6

GENE FLOW AND POPULATION SUBDIVISION

In deriving the Hardy–Weinberg law in Chapter 2, we assumed that the population was completely isolated. Isolation means that all individuals that contribute to the next generation come from the same population with no input at all from individuals from other populations. However, most species consist of not just one deme but rather many **local populations or subpopulations** consisting of the individuals inhabiting a geographic area from which most mating pairs are drawn that is generally small relative to the species' total geographic distribution. Although most matings may occur within a local population, in many species there is at least some interbreeding between individuals born into different local populations. Genetic interchange between local populations is called **gene flow**. In Chapter 1 we noted that DNA replication implies that genes have an existence in space and time that transcends the individuals that temporally bear them. Up to now, we have been primarily focused upon a gene's temporal existence, but with gene flow we begin to study a gene's spatial existence. In this chapter, we will study the evolutionary implications of gene flow and investigate how a species can become subdivided into genetically distinct local populations when gene flow is restricted. Restricted gene flow leads to variation in the frequency of a gene over space.

GENE FLOW

Gene Flow Between Two Local Populations

We start with a simple model in which two infinitely large local populations experience gene flow by symmetrically exchanging a portion m of their gametes each generation. We will monitor the evolution of these two populations at a single autosomal locus with two neutral alleles (A and a). The basic model is illustrated in Figure 6.1. In this simple model there is no mutation, selection, or genetic drift. For any given local population, we assume that a portion $1 - m$ of the gametes are sampled at random from the same local area and

Population Genetics and Microevolutionary Theory, By Alan R. Templeton

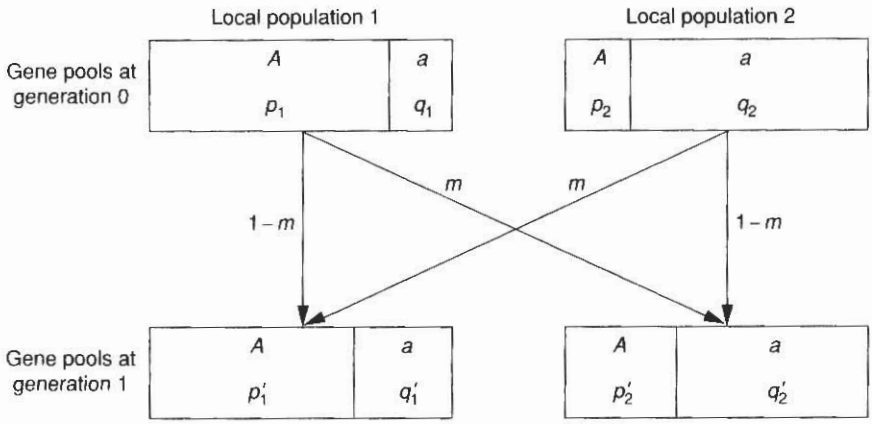

Figure 6.1. Model of symmetrical gene flow between two populations. The boxes represent the gene pools at an autosomal locus with two alleles, *A* and *a*, for the two populations over two successive generations, with m of the genes being interchanged between the two localities and $1 - m$ staying within the same locality.

that a portion m of the gametes are sampled at random from the other local population's gene pool (that is, gene flow). Letting p_1 be the initial frequency of the *A* allele in local population 1 and p_2 be the initial frequency of *A* in local population 2, the allele frequencies in the next generation in the two local populations are

$$p_1' = (1 - m)p_1 + mp_2 \qquad p_2' = (1 - m)p_2 + mp_1 \tag{6.1}$$

We can now see if evolution occurred by examining whether or not the allele frequencies in either local population change across the generations:

$$\begin{aligned} \Delta p_1 &= p_1' - p_1 = (1 - m)p_1 + mp_2 - p_1 = -m(p_1 - p_2) \\ \Delta p_2 &= -m(p_2 - p_1) \end{aligned} \tag{6.2}$$

Equations 6.2 show that gene flow acts as an evolutionary force (that is, gene flow alters allele frequencies) if the following two conditions are satisfied:

- $m > 0$ (the local populations have some genetic exchange and are not completely reproductively isolated) and
- $p_1 \neq p_2$ (the local populations have genetically distinct gene pools).

In other words, gene flow is an evolutionary force when it occurs between populations with distinct gene pools.

It is important to keep in mind that m is defined in terms of the gene pools, and therefore m represents the amount of exchange of *gametes* between the local populations and not necessarily individuals. In some species, gametes are exchanged directly without the individuals moving at all. For example, most trees are wind pollinated, and the pollen (regarding these haploid gametophytes as essentially being gametes) can be blown for hundreds of miles by the wind. Hence, tree populations that are quite distant can still experience gene flow, yet no diploid trees are walking back and forth! For many other species, m requires

that individuals move from their local population of birth to a different local population, followed by reproduction in their new location. Because gene flow requires both movement and reproduction, m is not just the amount of dispersal of individuals between local populations, but instead m represents a *complex interaction* between the pattern of dispersal and the system of mating. For example, system-of-mating inbreeding can greatly reduce gene flow, even if the individuals are in physical proximity. As mentioned in Chapter 3, the Tamils of India preferentially marry cousins. As a result of this inbreeding system of mating, the Tamils have little gene flow with other peoples with whom they physically intermingle.

Assortative mating can also greatly reduce the amount of gene flow. The European corn borer, an insect pest, has two pheromone races that apparently had once been geographically separated but are now broadly overlapping (Harrison and Vawter 1977). There is strong assortative mating for pheromone phenotype in these insects with greater than 95% of the matings occurring within the pheromone types (Malausa et al. 2005). Moreover, these races have allele frequency differences at many isozyme loci (Appendix 1) because of their historical isolation. Recall from Chapter 3 that when two previously isolated, genetically differentiated populations make genetic contact with one another, extensive linkage disequilibrium is created in the mixed population:

$$D_{\text{admixture}} = m(1 - m)(p_1 - p_2)(k_1 - k_2) \tag{6.3}$$

where the p's refer to the allele frequencies in the two populations at one locus and the k's are the respective frequencies at a second locus. Thus, in the areas of overlap of the pheromone races, there is linkage disequilibrium between the pheromone loci and all other loci having allele frequency differences between the historical races. Because assortative mating reduces the chances that individuals from the different pheromone races will mate with one another, it also reduces the effective gene flow m for all loci that had different allele frequencies in the historical races. As a result, assortative mating for pheromone type greatly reduces gene flow as an evolutionary force for all differentiated loci. Despite close physical proximity of individual corn borers, the effective m is very small and the races have maintained their differentiation even at isozyme loci that have no direct impact on the pheromone phenotype.

In contrast, disassortative mating enhances m for all loci. As mentioned in Chapter 3, disassortative mating systems give a reproductive advantage to individuals who are phenotypically dissimilar to the majority of individuals in the population. Often, dispersing individuals tend to deviate more on the average from the phenotypic means of the population into which they have dispersed. This gives dispersing individuals a reproductive advantage in their new population, thereby enhancing gene flow for all the genes borne by the dispersing individuals. For example, *Drosophila melanogaster* has a strong disassortative mating pheromone system (Averhoff and Richardson 1974, 1975), just the opposite of the European corn borer. Across the globe *D. melanogaster* is predominately a single, cosmopolitan species showing only modest geographical differentiation (except for some selected loci) even on a continental basis (Singh and Rhomberg 1987).

It is also important to note that assortative or disassortative mating on a *nongenetic* phenotype can influence m for many loci as long as the phenotypic differences influencing system of mating are correlated with the historical gene pool differences. This was already noted for the Amish human populations (Chapter 3), who have assortative mating based on religion and who, as a consequence, have little gene flow with surrounding populations and maintain genetic distinctiveness from their neighbors.

An example involving assortative mating on both genetic and nongenetic phenotypes is provided by gene flow between human populations of European origin and of African origin in the United States and in northeastern Brazil. In North America, European settlers brought in African slaves mainly from 1700 to 1808, with 98% of Africans coming from West and West-Central Africa. Once in North America, gene flow occurred between peoples of European and African origin, even though there was a tendency for assortative mating on the basis of skin color (Chapter 3). The people resulting from matings between individuals of European and African origin have been *socially* classified as "blacks" in North America. Genetically, and phenotypically for skin color, such people of mixed ancestry are intermediate and are no more black than they are "white." The social recognition of just two primary skin color categories is therefore a cultural decision, not a biological one. Nevertheless, this cultural classification has a direct and strong biological impact because it is coupled with assortative mating. The factors of assortative mating by the cultural "skin color" category, the cultural decision to classify people of mixed ancestry as blacks, and the numerical predominance of whites all combine to create an asymmetrical gene flow pattern. Effectively, much more gene flow occurred from whites to blacks than in the other direction in North America. A simplified version of this asymmetrical gene flow is given in Figure 6.2. In that figure, M is the effective amount of gene flow over the entire relevant period of North American history (in contrast to m, which is a per-generation gene flow parameter).

As can be seen from Figure 6.2, this pattern of gene flow (simplified relative to the actual pattern) results in an allele frequency of current African Americans (p_A) of $Mp_E + (1 - M)p_W$, where p_E and p_W are the allele frequencies in the ancestral European and West African populations, respectively. Solving for M in Figure 6.2, we can estimate M from the allele frequencies as

$$M = \frac{p_A - p_W}{p_E - p_W} = \frac{\text{change in allele frequency in African Americans from West Africans}}{\text{initial difference in allele frequency between Europeans and West Africans}} \tag{6.4}$$

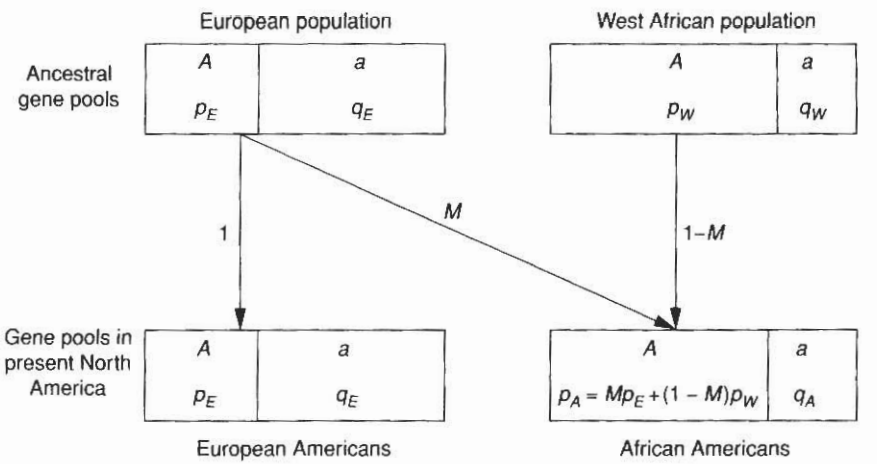

Figure 6.2. Model of asymmetrical gene flow between two populations representing simplified version of gene flow between Europeans and West Africans in North America to produce current African American population. In this model, M represents the cumulative impact of gene flow over several centuries.

For example, at the autosomal *Rh* blood group locus, the frequency of the *Rh*+ allele is 0.4381 in African Americans (p_A), 0.5512 in West Africans, and 0.0279 in European Americans. Assuming the current West African allele frequencies have not changed much over the last few centuries, $p_W = 0.5512$. Assuming that the current European American allele frequencies are close to the ancestral European immigrant allele frequencies because European Americans have not been so strongly influenced by gene flow, then $p_E = 0.0279$. These *Rh* allele frequencies yield $M = 0.216$ from equation 6.4. This value is typical of African Americans in North America (Reed 1969). What this number tells us is that the African American gene pool has been affected by gene flow such that it was about 20% European in origin and 80% West African at the time of these studies. However, there is much variation in the degree of admixture among different local populations of African Americans. For example, African Americans living in Columbia, South Carolina, have $M = 0.18$. In contrast, the Gullah-speaking Sea Island African Americans that live in nearby coastal South Carolina have $M = 0.035$. This low amount of admixture is consistent with the history of the Gullah-speaking African Americans, who have been relatively isolated throughout their entire history, have lived in an area that has always had an African American majority, and have retained many aspects of African culture, including their language (Parra et al. 2001).

European and West African populations were also brought into physical contact in northeastern Brazil at about the same time as North America. However, the social definitions of "race," particularly for individuals of mixed ancestry, were and are different from those used in North America. In northeastern Brazil, a number of alternative categories are available for individuals of mixed ancestry, and many individuals who would be socially classified as blacks in North America would not be considered blacks in Brazil. In a study of Brazilians (Franco et al. 1982), the Brazilian authors used the term white in the context of Brazilian culture. The gene pool of these whites was estimated to be 67% of European origin, 20% of West African origin, and 13% of Amerindian origin, using an equation similar to 6.4. In contrast, the "nonwhites" were 58% European, 25% African, and 17% Amerindian. Using just skin color and not social classification, the Brazilian subjects were also characterized from "most Caucasoid" to "most Negroid." In northeast Brazil, the most Caucasoid group is 71% European, in contrast to the nearly 100% found in North American whites. The most Negroid Brazilian group is 28% European—an amount of admixture greater than that of the average African American from North America. No matter how one categorizes the Brazilians in this study, it is obvious that there has been much more gene flow between the European and African gene pools in northeastern Brazil as compared to North America. Hence, the *cultural* systems of mating in the two countries have had a major genetic impact on the composition of their present-day populations despite similar initial founding populations and proportions. This example shows that m or its multigenerational cumulative analogue M is determined not just by physical movement of individuals but also by system of mating as influenced by genetic and nongenetic factors.

Genetic Impact of Gene Flow

We have already seen that allele frequencies are altered when gene flow occurs between genetically distinct populations. Gene flow therefore can be an evolutionary force. In this section, we will see that gene flow causes evolution in a nonrandom, predictable fashion. To show this, we will return to our simple model of symmetrical gene flow given in Figure 6.1. Starting with the initial populations prior to gene flow, their genetic distinctiveness is

measured by the difference in their allele frequencies; that is, $d_0 = p_1 - p_2$. After one generation of gene flow, Figure 6.1 shows that

$$p_1' = (1 - m)p_1 + mp_2 = p_1 - m(p_1 - p_2) = p_1 - md_0 \tag{6.5}$$

and similarly,

$$p_2' = p_2 + md_0 \tag{6.6}$$

Hence, the difference in gene pools between the two local populations after a single generation of gene flow is

$$d_1 = p_1' - p_2' = p_1 - md_0 - p_2 - md_0 = d_0(1 - 2m) \tag{6.7}$$

Note that equation 6.7 implies that $|d_1| < |d_0|$ for all $m > 0$ and $d_0 \neq 0$. By using the above equations recursively, the difference in allele frequencies between the two local populations after t generations of gene flow is

$$d_t = d_0(1 - 2m)^t \to 0 \quad \text{as } t \to \infty \tag{6.8}$$

Therefore, *gene flow decreases the allele frequency differences between local populations.*

Now consider a special case of Figure 6.1 in which $p_1 = 0$ and $p_2 = 1$. In this case, the frequency of the A allele in the population 1 gene pool will go from being completely absent to being present with a frequency of m. This evolutionary change caused by gene flow mimics that of mutation. If the mutation rate from a to A were μ, then the evolutionary change caused by mutation in a population initially lacking the A allele would be to introduce that allele with a frequency of μ. Hence, *gene flow can introduce new alleles into a population,* with m being the analog of the mutation rate. One major difference between gene flow and mutation as sources of new genetic variation for a local deme is that in general μ is constrained to take on only very small values, whereas m can be either small or large. A second major difference is that gene flow can introduce variation at many loci simultaneously, whereas mutation generally affects only one locus or nucleotide site at a time. A third major difference is that many new mutations are deleterious (Figure 5.3) and initially occur as single copies, thereby ensuring that many are rapidly lost from the population. In contrast, gene flow introduces genetic variation that has usually been around for more than one generation and can introduce multiple copies of new variants. Hence, there is the potential for a massive influx of new genetic variability through gene flow that can drastically alter a local gene pool, even in a single generation.

The effects of gene flow on genetic variation between and within local populations described above can be summarized as *gene flow decreases genetic variability between local populations and increases genetic variability within a local population.* Recall from Chapter 4 that genetic drift causes an increase in genetic variability between populations (their allele frequencies diverge) and decreases genetic variability within a population (loss and fixation of alleles). Hence, the effects of gene flow on within- and between-population genetic variability are the *opposite* of those of genetic drift.

In Chapter 2, we introduced the idea of population structure as the mechanisms or rules by which gametes are paired together in the reproducing population. We now include in those rules the exchange of gametes among local populations (gene flow). Parallel to this

process-oriented definition of population structure, there is also a pattern-oriented definition. **Population structure** is the amount of genetic variability and its distribution within and among local populations and individuals within a species. This definition emphasizes the spatial *patterns* of genetic variation that emerge from the rules of gametic exchange. The pattern of genotypic variability (heterozygosity versus homozygosity) among individuals within a local population is highly dependent upon the system of mating, as we saw in Chapter 3. As mentioned above, the distribution of allelic variation within and among local demes is influenced by both gene flow and genetic drift. Therefore, genetic population structure has three major components:

- System of mating
- Genetic drift
- Gene flow.

Because of the opposite effects of gene flow and genetic drift, the *balance* between drift and gene flow is a *primary determinant* of the genetic population structure of a species.

The concept of genetic population structure (hereafter called population structure) is critical for the remainder of this book. Genotypic variability provides the raw material for all evolutionary change, including that caused by natural selection. Population structure therefore determines the pattern and amount of genetic variability that is available for evolution within a species. As will be seen later, natural selection and other evolutionary forces operate within the constraints imposed by the population structure. Hence, virtually all evolutionary predictions, particularly those related to adaptive evolution, must always be placed in the context of population structure.

Given the central importance of population structure to microevolutionary processes, we need additional tools to measure and quantify it. The tools for measuring system of mating have already been discussed in Chapter 3 and those for drift in Chapters 4 and 5, so now we need to develop measures for the balance between gene flow and drift.

BALANCE OF GENE FLOW AND DRIFT

Recall from Chapter 4 that to measure the impact of genetic drift upon identity by descent, we started with equation 4.3:

$$\bar{F}(t) = \frac{1}{2N} + \left(1 - \frac{1}{2N}\right)\bar{F}(t-1)$$

where N is replaced by the inbreeding effective size for nonideal populations. To examine the balance between drift and mutation, we modified the above equation to yield equation 5.4:

$$\bar{F}(t) = \left[\frac{1}{2N} + \left(1 - \frac{1}{2N}\right)\bar{F}(t-1)\right](1-\mu)^2$$

Because gene flow and mutation behave in an analogous manner with respect to genetic variation within a local deme, a similar modification of equation 4.3 can be used to address the following question: Suppose a local deme of inbreeding effective size N_{ef} is experiencing gene flow at a rate of m per generation from some outside source. What