Hardy–Weinberg Equilibrium and the Foundations of Evolutionary Genetics

3. Incorporating Mutation and Migration

Amitabh Joshi

Amitabh Joshi studies and teaches evolutionary genetics and population ecology at the Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore. His current research interests are in life-history evolution and population dynamics. He also enjoys music (especially traditional qawwali in Braj, Farsi, Punjabi and Urdu), history, philosophy, and reading and writing poetry in Urdu, and English.

Previous parts:
Resonance, Vol.13:
Part 1. The Dance of the Genes, No.9, pp.812–835, 2008;
Part 2. The Inertial State in Population Genetics, No.10, pp.951–970, 2008.

Keywords
Mutation, migration, gene flow, equilibrium, dynamics.

The year 2008 marked the 150th anniversary of the debut of the concept of natural selection as a mechanism of adaptive evolution through the reading of papers by Darwin and Wallace to the Linnaean Society. It also marked the 100th anniversary of the enunciation of the principle defining the inertial state of populations from a genetic viewpoint through the independent publication of papers on the topic by G H Hardy and W Weinberg. In this series, we examine the significance of the Hardy–Weinberg Equilibrium as a basic model of population genetics that forms the foundation for evolutionary genetics. In Part 3 of this series, we begin to relax some of the assumptions of the basic model underlying the Hardy–Weinberg Equilibrium, and analyse the situations where either mutation or migration occur in an otherwise ideal large population.

We shall now begin to investigate the consequences of relaxing the assumptions of the ideal large population. We begin with a schematic depiction of the ideal large population (*Figure* 1) which captures the essence of the scheme described in the beginning of Part 2. This figure represents the inertial state for a single locus: allele frequencies do not change from one generation to the next, and genotypic frequencies can change to Hardy–Weinberg Equilibrium values during the transition from the pool of gametes to the pool of possible zygotes. Once the genotypic frequencies are at their equilibrium values, they too will not change over generations. Henceforth,

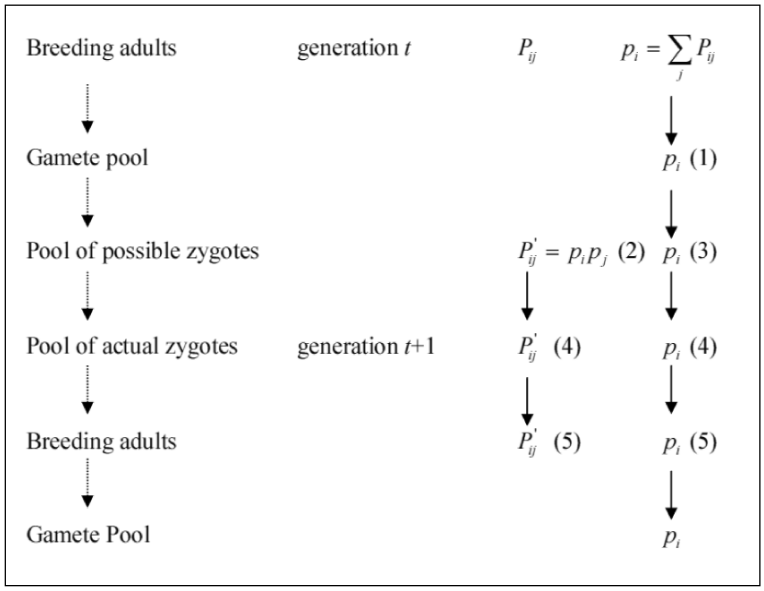

Figure 1. A schematic depiction of the important genetic events happening in an ideal large population. Dotted arrows show the different life-stages in the transition between generations, whereas solid arrows indicate no change in genotypic or allelic frequency between the two life-stages being considered. Thus, allele frequencies do not change at all while genotypic frequency changes during the transition from breeding adults to possible zygotes. The numbers in parentheses indicate the transitions for which the equality of frequencies would be violated if a certain assumption of the ideal large population were to be relaxed:* (1) *mutation, meiotic drive, migration, fertility selection, sexual selection* (2) *non-random mating,* (3) *gametic selection,* (4) *small population size,* (5) *viability selection (see text for a detailed explanation).

we will use this depiction repeatedly as a framework within which we analyse the consequences of relaxing the various assumptions going into the formulation of the ideal large population. The major effects of different evolutionarily important phenomena are also summarized in *Figure* 1, by pin-pointing the life-stage transition at which their effect is seen. Thus, mutations in germ-line cells and migration of individuals (or hapoloid propagules such as pollen) can both be assumed to alter the allele frequency in the gamete pool relative to that among the breeding adults.

Mutations in germ-line cells and migration of individuals can both be assumed to alter the allele frequency in the gamete pool relative to that among the breeding adults.

Often, certain alleles at a locus are able to obtain greater representation than expected based on their frequency. For example, among gametes formed by A_1A_2 individuals, we may find instances where A_1 gametes consistently account for more than half the number of gametes. This is a phenomenon called *meiotic drive*, and

Genotypic and, therefore, allelic frequencies can also change due to sampling errors during the transition from the pool of possible zygotes to that of actual zygotes, if a population is small in numbers (*random genetic drift*).

it also tends to alter allele frequencies during the breeding adult–gamete pool transition. Similarly, allele frequencies can change between the breeding adult and gamete pool stages if some genotypes are more likely to obtain matings than others (*sexual selection*), or if certain genotypes are likely to produce relatively more offspring per mating (*fertility selection*). If mating is not random, as for example in the case of preferential mating among either relatives (*inbreeding*) or individuals of similar phenotype (*assortative mating*), genotypic frequencies may no longer follow the relationship $P'_{ij} = p_i p_j$. Non-random mating in itself, however, does not alter allele frequencies in large populations, unless accompanied by some kind of sexual selection. If certain gametes are more successful at fertilization than others, allele frequencies can change between the gamete pool and pool of possible zygotes (*gametic selection*). Genotypic and, therefore, allelic frequencies can also change due to sampling errors during the transition from the pool of possible zygotes to that of actual zygotes, if a population is small in numbers (*random genetic drift*). And, finally, if certain genotypes are more likely to survive to become breeding adults, both genotypic and allelic frequencies can change during the actual zygotes–breeding adults transition. This is a very common form of natural selection, called *viability selection*. There is substantial empirical evidence for the operation of all these phenomena in real populations. Indeed, it is through these phenomena that the ecology, environment and genetic systems of organisms interact to give rise to evolutionary change which is then manifested as the diversity of living forms we marvel at. We will undertake more detailed discussions of many of these phenomena as we go along, but for now we will focus on mutation and migration.

Mutation

It is well known in genetics that an allele can mutate (change) into another allele. There can be many types

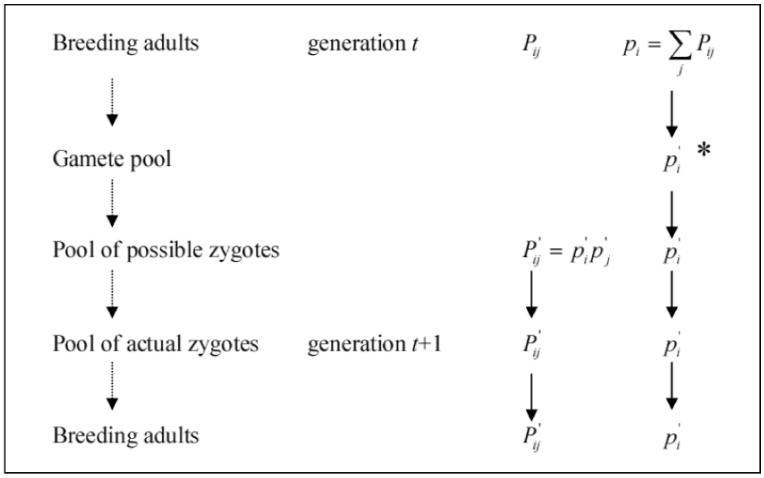

Figure 2. A schematic depiction of the consequence of allowing mutation (or migration) in an otherwise ideal large population. The asterisk indicates the effect of allowing mutation: allele frequencies in the gamete pool can change, relative to what they were in the breeding adults.

of mutations, based on the underlying mechanism, and the precise nature of the change in the DNA sequence, but we will not concern ourselves with those details. Typically mutations occur at rates of about 10^{-6} (one in a million) per locus per generation, although there is considerable variation around this very approximate average rate, and often different parts of the genome have very different rates of mutation. Mutations can occur in non-reproductive tissue and lead to severe consequences (e.g., cancer) but we restrict ourselves to germ-line mutations that occur in the sequence of cell divisions leading up to gamete formation. The consequence of these mutations is that the allele frequency is liable to change when going from breeding adults to the gamete pool (*Figure* 2).

Let us imagine a one-locus system with two alleles that follows all assumptions of the ideal large population, except that we now permit mutation from A_1 to A_2 at a rate of u per generation, and from A_2 to A_1 at a rate of v per generation. Random mating allows us to obtain genotypic frequencies from allele frequencies, and since there are only two alleles, we need to track just one allele frequency to have a description of the genetic composition of the system. Let the frequency of allele A_1 among breeding adults in generation t be p_1, and in

the gamete pool (and therefore in generation $t + 1$) be p_1'. All A_1 alleles in generation $t + 1$ must arise in one of two ways: they can be copies of A_1 alleles in generation t that did not mutate (probability $1 - u$), or they can have arisen by mutation from A_2 alleles in generation t (probability v). We can write the recursion for allele frequencies (using the relationship $p_2 = 1 - p_1$) as

$$\begin{aligned} p_1' &= p_1(1-u) + (1-p_1)v, \\ &= p_1 - p_1u - p_1v + v = p_1(1-u-v) + v. \end{aligned}$$

We now have our recursion, but what we would like to understand is the long-term behaviour of this system. For all practical purposes, we can describe our system by the allele frequency p_1, because knowing this we can obtain p_2, as well as all genotypic frequencies P_{ij}. In the terminology of dynamic systems, the allele frequency p_1 can, therefore, be considered the *state variable* for this system because its value at any point in time defines the composition or *state* of the system. For any dynamic system (i.e., a system whose state changes with time) that one is trying to understand, one would like to know primarily whether the state has an equilibrium or not,

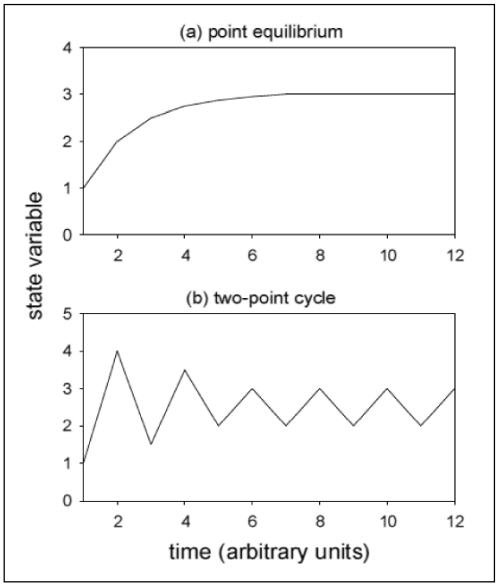

Figure 3. Illustration of two types of equilibria for a system described by a single state variable plotted here against time. In panel (a) the state variable settles down at a value of 3.0 and remains at it thereafter; this is a point equilibrium. In panel (b) the system settles down into a particular repeating sequence of state variable values. Such equilibria are also called limit cycles. In the example shown, the equilibrium sequence of states is for the state variable to alternate between the values of 2.0 and 3.0 at each successive time step. This is an example of a two-point cycle, because the system alternates between two states.

and, if so, whether the equilibrium is stable or unstable. In addition, it is usually also of interest to know how system approaches the equilibrium if it has one.

An equilibrium is a state of the system such that once the system has attained that state, it no longer changes with time. Thus, at an equilibrium, the change over time in the value of the state variable(s) should be zero. Often an equilibrium is a point equilibrium, meaning that there is one particular state of the system (described by a set of one value of each state variable) that, once attained, does not change (*Figure* 3a). In other situations, the equilibrium may be a repeating sequence of states, as in a system exhibiting regular oscillations (*Figure* 3b). The notion of stability of an equilibrium is connected with what happens if the system is shifted slightly from the equilibrium state: if it tends to return to the equilibrium following a perturbation, the equilibrium is stable; if it tends to go further away, the equilibirum is unstable. We will discuss these concepts in a little more detail as we encounter examples of different kinds of equilibria.

An equilibrium is a state of the system such that once the system has attained that state, it no longer changes with time. Thus, at an equilibrium, the change over time in the value of the state variable(s) should be zero.

To return to our population with mutation, once we have the recursion for allele frequency, how do we decide whether there will be an equilibrium or not? By definition, at equilibrium, the system should not change in state and, thus, the change in the value of the state variable should be zero. In other words, if there is a point equilibrium, there should be a value of allele frequency for which the change in allele frequency between generations ($\Delta p_1 = p_1' - p_1$) is zero. This situation can be easily explored as follows:

$$\begin{aligned} \Delta p_1 &= p_1' - p_1 = p_1(1-u-v) + v - p_1 \\ &= p_1(1-u-v-1) + v = v - p_1(u+v). \end{aligned}$$

The question to be asked now is whether there is a value of p_1 for which $\Delta p_1 = 0$. Clearly, $\Delta p_1 = 0$ if $v = p_1(u+v)$; the equilibrium frequency of the allele A_1, therefore, is the value satisfying $v = p_1(u+v)$, and this

The notion of stability of an equilibrium is connected with what happens if the system is shifted slightly from the equilibrium state: if it tends to return to the equilibrium following a perturbation, the equilibrium is stable; if it tends to go further away, the equilibirum is unstable.

is easily obtained by rearrangement as $\hat{p}_1 = v/(u+v)$ (the little 'hat' or caret above p_1 indicates that this is an equilibrium frequency). We would next like to examine the question of whether this equilibrium is stable. In order to do that more easily, we will first derive an expression linking allele frequencies any arbitrary number of generations apart. This formulation will not only allow us to test for the stability of the equilibrium allele frequency in a straightforward manner, but will also permit a discussion of the time dynamics of a one-locus two-allele system under mutation.

We start with the equation we derived earlier, $p_1(t+1) = p_1(t)\,(1-u-v)+v$. Note that we have substituted an explicit noting of generation within parentheses for the 'prime' notation used earlier because we want to extrapolate to an arbitrary number of generations. Now, we define two new constants X and Y, that depend only on the value of u and v, as $X = v/(u+c)$, and $Y = 1-u-v$. Note that X can, therefore, also be written as $X = v/(1-y)$. The reason for this seemingly unnecessary algebraic jugglery will shortly become clear. So, if we now use X and Y, we can reframe the recursion for allele frequency as follows:

$$\begin{aligned} p_1(t+1) &= p_1(t)\,(1-u-v)+v \\ &= p_1(t)Y + X(1-Y) \\ &= p_1(t)Y - XY + X \\ &= X + Y\,(p_1(t)-X)\,. \end{aligned}$$

If we now substitute back the values of X and Yinto this expression, we get the recursion in the following form:

$$p_1(t+1) = \frac{v}{(u+v)} + \left[p_1(t) - \frac{v}{(u+v)}\right](1-u-v)\,. \quad (1)$$

In this form, it is easy to see that given an initial allele frequency of $p_1(0)$, the frequency after an arbitrary

number of generations, t, can be obtained by

$$p_1(t) = \frac{v}{(u+v)} + \left[p_1(0) - \frac{v}{(u+v)}\right](1-u-v)^t. \quad (2)$$

To examine the stability of the equilibrium allele frequency, consider (1). At equilibrium, $\hat{p}_1 = v/(u+v)$, which implies $\left[p_1 - \frac{v}{(u+v)}\right](1-u-v) = 0$, and the frequency in the next generation will, therefore, remain $\hat{p}_1 = v/(u+v)$. Now imagine increasing the allele frequency in generation t, so that $p_1(t) = x + v/(u+v)$. Then, applying (1), the frequency in the next generation will be given by $p_1(t+1) = x(1-u-v) + (v/(u+v))$, and because $(1-u-v) < 1$ for non-zero u and v, it follows that $p_1(t+1) < p_1(t)$. It can similarly be shown that if we perturb the allele frequency downwards from the equilibrium value, such that $p_1(t) < v/(u+v)$, then $p_1(t+1) > p_1(t)$. In other words, if we reduce the frequency of the A_1 allele below the equilibrium, it tends to increase, and if we increase it above the equilibrium value, it tends to decrease: the equilibrium for allele frequency $\hat{p}_1 = v/(u+v)$ is, therefore, a stable one, because there is a tendency for allele frequencies to go back toward the equilibrium value following a perturbation.

The dynamic behaviour of this system is determined by (2), and the way in which allele frequency changes with time for a particular numerical example is depicted in *Figure* 4. Note in this figure that (i) it takes a really

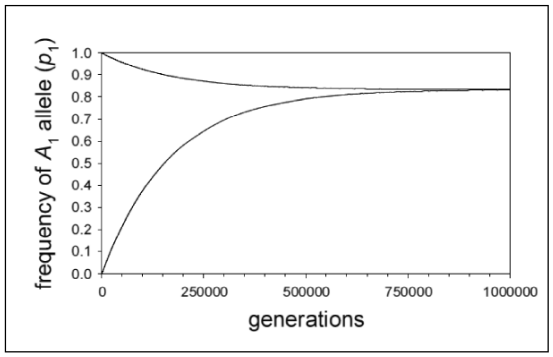

Figure 4. A plot of allele frequencies over generations in a one locus two allele system undergoing mutation from A_1 to A_2 at rate $u = 10^{-6}$, and from A_2 to A_1 at rate $v = 5 \times 10^{-6}$ per generation. The equilibrium frequency of A_1 is $v/(u+v) = 0.8333$.

long time for the system to get close to equilibrium (strictly speaking, equilibrium is actually reached only when $t = \infty$), and (ii) the rate of change of allele frequency over the first 250,000 generations or so is much higher when starting from a frequency of zero, as compared to when starting from a frequency of one. Both these features can be understood clearly in terms of (1) and (2). Let us first consider the differing rates of allele frequency change during the first 250,000 generations. The change in the frequency of the A_1 allele in one generation is

$$\begin{aligned}\Delta p_1 &= p_1' - p_1 \\ &= \frac{v}{(u+v)} + \left[p_1 - \frac{v}{(u+v)}\right](1-u-v) - p_1.\end{aligned}$$

Once again, for simplicity we can use $X = v/(u+v)$ and $Y = 1-u-v$ to designate the constant terms in this equation, which can, then, be re-written as follows.

$$\begin{aligned}\Delta p_1 &= p_1' - p_1 = X + (p_1 - X)Y - p_1 \\ &= X - XY + p_1 Y - p_1 = X(1-Y) - p_1(1-Y) \\ &= (1-Y)(X - p_1).\end{aligned}$$

In the long term, mutation is ultimately the only mechanism for generating new genetic variation, although if many loci are considered, the reshuffling of genomes that takes place every generation due to recombination itself generates a wide variety of multi-locus genotypes.

Since $(1 - Y)$ is a constant, the magnitude of Δp_1 is determined by that of $(X - p_1)$. Thus, for given mutation rates u and v, the magnitude change in allele frequency from one generation to the next will be greater when the allele frequency is farther from the equilibrium.

The factor in equation (2) determining the time taken to reach equilibrium for this system is the magnitude of $(1 - u - v)$, because as $(1 - u - v)^t$ approaches zero, the system approaches equilibrium, and how large t must be for $(1 - u - v)^t$ to approach zero depends on the magnitude of $(1 - u - v)$. The smaller the magnitude of u and v, the greater will be the value of t necessary for $(1 - u - v)^t$ to approach zero. In the example shown in *Figure* 4, u and v are of the order of 10^{-6}, and it takes

about 750,000 generations to get really close to the equilibrium allele frequency. If v and u were of the order of 10^{-4}, the system would approach close to equilibrium in about 35,000 generations, which, it should be noted, is still a really long time unless we are talking about microbial populations with a generation time in hours. Thus, mutation, even at slightly higher than average rates, is a rather weak evolutionary force in that it induces genetic change at fairly low rates. Yet, in the long term, mutation is ultimately the only mechanism for generating new genetic variation, although if many loci are considered, the reshuffling of genomes that takes place every generation due to recombination itself generates a wide variety of multi-locus genotypes. We have restricted our discussion to a simple one-locus two-allele case, because it permits us to introduce the way in which one can incorporate mutation into the framework of the ideal large population. There are, however, many further subtleties to the issue of mutation as an evolutionary force that interested readers may find stimulating.

In the real world, organisms move around: some individuals may leave a population, others may join it.

Migration

In the real world, organisms move around: some individuals may leave a population, others may join it. In the context of migration (often also termed *gene flow*), an individual could be an organism, a seed, a spore, or even a pollen grain: any entity that can carry genetic material into or out of a population. Depending on what the entity happens to be, the point in the generation cycle at which the migration occurs can vary. Juvenile or adult individuals may immigrate, altering adult (or post-zygotic) genotypic frequencies, fertilized eggs of aquatic invertebrates may float into a distant region, altering zygotic genotypic frequencies, or pollen may blow into a population, altering gametic frequencies. Nevertheless, gene flow into or out of populations, regardless of the life-stage at which it actually occurs, can be treated as altering the allele frequency during

Gene flow into or out of populations, regardless of the life-stage at which it actually occurs, can be treated as altering the allele frequency during the breeding adult – gamete pool transition.

Even if migration occurs during the zygotic or adult stage and, therefore, alters the genotypic frequencies, ultimately it is not the genotypes that are transmitted to the next generation but haploid genomes.

the breeding adult–gamete pool transition. Even if migration occurs during the zygotic or adult stage and, therefore, alters the genotypic frequencies, ultimately it is not the genotypes that are transmitted to the next generation but haploid genomes (alleles if we consider a single locus). So the net result is still an alteration in the allele frequency at the gamete pool stage, and that is how we shall model the process.

Imagine a number of large populations of equal size that exchange migrants among themselves each generation, such that in any generation a fraction m of the breeding adults in any population consists of immigrants. You can visualize this as a situation where, prior to breeding, each population contributes a fraction m of its adults to a migrant pool from which an identical number of adults is then returned to it at random (note that this implies some individuals could return back to their native population). We consider a one-locus, biallelic system, with all other assumptions of the ideal large population holding good, and allow genotypic and allelic frequencies to differ among the populations. Now consider any one population in this assemblage, that prior to migration has genotypic frequencies P_{ij} and, therefore, allele frequencies $p_i = \sum_j P_{ij}$. The genotypic frequencies in the zygotes formed after breeding will depend on the allele frequencies p'_i after migration (according to the relationship $P'_{ij} = p'_i p'_j$), rather than on the pre-migration genotypic frequencies P_{ij}, which can, consequently, be ignored. What we need to focus on is, thus, the change in allele frequencies from the pre-migration adults to the post-migration adults, which is for all practical purposes the same thing as the transition from breeding adults to gametes (*Figure* 2).

The recursion for migration is the same as that for mutation with $\overline{p}_1$ equivalent to $v/(u+v)$, and m equivalent to $(u+v)$.

Now, all A_1 alleles in the gamete pool must have arisen in one of two ways: they can be A_1 alleles from the individuals that did not migrate (probability $1 - m$), or

they can be A_1 alleles brought into the population by migrants (probability m). Among migrants, the frequency of A_1 alleles will be $\bar{p}_1$, the mean frequency of A_1 in pre-migration adults across all the populations. Therefore, the frequency of A_1 alleles in the gamete pool of the particular population we are discussing can be related to the frequencies in the pre-migration adults as follows:

$$p_1' = p_1(1-m) + \bar{p}_1 m.$$

The change in allele frequency due to migration is, therefore, given by

$$\Delta p_1 = p_1' - p_1 = m(\bar{p}_1 - p_1).$$

Note that $\bar{p}_1$ is a constant, because no alleles are mutating or going in or out of the assemblage as a whole; they are just getting reallocated among the populations that make up the assemblage. Hopefully, by now this is beginning to look a bit familiar. In fact, the recursion for migration is the same as that for mutation with $\bar{p}_1$ equivalent to $v/(u+v)$, and m equivalent to $(u+v)$. Basically, under both migration and mutation, $p_1(t+1)$ is a linear function of $p_1(t)$. Given this similarity between the recursions for mutation and migration, it can fairly easily be shown that

$$p_1(t) = \bar{p}_1 + [p_1(0) - \bar{p}_1]\,(1-m)^t\,. \qquad (3)$$

Putting the expression for $p_1(t)$ in this form is very useful because we can now apply to the migration case all the conclusions we drew from our analysis of the mutation case. The system has a stable equilibrium at which point all populations will have the same allele frequency $\hat{p}_1 = \bar{p}_1$, and the time taken to approach this equilibrium will depend upon m. A contrast to the case of mutation is that m is often much higher than mutation rates tend to be. Approach to equilibrium can, therefore, be relatively rapid under migration, and migration among populations (even at fairly low levels, e.g., $m = 0.01$) can homogenize populations quite rapidly.

Approach to equilibrium can, therefore, be relatively rapid under migration, and migration among populations (even at fairly low levels, e.g., m = 0.01) can homogenize populations quite rapidly.

Socially enforced patterns of caste or clan endogamy, in the long run, are not likely to be strong enough barriers to keep the endogamous groups genetically distinct.

can homogenize populations quite rapidly. The model we discussed here is formally called the 'Island Model' of migration. There are various other ways in which migration can be modeled, and discussions of these can be found in any population genetics textbook. The main point to remember about migration is that, given time, it is an effective homogenizing force even at fairly low rates. This has implications for human societies where despite social barriers on marriage between some groups (e.g., castes in the Indian context), some amount of gene flow always occurs between groups. The consequence is that socially enforced patterns of caste or clan endogamy, in the long run, are not likely to be strong enough barriers to keep the endogamous groups genetically distinct. Geographical distance, in general, tends to be a more important factor in maintaining the genetic differentiation of groups.

Suggested Reading

[1] **More information on meiotic drive can be found in practically any standard textbook of introductory genetics, e.g., F A Griffiths *et al*, *An Introduction to Genetic Analysis*, 6th Ed. W H Freeman & Co., New York, 1996.**

[2] **A detailed description of the different types of mutations, and the mechanisms by which they arise, can be found in practically any standard textbook of introductory genetics, e.g., Griffiths *et al*, A good summary can also be found in D L Hartl and A G Clark, *Principles of Population Genetics*, 2nd Ed. Sinauer, Sunderland, pp.97–109, 1989.**

[3] **A good summary of the infinite alleles model of mutation, and some of its consequences/related issues can be found in D L Hartl and A G Clark, *Principles of Population Genetics*, 2nd Ed. Sinauer, Sunderland, pp.122–143 and 349–383, 1989. (A slightly more detailed and up-to-date treatment can be found in the 3rd Edition, Sinauer, Sunderland, pp.174–189 and 315–360, 1997). If you are deeply fascinated by this topic, and do not mind fairly elaborate mathematical treatments, you may want to read M Kimura, *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge, 1983.**

[4] **An introductory discussion of other models of migration can be found in D L Hartl and A G Clark, *Principles of Population Genetics*, 2nd Ed. Sinauer, Sunderland, pp.308–322, 1989. The treatment in the 3rd Edition, Sinauer, Sunderland, pp.189–198, 1997, is more cursory.**

Address for Correspondence
Amitabh Joshi
Jawaharlal Nehru Centre for Advanced Scientific Research, Jakkur
Bangalore 560 064, India.
Email: ajoshi@jncasr.ac.in