Gene Flow and the Geographic Structure of Natural Populations

MONTGOMERY SLATKIN

There is abundant geographic variation in both morphology and gene frequency in most species. The extent of geographic variation results from a balance of forces tending to produce local genetic differentiation and forces tending to produce genetic homogeneity. Mutation, genetic drift due to finite population size, and natural selection favoring adaptations to local environmental conditions will all lead to the genetic differentiation of local populations, and the movement of gametes, individuals, and even entire populations—collectively called gene flow—will oppose that differentiation. Gene flow may either constrain evolution by preventing adaptation to local conditions or promote evolution by spreading new genes and combinations of genes throughout a species' range. Several methods are available for estimating the amount of gene flow. Direct methods monitor ongoing gene flow, and indirect methods use spatial distributions of gene frequencies to infer past gene flow. Applications of these methods show that species differ widely in the gene flow that they experience. Of particular interest are those species for which direct methods indicate little current gene flow but indirect methods indicate much higher levels of gene flow in the recent past. Such species probably have undergone large-scale demographic changes relatively frequently.

EVOLUTIONARY THEORY ATTEMPTS TO EXPLAIN PAST change in terms of the relatively few mechanisms that can cause genetic evolution. The mechanism central to Darwin's theory was natural selection, although Darwin recognized that accident and interbreeding between populations could oppose natural selection. In modern terminology, "genetic drift" is the unpredictable change in gene frequency due to finite population size, and "gene flow" is the change due to movement of gametes, individuals, or groups of individuals from one place to another.

Gene flow is often regarded as a constraining force in evolution. Natural selection will tend to adapt a population to local environmental conditions but immigrants from other populations will introduce genes adapted to other conditions. In fact, gene flow between populations may prevent them from evolving into different species. But as emphasized by Sewall Wright in particular, gene flow can also be a creative force in evolution. The movement of individuals and even entire populations may spread superior genes and combinations of genes throughout a species once they become common in one location. What role gene flow plays in a particular species depends both on the geographic distribution of that species and on the importance of other evolutionary forces. Population genetics theory now makes clear predictions about how gene flow can influence genetic evolution, and recent studies of natural populations are beginning to estimate how much gene flow occurs in different species.

Geographic Distributions of Species

The overall geographic range of a species is determined largely by a series of historical accidents. A species will extend its range until it is stopped by barriers to dispersal. These barriers are sometimes large and conspicuous and will stop most species that reach them, and they are sometimes so small that it is difficult to see why a particular species has not crossed them. Mountain ranges, deserts, oceans, and other major geographic features together form a network of barriers that separate potentially isolated regions. Which regions a species occupies depends on when it originated and what barriers to its dispersal it encountered. Although the principle is simple, the possibilities are numerous because species are formed at different times, because barriers appear and disappear, and because barriers differ in importance to different types of organisms. Historical biogeography is partly devoted to identifying barriers to dispersal and describing their effects on the geographic distributions of species.

On a smaller scale, where a species is found is determined primarily by ecological factors, including climate, predators, competitors, and usable resources. The resulting population structure may appear essentially continuous over large geographic areas or be very patchy, with areas of high abundance separated by areas in which a species is rarely or never found. The demographic structure may be stable, with local populations persisting in each area continuously for long times, or unstable, with large-scale demographic changes occurring frequently in the evolutionary history of a species. Demographic instability occurs in a variety of ways. Weeds and other "colonizing species" usually comprise numerous local populations that persist for relatively short times, perhaps only a few generations. Such species depend on regularly finding new suitable habitats. Demographic instability can also result from large-scale expansions in geographic range, as could occur during major climatic changes or after crossing former barriers to dispersal. Such range expansions may be rare on the time scale set by human observation, but they may be frequent on the much longer time scale of genetic evolution.

Gene Flow as a Constraining Force

Darwin emphasized that isolation of populations was one factor promoting evolution. He noted how plant and animal breeders would separate individuals with desirable characteristics in order to prevent interbreeding with the parental stock. Extrapolating to natural populations, he accounted for the unusual characteristics of

---

The author is in the Department of Zoology and the Museum of Vertebrate Zoology, University of California, Berkeley, CA 94720.

species on remote islands and in other isolated habitats such as caves in part by their extreme isolation. Darwin did not say that isolation was a necessary first step in the evolution of novel traits. That view came later, being first suggested by Wagner in 1868 and promoted by Karl Jordan at the end of the 19th century (*1*). In more recent evolutionary discussions, Mayr has been the strongest advocate of the idea that gene flow is a strong constraining force to evolutionary change, although his views on this subject have moderated considerably in the past 15 years (*2*). Mayr's views of genetic evolution are based on his "biological" definition of species as a group of actually or potentially interbreeding organisms. Because of the faculty for interbreeding, gene flow between populations of the same species is possible and can prevent local differentiation. If gene flow is interrupted, populations can evolve independently and eventually form distinct species, what Mayr called the "allopatric" mode of speciation.

Mayr's view of the importance of gene flow was challenged in 1969 by Ehrlich and Raven (*3*) on two grounds. First, Ehrlich and Raven argued that gene flow in natural populations is too rare and restricted to possibly be able to bind a species into a single evolutionary unit. Second, they noted that strong selection can produce adaptations to local conditions in the presence of substantial gene flow. The difference between their view and Mayr's is one of degree: how much gene flow is sufficient to prevent genetic differentiation and speciation? Classical population genetics theory points to a few general principles concerning the relative strengths of gene flow and other forces, and new developments suggest that there are useful additions to those principles.

The balance achieved between gene flow and genetic drift provides a background against which to consider the effects of different kinds of selection. Genetic drift, like gene flow, has the same average effect on all nuclear genes. In a group of completely isolated populations, genetic drift alone would tend to fix different alleles in different local populations. Any gene flow at all among populations will prevent complete fixation, but as shown first by Wright (*4*) gene flow must exceed a certain level to prevent substantial genetic differentiation due to genetic drift. Roughly speaking, an average of one individual or more exchanged between two populations will prevent different neutral alleles at the same locus from being nearly fixed in two populations. What is surprising about this result is that it is independent of population size. In larger populations, the force of gene flow as measured by the fraction of individuals that are immigrants (often denoted by m) is smaller but the force of genetic drift which is proportional to the inverse of the population size (N) is weaker, so the two forces remain in the same balance.

Recently theories have shown that frequent extinctions and recolonizations of local populations can also be an important source of gene flow. Even if there is no exchange of individuals between established populations, there will be little differentiation of local populations due to genetic drift if the average time that a population persists in one area is less than the time it takes for genetic drift to fix neutral alleles, which is of the same order of magnitude as the effective population size (*5*). This provides a rule for extinctions and recolonizations that corresponds to Wright's rule for the exchange between permanent local populations.

Natural selection can be much more effective than genetic drift in either preventing or establishing local differences. Selection in favor of the same alleles or the same traits would produce geographic uniformity regardless of any gene flow. Selection favoring different alleles in different locations will succeed in producing local differences reflecting genetic adaptations to local conditions if, roughly speaking, the fitness differences, measured by s, exceed the fraction of immigrants, m (*6*).

Although genetic drift affects all loci in the same way, natural selection does not. Natural selection for locally important adaptations could cause substantial differences at a few loci, with other loci that are neutral or only weakly selected being relatively uniform throughout a species' range. In terms of the relative strengths of evolutionary forces, gene flow might be weaker than selection at some loci yet much stronger than genetic drift at other loci.

These general principles follow from the classical population genetics approach of examining changes in allele frequencies at a single genetic locus. Recent theory has concentrated on the interactions among these forces, especially as they are complicated by genetic linkage. For example, if selection affecting different loci follows the same geographic pattern, as might be expected in a species experiencing different ecological conditions in different areas, then selection on those loci is reinforced by linkage (*7*). In addition, the effective level of gene flow at linked neutral loci is reduced, implying that geographic variation in selection acts as a partial barrier to gene flow (*8*).

Gene Flow as a Creative Force

Gene flow can inhibit genetic evolution by preventing natural selection and genetic drift from establishing and maintaining local genetic differences. Gene flow can also promote genetic evolution, although this theory is less well developed. In the early 1930s, Wright (*9*) introduced his "shifting balance theory" in which gene flow and population subdivision played a central role. Wright had developed the mathematical theory of genetic drift and had recognized that drift could cause genetic evolution that would not occur under the influence of natural selection alone. To visualize the problem, Wright introduced the powerful metaphor of the "adaptive landscape," in which a population was represented as a point on a surface with the axes being measurements of phenotypic characters and the height being the mean fitness of a population with that combination of characters. Under natural selection alone, a population would move "uphill" on the adaptive surface and stop when it reached a local maximum of mean fitness, an "adaptive peak" (*10*). Wright argued that, because of the pleiotropy of most genes and the epistatic interactions among genes, the adaptive surface for most species would have numerous peaks separated by adaptive valleys. Under natural selection alone, a species would be trapped on one adaptive peak even if there were higher adaptive peaks representing better adapted combinations of characters.

According to the shifting balance theory, many species comprise small, partially isolated populations. Genetic drift in one of these populations could fix genes or combinations of genes that would carry that population to a higher adaptive peak. Then gene flow would spread those genes to other populations. Wright emphasized that the spread could be due both to gene flow between established populations and to the colonization of new populations.

Although Wright discussed the shifting balance theory extensively, he did not delimit the conditions under which it would work. The kinds of models needed are particularly difficult because they must account for several forces acting simultaneously. For relatively simple models, population subdivision will promote genetic evolution under some conditions, as Wright predicted (*11*). Local population sizes must be sufficiently small and immigration sufficiently uncommon that genetic drift can overcome selection and fix single genes or combinations of genes that would otherwise remain rare. For a demographically stable species, that low level of gene flow is probably too weak to spread those new genes or combinations to other populations in a reasonable time. In a demographically unstable species, however, the movement of entire populations could easily promote rapid genetic evolution (*12*).

Group Selection

If there are genetic differences among local populations, any differences that either decrease the likelihood of local extinction or increase the likelihood of a population producing emigrants or colonists will affect the genetic composition of a species. Wright (*13*) called selection due to differences among local populations "interdemic selection" to distinguish it from selection acting within each local population or "deme." Interdemic selection was part of the shifting balance theory because it would be effective in spreading genes fixed initially in one population. More recently, this process has been called "group selection." At one time, group selection was contrasted with "kin selection," which depends on the genetic relatedness of individuals, as an explanation for the evolution of altruistic behavior (*14*). The more current view is that the group and kin selection represent ends of a continuum of processes that depend on genetic variation within and between populations. Selection on a particular allele or trait can be partitioned into the components due to the effects on different hierarchical levels of a species, from the individual to the family, the local population, and assemblages of local populations (*15*). This approach has proved to be especially fruitful for modeling the effects of complex mating systems and population structures for which it is impossible to say precisely what constitutes a local population.

Group selection is another way in which gene flow and population differentiation may play a creative role in evolution. In Wright's shifting balance theory, the movement of individuals between permanent populations is one way that adaptations in one local population may spread to others, but a much more effective way for spreading new adaptations is through the establishment of new local populations. This intuitive argument is supported by recent theoretical analysis. If extinctions and recolonizations are frequent, then group selection can lead to the fixation of genes in a species even if those genes were opposed by natural selection within each local population (*15*).

Group selection is likely to be particularly important for traits affecting dispersal because the evolution of dispersal ability necessarily depends on what happens to individuals that disperse. Dispersal is an important part of the life cycle of many species and is especially so for parasites. Group selection is often invoked in discussing the evolution of parasitic diseases. Viral and bacterial diseases sometimes become less virulent, and this is often attributed to group selection: genotypes of the parasite that prolong the survival of the host will be favored by group selection because greater host longevity will provide greater opportunities for dispersal of those genotypes. This view is widespread among parasitologists, but there is currently no evidence that group selection due to differential survival of hosts is in fact responsible for the loss of virulence of parasitic diseases (*16*). That is a plausible explanation, but other explanations are equally plausible. Because of the very short generation time of parasites, even an extremely virulent disease agent that kills a host within a few days may still have ample opportunity for evolution within each host. The virulence to a host is intimately tied to the response of individual parasites to the host's immune system, so selection within each host could easily result in a reduction in virulence. Evolutionary models of host-parasite coevolution are beginning to take realistic account of the population biology of both the parasites and hosts (*17*). The population structure of the parasites imposed by the usually temporary association with hosts will necessarily be central to realistic theories of host-parasite coevolution. This problem is not only of evolutionary interest. The treatment and control of parasitic diseases will have to be based on a correct understanding of the forces governing host-parasite coevolution.

Speciation

The formation of new species from existing ones is an elusive process. By definition, species are "reproductively isolated," meaning that they do not freely interbreed under natural conditions (*2*). Therefore, there is no, or essentially no, gene flow between species. Besides being reproductively isolated, species differ in other ways from one another. We can usually distinguish individuals belonging to different species by many differences in behavior and morphology that may have nothing to do with reproductive isolation. There are then three components to species formation: the cessation of gene flow, the evolution of reproductive isolating mechanisms (that is, characteristics that prevent interbreeding), and the accumulation of other morphological and behavioral differences. The question is in what order these events occur. According to Mayr's allopatric theory (*2*), gene flow between existing populations plays a conservative role and a cessation of gene flow due to the appearance of a barrier to dispersal will precede the evolution of reproductive isolation and other differences. Many examples of this process are provided by species on different continents that have been gradually separated by tectonic movements.

But is a complete absence of gene flow a necessary first step? Or, more realistically, how much does gene flow need to be reduced before speciation can proceed? If gene flow is not completely stopped between two populations, then there is an initial disadvantage to individuals that preferentially mate with members of their own population (which is the first step toward reproductive isolation and speciation), because those individuals are restricting their pool of potential mates. For speciation to proceed, that disadvantage has to be offset by some advantage. In most theories of speciation in the presence of gene flow—theories of "sympatric" or "parapatric" speciation—the offsetting advantage is the greater opportunity for adapting to particular local conditions. In such theories, reproductive isolation evolves because natural selection favors mechanisms that reduce and eliminate gene flow, thereby permitting more precise local adaptations.

Theories of sympatric speciation show that the conditions for the evolution of reproductive isolation in the presence of substantial gene flow are rather restrictive. For sympatric speciation to occur readily, there must be a high degree of genetic correlation between traits conferring local adaptations and traits causing reproductive isolation (*18*), those genetic correlations being due either to genetic linkage or pleiotropy. Otherwise, recombination will tend to disassociate alleles producing local adaptations from those causing reproductive isolation.

Gene flow between established populations always plays a constraining role in speciation. However, the founding of new populations, which is another form of gene flow, may lead to speciation. This idea was first proposed by Mayr as another mode of allopatric speciation (*19*), and he has since called it "peripatric speciation" to distinguish it from allopatric speciation that results from the appearance of a barrier to dispersal between existing populations. According to this theory, speciation can occur because, in a newly founded small population, rapid genetic evolution can be caused by the combined effects of genetic drift and strong natural selection under new environmental conditions, what Mayr called "genetic revolutions." The "founder flush" and "genetic transilience" theories are in the same spirit but differ in detail (*20*).

Mayr's theory of peripatric speciation is similar to Wright's shifting balance theory in several ways, although the two theories are couched in very different terms. Mayr, like Wright, emphasized that small populations could evolve rapidly and contain new genetic combinations that might not appear in the rest of the species. And Mayr, like Wright, emphasized that pleiotropy and epistasis could foster rapid evolution. The difference is that Mayr, unlike Wright, emphasized that reproductive isolation could be an additional consequence of this rapid evolution. Some of such nascent species, possibly only a small fraction, could then expand their geographic range while others might go extinct before being noticed. In both the shifting balance theory and the peripatric theory of speciation, the establishment of small populations and the subsequent spread of successful populations allows for evolution that could not occur in a single large population under the influence of natural selection alone.

Peripatric speciation and related theories all assume that the reduction in gene flow occurs first. They differ from the more traditional theories of allopatric speciation primarily in the rate of formation of new species. The divergence of species after the appearance of major geographic barriers, such as the separation of continents by tectonic movements, is thought to be relatively slow, or at least there is no reason to assume it is rapid. Genetic revolutions and similar processes are supposed to occur much more quickly because genetic drift in small populations possibly combined with strong selection due to peculiar local conditions can produce rapid evolution. At present, there are doubts about whether these processes account for the formation of any species (*21*), but if they are important, the demographic instability of a widespread species would facilitate their operation.

Estimating Levels of Gene Flow

Theoretical studies are clear about what will happen under different amounts of gene flow, but how much gene flow is occurring is much less clear. The problem is that gene flow is intrinsically difficult to measure. Individual movements can be observed in some species, particularly birds and mammals, but, in other species, gametes or newly formed zygotes disperse and their movements are difficult to follow. Furthermore, gene flow depends not only on dispersal but also on successful breeding, and that may also be difficult to assess.

Population biologists have two classes of methods to estimate how much gene flow occurs in natural populations. "Direct methods" use estimates of dispersal distances and breeding success of dispersers to infer how much gene flow is occurring at the time the observations are made. "Indirect methods" use allele frequencies, and more recently DNA sequence differences, to estimate the levels of gene flow that must have been occurring in order to produce the observed patterns. Both classes of methods depend on assumptions about the species being studied and each class has its strengths. Ideally, both classes of methods should be used on a particular species because they yield different information.

Direct methods. The mobility of a species is often one of its most conspicuous features and casual observations often suggest how much gene flow a species experiences. Some species of birds have a worldwide distribution with individuals being known to fly hundreds or thousands of miles. Such species can reasonably be supposed to form an almost panmictic unit. Planktonic larvae of many marine species can survive for months in the ocean and disperse passively with currents. Although tracking an individual larva is impossible, the capacity for long-distance dispersal and the wide geographic range of many marine species suggest that gene flow over long distances is common.

Casual observations of dispersal can sometimes be misleading. The capacity for dispersal does not always predict how much gene flow actually occurs. One reason is that dispersal entails considerable risk. For most species, suitable habitats are rare, and passive dispersers, such as wind-dispersed seeds and planktonic larvae, may not find them. Also, suitable habitats may be sufficiently crowded that dispersers cannot establish themselves. Ehrlich and Raven (*3*) emphasized that dispersal of individuals and hence gene flow is over much shorter distances than individuals are capable of moving and that view is widely supported by other empirical studies (*22*).

Direct observations of dispersal are necessarily limited in both space and time. Quantitative studies depend on either recapturing marked individuals or monitoring all dispersers in a restricted area. Individuals that move beyond the study area are missed, as are individuals that disperse by unexpected means. It is possible to use a net to collect seeds that are dispersed by air currents but not those that stick to the feet of rodents.

Another and possibly more serious source of bias is due to the limited time scale of direct observations. The evolutionary importance of gene flow depends on its effects averaged over a large number of generations. The time scale of change associated with a particular evolutionary force is the time needed for that force to cause a substantial change in gene frequency. Roughly speaking, the time scale associated with natural selection is the inverse of the difference in relative fitnesses. For example, a 1% difference in the fitnesses of different genotypes would cause a significant change in gene frequency in approximately 100 generations. The time scale associated with genetic drift is approximately the number of individuals in a local population. The importance of gene flow relative to these other forces is determined by the amount of gene flow averaged over the time scale of change due to the other force. For gene flow to offset a fitness difference of 1%, an average over 100 generations of roughly 1% of a population would have to be replaced by immigrants. That average could be achieved by only three or four episodes of significant gene flow. If direct estimates of gene flow are made for only one or a few generations, those episodes could be missed completely.

Bias in direct estimates of gene flow may be compounded by the tendency to observe dispersal under what appear to be "normal" conditions, with measurements made in relatively undisturbed populations. Under those conditions, there may be little dispersal. For example, the checkerspot butterfly *Euphydryas editha*, which has been studied extensively by Ehrlich and his colleagues for the past 25 years, moves little between populations that are almost adjacent (*23*). That would suggest that there is almost no gene flow even on the smallest geographic scale. Yet when a local population went extinct, a new population was established after only one year, indicating a potentially strong force preventing the genetic differentiation of neighboring populations. However, such observations are uncommon because they are rarely seen in short-term studies and because large-scale demographic changes in a species are more in the province of ecology than population genetics.

Direct measures of dispersal can indicate the gene flow at a particular time, but they do not necessarily reveal the level of gene flow over longer time scales that may encompass a variety of events not occurring during the period of observation. In contrast, indirect estimates of gene flow based on the analysis of gene frequencies necessarily depend on levels of gene flow averaged over long times. The agreement or lack of agreement between these two methods indicates the extent to which rare and unpredictable events, including large-scale dispersal and major changes in population structure, have been important in the recent history of a species.

Indirect methods. Indirect ways to estimate levels of gene flow use allele frequencies, usually as determined by electrophoretic surveys, to estimate levels and patterns of gene flow. Recently, restriction site polymorphisms and DNA sequence data have also been used, but at present there are fewer such data. Spatial distributions of allele frequencies do not themselves reveal how much gene flow is occurring. Models of gene flow and other forces of genetic evolution are used to predict how much gene flow must have occurred in order for the patterns in the data to be observed. An ideal indirect method would be one that detects patterns in allele frequencies that are due only to gene flow and ignores patterns due to natural selection, genetic drift, and mutation. That goal has not been reached, but there are methods that do indicate the balance achieved between gene flow and genetic drift and that are relatively insensitive to assumptions about natural selection and mutation. Furthermore, there are generalizations of these methods to data on restriction site polymorphisms and DNA sequences.

There are currently two indirect methods that can be used to estimate average levels of gene flow among populations. One method is Wright's statistic for estimating the standardized variance in allele frequencies among local populations, F_{ST} (*24*). If there are only two alleles at a locus, $F_{ST} = \sigma^2/p(1 - p)$, where p is the mean and σ^2 is the variance in frequency of either allele. If there are more than two alleles per locus and data for more than one locus, there are a variety of ways of combining information to yield a single estimate of F_{ST} (*25*). The reason for estimating F_{ST} is that Wright (*4*) showed that for neutral alleles, $F_{ST} \approx 1/(1 + 4Nm)$, where N is the local population size and m is the average rate of immigration in an "island" model of population structure. The island model assumes that every local population is equally accessible from every other, and it represents the extreme in dispersal over large distances. By inverting Wright's formula, the value of Nm can be estimated from F_{ST}. One reason for estimating Nm is that this combination of parameters indicates the relative strengths of gene flow and genetic drift. Genetic drift will result in substantial local differentiation if $Nm < 1$ but not if $Nm > 1$.

The other method for estimating Nm is my method, which depends on the frequencies of rare alleles, alleles found in only one or a few local populations (*26*). My method is based on the fact that the average frequency of alleles found in only a single population is a simple function of Nm: $\ln[p(1)] \approx a\ln(Nm) + b$, where $p(1)$ is the average frequency of alleles found in only one population sampled and a and b are constants that depend on the number of individuals sampled per population (*26, 27*).

Although these two methods appear to be very different and to have different properties, Barton and I have recently found them to be different ways of estimating the same essential properties of gene frequency distributions and, in extensive simulation studies, have found estimates using these two methods to be consistent over a wide range of assumptions about population structure, selection, and mutation (*27*).

Both F_{ST} and rare alleles have the same desirable properties as indicators of the level of gene flow. Loci with different mutation rates and loci exposed to different kinds of natural selection will tend to produce similar estimates of Nm. Therefore, information from different loci can be combined without making any restrictive assumptions about those loci. The only important exception is a locus at which natural selection maintains different alleles in high frequencies in different local populations despite any gene flow occurring. That kind of selection would result in a pattern indicating little or no gene flow ($Nm << 1$) even if the actual level of gene flow were much larger. That would be a serious problem for these methods were it not for the fact that data are available for numerous genetic loci, and estimates based on different loci or subsets of loci in the same species are usually consistent. In a few species, allele frequencies at one locus are inconsistent with all the others, which suggests that selection is affecting that locus.

To illustrate the application of one of these methods, Table 1 shows value of F_{ST} and derived estimates of Nm for eight polymorphic loci from 21 local populations of *E. editha* (*28*). This example illustrates several points about these methods. First, estimates of Nm based on data from most loci are consistent and indicate gene flow is sufficiently strong that it prevents genetic drift from causing local genetic differentiation. The average estimate of Nm based on seven of the eight loci (excluding *hk*) is 7.8, and the variation about that mean is consistent with theoretical expectations. Second, one locus, *hk* (hexokinase), differs sufficiently from the others that it is reasonable to conclude that it or a locus closely linked to it is subject to strong natural selection favoring different alleles in different populations.

Table 1. Analysis of data of McKechnie *et al.* (*23*, table 1) on *Euphydryas editha*. The F_{ST} values are the averages for a locus of the values computed for each allele. The estimate of Nm is based on Wright's formula (*25*) for the equilibrium in an island model of population structure under a balance between gene flow and genetic drift: $F_{ST} = 1/(1 + 4Nm)$.

Locus	F_{ST}	Nm (estimate)
pgm	0.028	8.7
pgi	0.052	4.6
hk	0.291	0.6
got	0.017	14.5
ak	0.062	3.8
bdh	0.034	7.1
α-*gpdh*	0.027	9.0
to	0.035	6.9

Third, this example illustrates how direct and indirect estimates of gene flow can differ. Ehrlich and his co-workers have persuasive evidence showing that gene flow during the past 25 years between even nearby populations is rare and has almost certainly not occurred between populations that are more widely separated (*23*). Marked individuals released in one population rarely move to adjacent populations despite their obvious capacity to fly much farther and despite the absence of any barriers preventing movement. They estimated that Nm between nearby populations is approximately 0.1. Moreover, the breeding of these butterflies in a particular area is synchronized and is controlled by the flowering times of their primary host plant. Host plants and times of breeding vary with habitat, so populations as close as 19 km are sufficiently different that an individual moving between them would be unable to mate. Yet the results in Table 1, which are based on data from samples taken throughout central California, show that gene flow must have occurred sufficiently often, averaged over a long time, to prevent local differentiation. If gene flow were not responsible for these patterns, we would have to conclude that natural selection, genetic drift, and mutation had combined in precisely the right way to mimic the effects of gene flow at seven different loci in 21 independent local populations.

The difference between direct and indirect estimates of gene flow in *E. editha* indicates that movement of individuals between existing populations cannot account for their genetic similarity. Instead, the current patterns are probably due to substantial gene flow in the recent past. That gene flow could possibly have been due to the large-scale movement between existing populations permitted by unusual environmental conditions or major range expansions producing the current geographic distribution.

Without further information, an estimate of Nm does not lead to an estimate m, the fraction of a population immigrating. If the average effective population size, N, can be estimated from census data, then m can be inferred. Sizes of local populations of *E. editha* vary between 200 and 3000 individuals. If, for the sake of discussion, we say that $N = 1000$, then $m \approx 0.0078$. Selection at the hexokinase locus would have to be at least that order of magnitude to account for the difference between the value of F_{ST} for it and the other loci. A potential problem with this approach is that estimates of N based on census data may suffer the same bias as do direct estimates of gene flow. Population sizes may differ greatly in time, so estimates of N based on current censuses may not reflect the average population sizes in the past.

Studies of *Drosophila pseudoobscura* show the same difference between direct and indirect estimates of levels of gene flow. This species is found throughout the western United States and most of Mexico and Central America and was studied extensively by Dobzhansky, in collaboration with Wright, with the intention of estimating dispersal rates and local population sizes and testing the shifting balance theory (*29*). Dispersal distances in *D. pseudoobscura* depend on habitat. Adults disperse an average of approximately 500 m during their lifetimes in what appears to be optimal conditions in forests and several kilometers in deserts, which are much poorer habitats (*30*). Yet there is little differentiation of any of the North American populations, which is difficult to account for with observed dispersal distances. As in *E. editha*, the pattern would be consistent with a recent dispersal of populations to the present range. For *Drosophila*, at least, we know populations can spread rapidly because *D. subobscura* and *D. ambigua*, both European species, have been found in North America only within the past 10 years, and yet have extended their geographic ranges from British Columbia to northern California (*30*). *Drosophila subobscura* has also spread rapidly in South America (*30*).

For some other species, direct and indirect methods give consistent results. The mussel *Mytilus edulis* has planktonic larvae that are capable of long-distance dispersal. An analysis of allele frequencies indicates that there are high levels of gene flow among populations throughout the eastern coast of North America (*26*). The patterns of allele frequencies in several species of plethodontid salamanders indicate that there is little or no gene flow among populations, which is in agreement with observations that these salamanders disperse very little. Although no one has attempted an exhaustive survey, there appear to be no species for which indirect estimates indicate substantially lower levels of gene flow than do direct estimates.

Evolutionary Implications

Our current understanding of gene flow in natural populations is far from complete, but population genetics theory does provide a guide to what can happen, and recent studies of natural populations are beginning to indicate what patterns are found in nature. Whether gene flow is a potentially constraining or creative force seems to depend on whether a species has a stable demographic structure over evolutionarily significant amounts of time. If the geographic distribution of a species remains the same and if local populations persist for long times, then gene flow occurs primarily through the movement of individuals between established populations, with the amount of gene flow dependent on the breeding biology of the species. In such species, gene flow generally plays a conservative role because it prevents the genetic differentiation of local populations and inhibits speciation. If gene flow is infrequent, as it appears to be in some salamanders, then each population evolves independently. The absence of gene flow, however, does not necessarily trigger rapid evolution; salamanders in particular have evolved very little in their long history (*31*).

The greatest opportunity for gene flow and population subdivision to play an important evolutionary role is in species with unstable population structures, either because of frequent extinction and recolonization of local populations or because of occasional large-scale changes in geographic range. Many kinds of species, especially parasites and "weedy species," are known to have unstable local populations. The application of indirect methods to gene frequency data is showing that some other species may also have unstable population structures, at least when considered on an evolutionary time scale.

Techniques from molecular biology hold the promise of providing much more detailed information about the genetic structure of natural populations than has been previously been available. The analysis of restriction site polymorphisms and DNA sequences has already been used extensively for reconstructing phylogenies of different species. There is the hope at least that these methods will reveal how recently closely related species have diverged, and on a fine scale, what sorts of genetic changes are associated with species formation. In the past, population geneticists have been limited in their abilities to characterize closely related species and populations of the same species. Furthermore, these methods also hold great promise for revealing more about the pattern of genetic variation within species (*32*). It is now possible to infer the phylogenies of individual DNA sequences, and the next step will be to trace the spread of those sequences through a species and between species.

REFERENCES AND NOTES

1. P. J. Bowler, *Evolution: The History of an Idea* (Univ. of California Press, Berkeley, CA, 1984).
2. E. Mayr, *Systematics and the Origin of Species* (Columbia Univ. Press, New York, 1942); *Animal Species and Evolution* (Harvard Univ. Press, Cambridge, MA, 1963); *Population, Species, and Evolution* (Harvard Univ. Press, Cambridge, MA, 1970).
3. P. R. Ehrlich and P. H. Raven, *Science* **165**, 1228 (1969).
4. S. Wright, *Genetics* **16**, 97 (1931).
5. M. Slatkin, *Theor. Pop. Biol.* **12**, 253 (1977); T. Maruyama and M. Kimura, *Proc. Natl. Acad. Sci. U.S.A.* 77, 6710 (1980).
6. J. B. S. Haldane, *Proc. Cambridge Phil. Soc.* **26**, 220 (1930); T. Nagylaki, *Genetics* **80**, 595 (1975).
7. N. H. Barton, *Evolution* **37**, 454 (1983).
8. ———, *Heredity* **43**, 333; B. O. Bengtsson, in *Evolution: Essays in Honour of John Maynard Smith*, P. J. Greenwood, P. H. Harvey, M. Slatkin, Eds. (Cambridge Univ. Press, Cambridge, 1985), pp. 31–42.
9. S. Wright, *Proc. Sixth Int. Cong. Genet.* **1**, 356 (1932). For more recent reviews of Wright's theory, see S. Wright [*Evolution* **36**, 427 (1982)] and W. Provine [*Sewall Wright and Evolutionary Biology* (Univ. of Chicago Press, Chicago, 1986)].
10. Wright's original definition of an adaptive surface was not presented in terms that could be easily related to population genetic theory [see Provine in (*9*)]. The definition of an adaptive surface used here came later and incorporates both R. A. Fisher's "Fundamental theorem of natural selection" [*The Genetical Theory of Natural Selection* (Clarendon Press, Oxford, 1931)] and Wright's (*4*) demonstration that the mean fitness is a potential function. If there is frequency-dependent selection or strong linkage disequilibrium among loci, the surface of mean fitnesses is not sufficient to predict the direction of evolutionary changes [W. J. Ewens, *Mathematical Population Genetics* (Springer-Verlag, New York, 1979)].
11. R. Lande, *Evolution* **33**, 234 (1979); M. Slatkin, *ibid.* **35**, 477 (1981); S. Rouhani and N. H. Barton, *Theor. Pop. Biol.*, in press.
12. R. Lande, *Evolution* **38**, 743 (1984).
13. S. Wright, *Am. Nat.* **90**, 5 (1956).
14. J. Maynard Smith, *Nature (London)* **201**, 1145 (1964).
15. M. J. Wade, *Q. Rev. Biol.* **53**, 101 (1978); *Science* **210**, 665 (1980).
16. B. R. Levin *et al.*, in *Population Biology of Infectious Diseases* (Springer-Verlag, Berlin, 1982), pp. 213–243.
17. R. M. May and R. M. Anderson, *Proc. R. Soc. London* **B219**, 281 (1983).
18. J. Felsenstein, *Evolution* **35**, 124 (1981).
19. E. Mayr, in *Evolution as a Process*, J. Huxley, A. C. Hardy, E. B. Ford, Eds. (Allen & Unwin, London, 1954), pp. 157–180.
20. H. L. Carson and A. R. Templeton, *Annu. Rev. Ecol. Syst.* **15**, 97 (1984).
21. N. H. Barton and B. Charlesworth, *ibid.*, p. 133.
22. M. Slatkin, *ibid.* **16**, 393 (1985).
23. P. R. Ehrlich, R. A. White, M. C. Singer, S. W. McKechnie, L. E. Gilbert, *Science* **188**, 221 (1975).
24. S. Wright, *Ann. Eugenics* **15**, 323 (1951); J. F. Crow and K. Aoki, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 6073 (1984).
25. B. S. Weir and C. C. Cockerham, *Evolution* **38**, 1358 (1984).
26. M. Slatkin, *ibid.* **39**, 53 (1985).
27. N. H. Barton and M. Slatkin, *Heredity* **56**, 409 (1986); unpublished manuscript.
28. S. W. McKechnie, P. R. Ehrlich, R. R. White, *Genetics* **81**, 571 (1975).
29. Th. Dobzhansky and S. Wright, *ibid.* **28**, 304 (1943).
30. A. T. Beckenbach and A. Prevosti, *Am. Midl. Nat.* **115**, 10 (1986).
31. D. B. Wake, G. Roth, M. H. Wake, *J. Theor. Biol.* **101**, 211 (1983).
32. J. C. Avise, E. Bermingham, L. G. Kessler, N. C. Saunders, *Evolution* **38**, 931 (1984); M. Kreitman, *Nature (London)* **304**, 412 (1983).
33. I thank R. E. Lenski, R. M. May, J. K. Moore, T. Prout, and M. J. Wade for helpful discussions of this topic, and N. H. Barton, J. J. Bull, and two anonymous reviewers for numerous useful comments on an earlier draft of this paper. NSF grant BSR-8500258 has provided partial financial support.