJ, Wilson V, Wong Z. 1988. Spontaneous mutation rates to new length alleles at tandem-repetitive hypervariable loci in human DNA. *Nature* 332:278–81

48. Jeffreys AJ, Tamaki K, Macleod A, Monckton DG, Neil DL, et al. 1994. Complex gene conversion events in germline mutations at human minisatellites. *Nature Genet.* 6:136–45

48a. Jeffreys AJ, Wilson V, Thein SL. 1985. Hypervariable "minisatellite" regions in human DNA. *Nature* 314:67–73

49. Kaplan N, Hudson RR, Izuka M. 1991. The coalescent process in models with selection, recombination and geographic subdivision. *Genet. Res. Camb.* 57: 83–91

50. Karl SA, Avise JC. 1992. Balancing selection at allozyme loci in oysters: implications from nuclear RFLPs. *Science* 256:100–2

51. Kim K-J, Jansen RK, Turner BL. 1992. Evolutionary implications of intraspecific chloroplast DNA variation in dwarf dandelions (Krigia; Asteraceae). *Am. J. Bot.* 79:708–15

52. Kimura M. 1953. "Stepping-stone" model of population. *Annu. Rep. Natl. Inst. Genet. Japan* 3:62–63

53. Kingman JFC. 1982. The coalescent. *Stochast. Proc. Theor. Appl.* 13:235–48

54. Kingman JFC. 1982. On the genealogy of large populations. *J. Appl. Prob.* 19A:27–43

55. Koehn RK, Gaffney PM. 1984. Genetic heterozygosity and growth rate in *Mytilus edulis. Mar. Biol.* 82:1–7

56. Kohen RR, Hilbish TJ. 1987. The adaptive importance of genetic variation. *Am. Sci.* 75:134–41

57. Kreitman M. 1983. Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster. Nature* 304:412–17

57a. Kreitman M, Akashi H. 1995. Molecular evidence for natural selection. *Annu. Rev. Ecol. Syst.* 26:403–22

58. Kreitman ME, Aguade M. 1986. Excess polymorphism at the *Adh* locus in *Drosophila melanogaster. Genetics* 114:93–110

59. Lansman RA, Shade RO, Shapira JF, Avise JC. 1981. The use of restriction endonucleases to measure mitochondrial DNA sequence relatedness in natural populations. III. Techniques and potential applications. *J. Mol. Evol.* 17:214–26

60. Latter BDH. 1973. The island model of population differentiation: a general solution. *Genetics* 73:147–57

61. Lewontin RC. 1974. *The Genetic Basis of Evolutionary Change.* New York: Columbia Univ. Press

62. Lewontin RC. 1985. Population genetics. In *Evolution: Essays in Honour of John Maynard Smith,* ed. PJ Greenwood, PH Harvey, M Slatkin, pp. 3–18. Cambridge: Cambridge Univ. Press

63. Lewontin RC. 1991. Twenty-five years ago in Genetics. Electrophoresis in the development of evolutionary genetics: milestone or millstone? *Genetics* 128:657–62

64. Lynch M, Crease TJ. 1990. The analysis of population survey data on DNA sequence variation. *Mol. Biol. Evol.* 7:377–94

65. Mason-Gamer RJ, Holsinger KE, Jansen RK. 1995. Chloroplast DNA haplotype variation within and among populations of *Coreopsis grandiflora* (Asteraceae). *Mol. Biol. Evol.* 12:371–81

66. McCauley DE. 1994. Contrasting the distribution of chloroplast DNA an allozyme polymorphism among local populations of *Silene alba*: implications for studies of gene flow in plants. *Proc. Natl. Acad. Sci. USA* 91:8127–31

67. McCauley DE. 1995. The use of chloroplast DNA polymorphism in studies of gene flow in plants. *Trends Ecol. Evol.* 10:198–202

68. Mitton JB. 1994. Molecular approaches to population biology. *Annu. Rev. Ecol. Syst.* 25:45–69

69. Murray V. 1989. Improved double-stranded DNA sequencing using the linear polymerase chain reaction. *Nucleic Acids Res.* 17:8889

70. Nagylaki T. 1978. The geographical structure of populations. In *Studies in Mathematics,* ed. SA Levin, pp. 588–624

71. Neale DB, Saghai-Maroof MA, Allard RW, Zhang Q, Jorgensen RA. 1986. Chloroplast DNA diversity in populations of wild and cultivated barley. *Genetics* 120:1105–10

72. Nei M. 1973. Analysis of gene diversity in subdivided populations. *Proc. Natl. Acad. Sci. USA* 70:3321–23

73. Nei M. 1982. Evolution of human races at the gene level. In *Human Genetics, Part A: The Unfolding Genome,* ed. B Bohhe-Tamir, P Cohen, RN Goodman, pp. 167–181. New York: Alan R. Liss

74. Nei M, Chesser RK. 1983. Estimation of fixation indices and gene diversities. *Ann. Hum. Genet.* 47:253–59

75. Neigel JE, Avise JC. 1986. Phylogenetic relationships of mitochondrial DNA under various demographic models of speciation. In *Evolutionary Processes and Theory,* ed. S Karlin, E Nevo. New York: Academic

76. Neigel JE, Avise JC. 1993. Application of a random walk model to geographic distributions of animal mitochondrial DNA variation. *Genetics* 135:1209–20

77. Neigel JE, Ball RM, Avise JC. 1991. Estimation of single generation migration distances from geographic variation in animal mitochondrial DNA. *Evolution* 45:423–32

78. O'Connell N, Slatkin M. 1993. High mutation rate loci in a subdivided population. *Theor. Pop. Biol.* 44:110–27

79. Okubo A. 1980. *Diffusion and Ecological Problems: Mathematical Models.* Berlin: Springer-Verlag

80. Palmer JD. 1991. Plastid chromosomes: structure and evolution. In *The Molecular Biology of Plastids,* ed. IK Vasil. 7A:5–53. New York: Academic

81. Petit RJ, Kremer A, Wagner DB. 1993. Geographic structure of chloroplast DNA polymorphisms in European oaks. *Theor. Appl. Genet.* 87:122–28

81a. Petit RJ, Kremer A, Wagner DB. 1993. Finite island model for organelle and nuclear genes in plants. *Heredity* 71:630–41

82. Pons O, Petit RJ. 1996. Measuring and testing genetic differentiation with ordered versus unordered alleles. *Genetics* 144:1237

83. Powers DA, DiMichele L, Place AR. 1983. The use of enzyme kinetics to predict differences in cellular metabolism, developmental rate and swimming performance between LDH-B genotypes of the fish *Fundulus heteroclitus. Curr. Top. Biol. Med. Res.* 10:147–70

84. Reeb CA, Avise JC. 1990. A genetic discontinuity in a continuously distributed species: mitochondrial DNA in the American Oyster, *Crassostrea virginica. Genetics* 124:397–406

85. Renshaw E. 1991. *Modelling Biological Populations in Space and Time.* Cambridge: Cambridge Univ. Press

86. Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, et al. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239:487–91

87. Sawyer S. 1967. Branching diffusion processes in population genetics. *Adv. Appl. Prob.* 8:659–89

88. Schaeffer SW, Aquadro CF, Anderson WW. 1987. Restriction-map variation in the alcohol dehydrogenase region in *Drosophila pseudoobscura. Mol. Biol. Evol.* 4:254–65

89. Schlotter C, Tautz D. 1993. Slippage synthesis of microsatellites. *Nucleic Acids Res.* 20:211–15

90. Sears BB. 1980. The eliminatino of plastids during spermatogenesis and fertilization in the plant kingdom. *Plasmid* 4:233–55

91. Shriver MD, Jin L, Chakraborty R, Boerwinkle E. 1993. VNTR allele frequency distributions under the stepwise mutation model: a computer simulation approach. *Genetics* 134:983–93

92. Skellam JG. 1951. Random dispersal in theoretical populations. *Biometrika* 38:196–218

93. Slade RW, Moritz C, Heideman A, Hale PT. 1993. Rapid assessment of single copy nuclear DNA variation in diverse species. *Mol. Ecol.* 2:359–73

94. Slatkin M. 1977. Gene flow and genetic drift in a species subject to frequent local extinction. *Theor. Popul. Biol.* 12:253–62

95. Slatkin M. 1985. Gene flow in natural populations. *Annu. Rev. Ecol. Syst.* 16:393–430

96. Slatkin M. 1985. Rare alleles as indicators of gene flow. *Evolution* 39:53–65

97. Slatkin M. 1987. Gene flow and the geographic structure of natural populations. *Science* 236:787–92

98. Slatkin M. 1991. Inbreeding coefficients and coalescence times. *Genet. Res. Camb.* 58:167–75

99. Slatkin M. 1993. Isolation by distance in equilibrium and non-equilibrium populations. *Evolution* 47:264–79

100. Slatkin M. 1995. A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 139:457–62

101. Slatkin M, Arter H. 1991. Spatial autocorrelation methods in population genetics. *Am. Nat.* 138:499–517

102. Slatkin M, Barton NH. 1989. A comparison of three indirect methods for estimating average levels of gene flow. *Evolution* 43:1349–68

103. Slatkin M, Maddison WP. 1989. A cladistic measure of gene flow inferred from the phylogenies of alleles. *Genetics* 123:603–13

104. Slatkin M, Maddison WP. 1990. Detecting isolation by distance using phylogenies of genes. *Genetics* 126:249–60

105. Sokal RR, Jacquez GM, Wooten MC. 1989. Spatial autocorrelation analysis of migration and selection. *Genetics* 121:845–55

106. Sokal RR, Oden NL. 1978. Spatial autocorrelation in biology. II. Some biological implications and four applications of evolutionary and ecological interest. *Biol. J. Linn. Soc.* 10:229–49

107. Sokal RR, Wartenberg DE. 1983. A test of spatial autocorrelation analysis using

an isolation-by-distance model. *Genetics* 105:219–37

108. Soltis DE, Mayer MS, Soltis PS, Edgerton M. 1991. Chloroplast-DNA variation in *Tellima grandiflora. Am. J. Bot.* 78:1379–90

109. Soltis DE, Soltis PS, Ranker TA, Ness BD. 1989. Chloroplast DNA variation in a wild plant, *Tolmiea menziesii. Genetics* 121:819–26

110. Swofford DL, Olsen GJ, Waddell PJ, Hillis DM. 1996. Phylogenetic inference. In *Molecular Systematics,* ed. DM Hillis, C Moritz, BK Mable, pp. 407–514. Sunderland, MA: Sinauer

111. Takahata N. 1988. The coalescent in two partially isolated diffusion populations. *Genet. Res.* 52:213–22

112. Takahata N, Palumbi SR. 1985. Extranuclear differentiation and gene flow in the finite island model. *Genetics* 109:441–57

113. Takahata N, Slatkin M. 1990. Genealogy of neutral genes in two partially isolated populations. *Theor. Popul. Biol.* 38:331–50

113a. Tautz D. 1989. Hypervariability of simple sequences as a general source for polymorphic markers. *Nucleic Acids Res.* 17:6463–571

114. Tautz D, Schlotterer C. 1994. Simple sequences. *Curr. Opin. Genet. Dev.* 4:832–37

115. Tavarae S. 1984. Line-of-descent and genealogical processes, and their applications in population genetic models. *Theor. Popul. Biol.* 26:119–64

116. Templeton AR, Routman E, Phillips CA. 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* 140:767–82

117. Valdes AM, Slatkin M, Friemer NB. 1993. Allele frequencies at microsatellite loci: the stepwise mutation model revisited. *Genetics* 133:737–49

118. Vawter L, Brown WM. 1986. Nuclear and mitochondrial DNA comparisons reveal extreme rate variation in the molecular clock. *Science* 234:194–96

119. Voelker RA, Schaffer HE, Mukai T.
1980. Spontaneous allozyme mutations in *Drosphila melanogaster*: rate of occurrence and nature of the mutants. *Genetics* 94:961–68

120. Wade MJ, McCauley DE. 1988. Extinction and recolonization: their effects on the genetic differentiation of local populations. *Evolution* 42:995–1005

121. Watt WB. 1983. Adaptation at specific loci. II. Demographic and biochemical elements in the maintenance of the Colias PGI polymorphism. *Genetics* 103:691–724

122. Weber JL, Wong C. 1993. Mutation of human short tandem repeats. *Hum. Mol. Genet.* 2:1123–28

123. Wehrhahn CF, Powell R. 1987. Electrophoretic variation, regional differences, and gene flow in the coho salmon (*Onchorhynchus kisutch*) of southern British Columbia. *Can. J. Fish. Aqaut. Sci.* 44:822–31

124. Weir BS, Cockerham CC. 1984. Estimating F-Statistics for the analysis of population structure. *Evolution* 38:1358–70

125. Whitlock MC, McCauley DE. 1990. Some population genetic consequences of colony formation and extinction: genetic correlations within founding groups. *Evolution* 44:1717–24

126. Wright S. 1921. Systems of mating. I. The biometric relations between parent and offspring. *Genetics* 6:111–23

127. Wright S. 1931. Evolution in Mendelian populations. *Genetics* 16:97–159

128. Wright S. 1938. The distribution of gene frequencies under irreversible mutation. *Proc. Natl. Acad. Sci. USA* 245:253–59

129. Wright S. 1940. Breeding structure of populations in relation to speciation. *Am. Nat.* 74:232–48

130. Wright S. 1943. Isolation by distance. *Genetics* 28:114–38

131. Wright S. 1946. Isolation by distance under diverse systems of mating. *Genetics* 28:139–56

132. Wright S. 1951. The genetical structure of populations. *Ann. Eugen.* 15:323–53

133. Wright S. 1965. The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution* 19:395–420