

Mutational Change in Evolution

Dan Graur, *Tel Aviv University, Tel Aviv, Israel*

Evolution is the process of change in the genetic makeup of populations. Two evolutionary forces are known to cause changes in gene frequencies with time: selection and random genetic drift.

Introduction

The chromosomal or genomic location of a gene is called a locus, and alternative forms of the gene at a locus are called alleles. In a population, more than one allele may be present at a locus, and their relative proportions are referred to as the allele frequencies. The set of all alleles existing in a population at all loci is called the gene pool.

A germ-line mutation (henceforth mutation) is a heritable change in the genetic material in an individual, which results in the generation of a new allele in the population. Thus, a mutation indicates that the genetic information that is transmitted from a parent to its offspring, that is, the genetic information that will be found in the somatic cells of the offspring, is different in some way from the genetic information in the somatic cells of the parent. Since mutations are rare events, it is reasonable to assume that not more than one mutation will be created at a locus at any given time. Thus, in a diploid population, the initial frequency of a new mutation will be $1/2N$, where N is the population size.

A convenient and unambiguous definition of evolution is 'a process of change in the genetic makeup of populations'. In other words, evolution can be viewed as a change in allele frequencies from generation to generation. Significant evolutionary events are those in which the change in allele frequencies is large and rapid: for example a change from an allele frequency of near zero to one in a relatively short time. Given this definition, an understanding of the evolutionary process is achieved through the study of the temporal dynamics of allele frequency changes.

In the following, we review some basic principles of population genetics that are essential for understanding the dynamics of temporal changes in the frequency of a new mutant allele under the effect of various evolutionary forces. In addition, from the long-term point of view, we will discuss the probability that a new mutant variant will completely replace an old one in the population, and we will estimate how fast and how frequent the replacement process will be. Most changes at the DNA level have only a small effect on the phenotype and, consequently, on the fitness of the

Introductory article

Article contents

- Introduction
- Changes in Allele Frequencies
- Natural Selection
- Random Genetic Drift
- Effective Population Size
- Gene Substitution
- Driving Forces in Evolution

organism, and so the frequencies of molecular variants are subject to strong chance effects. Therefore, the element of chance should be taken into consideration when dealing with molecular evolution.

Changes in Allele Frequencies

For a new mutation to become significant from an evolutionary point of view, it must increase its frequency and ultimately become fixed in the population (i.e. all the individuals in a subsequent generation will share the same mutant allele). Thus, in evolution, a mutation arising in an individual is not important *per se* even if it is passed on to the next generation. For a mutant allele to increase its frequency, factors other than mutation must come into play.

The major factors affecting the frequency of alleles in populations are natural selection and random genetic drift. To understand the process of evolution, we must study how these factors govern the changes of allele frequencies. In classical evolutionary studies involving morphological traits, natural selection has been considered as the major driving force in evolution. In contrast, random genetic drift is thought to have played a predominant role in evolution at the molecular level.

There are two mathematical approaches to studying genetic changes in populations. Deterministic models assume that changes in the frequencies of alleles in a population from generation to generation occur in a unique manner and can be unambiguously predicted from knowledge of initial conditions. In strict mathematical terms, this approach is valid only when the population is infinite in size and the environment either remains constant with time or changes according to deterministic rules. These conditions are obviously never met in nature and, therefore, a purely

deterministic approach may not be sufficient to describe the temporal changes in allele frequencies in populations. Random or unpredictable fluctuations in allele frequencies must also be taken into account. Stochastic models assume that changes in allele frequencies occur in a probabilistic manner. That is, from knowledge of the conditions in one generation one cannot predict unambiguously the allele frequencies in the next generation, but can only determine the probabilities with which certain genetic outcomes are likely to be attained. Obviously, stochastic models are preferable to deterministic ones, since they are based on more realistic assumptions. However, deterministic models are simpler and, under certain circumstances, they yield sufficiently accurate approximations.

Natural Selection

Natural selection is defined as the differential reproduction of genetically distinct individuals (genotypes) within a population. Differential reproduction is caused by differences among individuals in such traits as mortality, fertility, fecundity, mating success and the viability of the offspring. Natural selection is predicted on the availability of genetic variation among individuals in characters related to reproductive success. When a population consists of individuals that do not differ from one another in such traits, it is not subject to natural selection. Accordingly, without mutation, which introduces new variation into the population, there can be no selection.

Selection may lead to changes in allele frequencies over time. However, a mere change in allele frequencies from generation to generation does not necessarily indicate that selection is at work. Other processes, such as random genetic drift, can bring about temporal changes in allele frequencies as well. Interestingly, the opposite is also true, that is, a lack of change in allele frequencies does not necessarily indicate that selection is absent.

The fitness of a genotype (denoted as w) is a measure of the individual's ability to survive and reproduce. However, since the size of a population is usually constrained by the carrying capacity of the environment in which the population resides, the evolutionary success of an individual is determined not by its absolute fitness, but by its relative fitness in comparison to the other genotypes in the population. In nature, the fitness of a genotype is not expected to remain constant for all generations and under all environmental circumstances. However, by assigning a constant value of fitness to each genotype, we are able to formulate simple models that are useful for understanding the dynamics of change in the genetic structure of a population brought about by natural

selection. In the simplest class of models, we assume that the fitness of the organism is determined solely by its genetic makeup. We also assume that all loci contribute independently to the fitness of the individual (i.e. the different loci do not interact with one another in a manner that affects fitness), so that each locus can be dealt with separately.

A mutation may or may not affect the fitness of the organism that carries it. In haploid organisms, mutations may have three types of effects on fitness. Most new mutations reduce the fitness of their carriers, that is, they are deleterious and will be selected against, so that the vast majority of them will eventually be removed from the population. This type of selection is called purifying selection. Occasionally, a new mutation may be as fit as the best allele in the population. Such a mutation is selectively neutral, and its fate is not determined by selection. In exceedingly rare cases, a mutation that confers a selective advantage on its carriers may arise. Such a mutation is called advantageous, and it will be subjected to positive selection. The fitness in diploid organisms is determined by the interaction between the two alleles at a locus.

In the following, we consider the diploid case of one locus with two alleles, A_1 and A_2 . A_1 is assumed to be the original or 'old' allele in the population. Thus, the initial population consists of only one genotype, A_1A_1 . In particular, we are interested in the dynamics of change in allele frequencies following the appearance of a new mutant allele, A_2 , and the consequent creation of two new genotypes, A_1A_2 and A_2A_2 .

There are three possible diploid genotypes: A_1A_1 , A_1A_2 and A_2A_2 , and their fitnesses are denoted by w_{11} , w_{12} and w_{22} respectively. Given that the frequency of allele A_1 in a population is p , and the frequency of the complementary allele, A_2 , is $q = 1 - p$, we can show that, under random mating, the frequencies of A_1A_1 , A_1A_2 and A_2A_2 genotypes are p^2 , $2pq$, and q^2 respectively. A population in which such genotypic ratios are maintained is at Hardy–Weinberg equilibrium.

In the general case, the three genotypes are assigned the following fitness values and initial frequencies:

Genotype	A_1A_1	A_1A_2	A_2A_2
Fitness	w_{11}	w_{12}	w_{22}
Frequency	p^2	$2pq$	q^2

Let us now consider the dynamics of allele frequency changes following selection. The relative contribution of each genotype to the subsequent generation is the product of its initial frequency and its fitness. Given the frequencies of the three genotypes and their fitnesses as above, the relative contributions of the three genotypes to the next generation will be p^2w_{11} , $2pqw_{12}$ and q^2w_{22} for A_1A_1 , A_1A_2 and A_2A_2

respectively. Since half of the alleles carried by A_1A_2 individuals and all of the alleles carried by A_2A_2 individuals are A_2 , the frequency of allele A_2 in the next generation (q_{t+1}) will become

$$q_{t+1} = \frac{pqw_{12} + q^2w_{22}}{p^2w_{11} + 2pqw_{12} + q^2w_{22}} \quad (1)$$

The extent of change in the frequency of allele A_2 per generation is denoted as Δq . We can show that

$$\Delta q = q_t + 1 - q_t \frac{pq[p(w_{12} - w_{11}) + q(w_{22} - w_{12})]}{p^2w_{11} + 2pqw_{12} + q^2w_{22}} \quad (2)$$

For mathematical convenience, we shall assign a relative fitness value of 1 to the A_1A_1 genotype. Thus, eqn [2] reduces to

$$\Delta q = \frac{pq[p(w_{12} - 1) + q(w_{22} - w_{12})]}{p^2 + 2pqw_{12} + q^2w_{22}} \quad (3)$$

Five common modes of interaction will be considered: codominance, complete recessiveness, complete dominance, overdominance and underdominance.

Codominance

In *codominant selection*, the two homozygotes have different fitness values, whereas the fitness of the heterozygote is the mean of the fitnesses of the two homozygous genotypes. The relative fitness values for the three genotypes can be written as

Genotype	A_1A_1	A_1A_2	A_2A_2
Fitness	1	$1 + s$	$1 + 2s$

where s is the difference between the fitness of an A_2 -carrying genotype and the fitness of A_1A_1 . A positive value of s denotes an increase in fitness in comparison with A_1A_1 , while a negative value denotes a decrease in fitness. When $s=0$, the fitness of an A_2 -carrying genotype will be the same as that of A_1A_1 .

From eqn [3], we obtain the following change in the frequency of allele A_2 per generation under codominance:

$$\Delta q = \frac{spq}{1 + 2spq + 2sq^2} \quad (4)$$

By iteration, eqn [4] can be used to compute the frequency of A_2 at any generation. However, through approximation by a differential equation we obtain a convenient solution

$$q_t = \frac{1}{1 + \left(\frac{1 - q_0}{q_0}\right)e^{-st}} \quad (5)$$

where q_0 and q_t are the frequencies of A_2 at generations 0 and t respectively.

In eqn [5], the frequency q_t is expressed as a function of time t . Alternatively, t can be expressed as a function of q as

$$t = \frac{1}{s} \ln \frac{q_t(1 - q_0)}{q_0(1 - q_t)} \quad (6)$$

From this equation, one can calculate the number of generations required for the frequency of A_2 to change from one value (q_0) to another (q_t).

Figure 1a illustrates the increase in the frequency of allele A_2 for $s = 0.01$. We see that codominance always increases the frequency of one allele at the expense of the other, regardless of the relative allele frequencies in the population. Therefore, codominance is a type of directional selection. Note, however, that at low frequencies, selection for a codominant allele is not very efficient (i.e. the change in allele frequencies is slow). The reason is that, at low frequencies, the proportion of A_2 alleles residing in heterozygotes is large. For example, when the frequency of A_2 is 0.5, 50% of the A_2 alleles are carried by heterozygotes, whereas when the frequency of A_2 is 0.01, 99% of all such alleles reside in heterozygotes. Because heterozygotes, which contain both alleles, are subject to weaker selective pressures than are A_2A_2 homozygotes (s versus $2s$), the overall change in allele frequencies at low values of q will be small.

Dominance

In the dominant mode of selection, the two homozygotes have different fitness values, whereas the fitness of the heterozygote is the same as the fitness of one of the two homozygous genotypes. In the following, we distinguish between two cases. In the first case, the new allele, A_2 , is dominant over the old allele, A_1 , and the fitness of the heterozygote, A_1A_2 , is identical to the fitness of the A_2A_2 homozygote. The relative fitness values of the three genotypes are

Genotype	A_1A_1	A_1A_2	A_2A_2
Fitness	1	$1 + s$	$1 + s$

From eqn [2], we obtain the following change in the frequency of the A_2 allele per generation:

$$\Delta q = \frac{p^2q_s}{1 - s - p^2s} \quad (7)$$

In the second case, A_1 is dominant over A_2 , and the fitness of the heterozygote, A_1A_2 , is identical to that of the A_1A_1 homozygote. That is, the new allele, A_2 , is recessive. The relative fitness values are

Genotype	A_1A_1	A_1A_2	A_2A_2
Fitness	1	1	$1 + s$

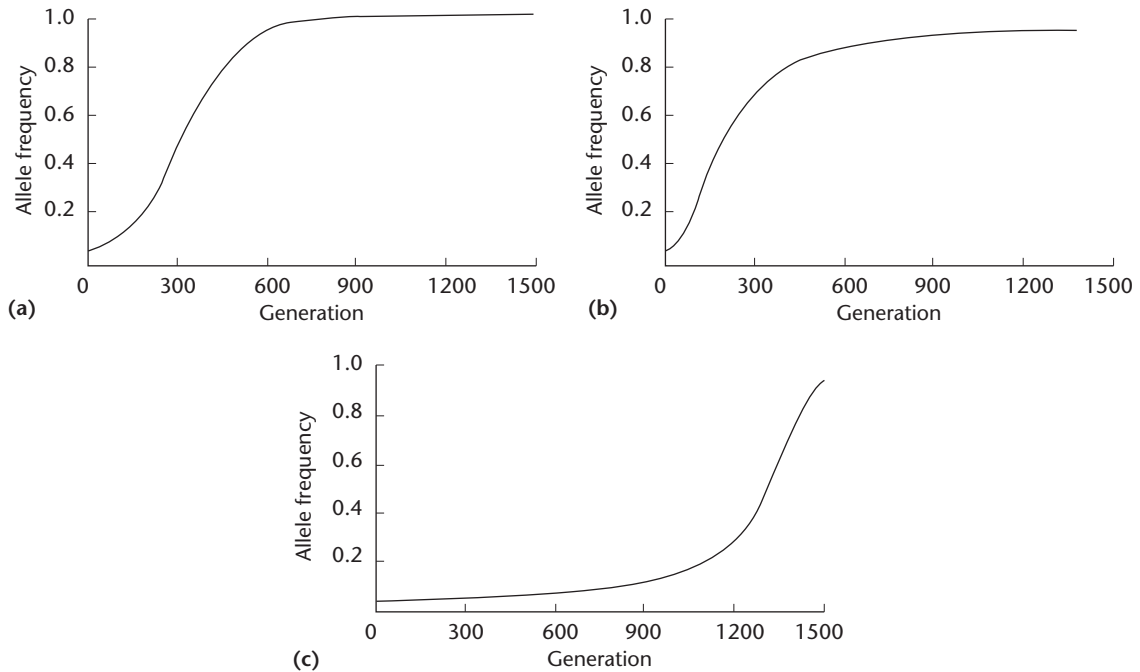


Figure 1 Changes in the frequency of an advantageous allele A_2 under (a) codominant selection, (b) dominant selection where A_2 is dominant over A_1 and (c) dominant selection where A_1 is dominant over A_2 . In all cases, the frequency of A_2 at time 0 is $q = 0.04$, and the selective advantage is $s = 0.01$.

From eqn [2], we obtain the following change in the frequency of A_2 per generation

$$\Delta q = \frac{pq^2s}{1 + q^2s} \quad (8)$$

Figure 1b and **c** illustrates the increase in the frequency of allele A_2 for $s = 0.01$ for dominant and recessive A_2 alleles respectively. Both types of selection are also directional because the frequency of A_2 increases at the expense of A_1 regardless of the allele frequencies in the population.

If the new allele is advantageous and dominant over the old one, then selection is very efficient and the frequency of A_2 increases rapidly. If the new dominant allele is deleterious, it will be eliminated from the population rapidly. This is the reason why dominant deleterious alleles are seldom seen in populations. Dominant pathological genes, such as Huntington disease, are not deleterious from an evolutionary point of view since the age of onset for the disease is in the postreproductive period and, therefore, the fitness of carriers of the Huntington allele is not significantly lower than that of individuals that do not carry the allele.

We note that the less frequent allele A_2 is, the larger the proportion of A_2 alleles in a heterozygous state will be. Thus, as far as selection against recessive homozygotes is concerned, at low frequencies only very few

alleles in each generation will be subject to selection and as a consequence their frequency in the population will not change much. In other words, selection will not be very efficient in increasing the frequency of beneficial alleles or reducing the frequency of deleterious ones, as long as their frequency in a population is low. For this reason, it is almost impossible to rid a population of recessive deleterious alleles, even when such alleles are lethal in a homozygous state (i.e. $s = -1$). The persistence of such genetic diseases as Tay–Sachs syndrome and cystic fibrosis in human populations attests to the inefficiency of the directional selection regime at low allele frequencies. It is possible, however, that some recessive deleterious alleles are maintained by heterozygote advantage (see below).

Overdominance and underdominance

In the overdominant mode of selection, the heterozygote has the highest fitness. Thus

Genotype	A_1A_1	A_1A_2	A_2A_2
Fitness	1	$1 + s$	$1 + t$

In this case, $s > 0$ and $s > t$. Depending on whether the fitness of A_2A_2 is higher than, equal to or lower than that of A_1A_1 , t can be positive, zero or negative

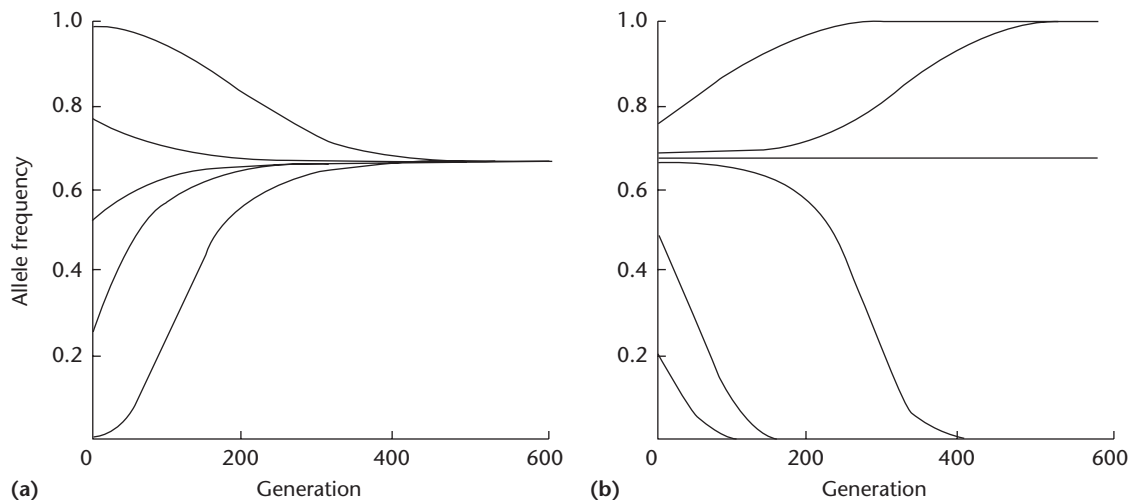


Figure 2 (a) Changes in the frequency of an allele subject to overdominant selection. Initial frequencies from top to bottom curves: 0.99, 0.75, 0.50, 0.25 and 0.01; $s = 0.04$ and $t = 0.02$. Since the s and t values are exceptionally large, the change in allele frequency is rapid. Note that there is a stable equilibrium at $q = 0.667$. (b) Changes in the frequency of an allele subject to underdominant selection. Initial frequencies from top to bottom curves: 0.75, 0.668, 0.667, 0.666, 0.50 and 0.20; $s = -0.02$ and $t = -0.01$. Again, because of the large values of s and t , the change in allele frequency is rapid. Note that there is an equilibrium at $q = 0.667$. This equilibrium, however, is unstable since even the slightest deviation from it will cause one of the alleles to be eliminated from the population.

respectively. The change in allele frequencies is expressed as

$$\Delta q = \frac{pq(2sq - tq - s)}{1 + 2spq + tq^2} \quad (9)$$

Figure 2a illustrates the changes in the frequency of allele A_2 . In contrast to the codominant or dominant selection regimes, in which allele A_1 is eventually eliminated from the population, under overdominant selection the population sooner or later reaches a stable equilibrium in which alleles A_1 and A_2 coexist. The equilibrium is stable because in case of a deviation from equilibrium, selection will quickly restore the equilibrium frequencies. After equilibrium is reached, no further change in allele frequencies will be observed (i.e. $\Delta q = 0$). Thus, overdominant selection belongs to a class of selection regimes called balancing selection.

The frequency of allele A_2 at equilibrium is obtained by solving eqn [9] for $\Delta q = 0$:

$$\hat{q} = \frac{s}{2s - t} \quad (10)$$

When $t = 0$ (i.e. both homozygotes have identical fitness values), the equilibrium frequencies of both alleles will be 50%.

In the underdominant mode of selection, the heterozygote has the lowest fitness, that is, $s < 0$ and $s < t$. As in the case of overdominant selection, here too the change in the frequency of A_2 is described by eqn [9], and the frequency of allele A_2 at equilibrium by eqn [10]. However, in this case, the population reaches

an unstable equilibrium, that is, any deviation from the equilibrium frequencies will drive the allele below the equilibrium frequency to extinction, while the other allele will be fixed (**Figure 2b**). An allele may, thus, be eliminated from the population even if the fitness of its homozygote (A_2A_2) is much higher than that of the prevailing homozygote (A_1A_1). Since a new allele arising in the population by mutation always has a very low frequency (i.e. much below the equilibrium frequency), new underdominant mutants will always be eliminated from the population. Thus, underdominant selection can serve as a powerful illustration of the fallacy of intuitive evolutionary thinking, according to which an allele that increases the fitness of its carrier will always replace one that does not.

Random Genetic Drift

In addition to natural selection, allele frequency changes can also occur by chance. An important factor in producing random fluctuations in allele frequencies is the random sampling of gametes in the process of reproduction (**Figure 3**). Sampling occurs because, in the vast majority of cases in nature, the number of gametes available in any generation is much larger than the number of adult individuals produced in the next generation. In other words, only a minute fraction of gametes succeeds in developing into adults.

To see the effect of sampling, let us consider an idealized situation in which all individuals in the population have the same fitness and selection does

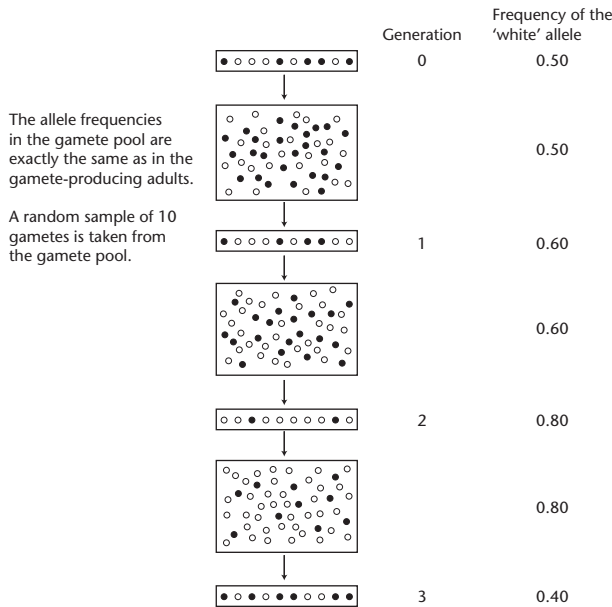


Figure 3 Random sampling of gametes. Allele frequencies in the gamete pools (large boxes) in each generation are assumed to reflect exactly the allele frequencies in the adults of the parental generation (small boxes). Since the population size is finite, allele frequencies fluctuate up and down. (Bodmer WF and Cavalli-Sforza (1976) *Genetics, Evolution, and Man*. San Francisco: Freeman.)

not operate. We further simplify the problem by considering a population with nonoverlapping generations (i.e. a group of individuals that reproduce simultaneously and die immediately afterwards), such that any given generation can be unambiguously distinguished from both previous and subsequent generations. Finally, we assume that the population size does not change from generation to generation. The population under consideration is diploid and consists of N individuals, so that at any given locus, the population contains $2N$ alleles. Let us again consider the simple case of one locus with two alleles, A_1 and A_2 , with frequencies p and $q = 1 - p$ respectively. When $2N$ gametes are sampled from the infinite gamete pool, the probability, P_i , that the sample contains exactly i alleles of type A_1 is given by the binomial probability function:

$$P_i = \frac{(2N)!}{i!(2N-i)!} p^i q^{2N-i} \quad (11)$$

Since P_i is always greater than 0 for populations in which the two alleles coexist (i.e. $0 < p < 1$), the allele frequencies may change from generation to generation without the aid of selection.

The process of change in allele frequency solely due to chance effects is called random genetic drift. One should note, however, that random genetic drift can also be caused by processes other than the sampling of

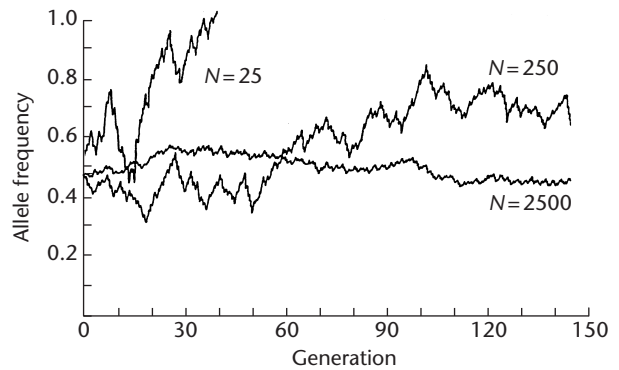


Figure 4 Changes in frequencies of alleles subject to random genetic drift in populations of different sizes. The smallest population reached fixation after 42 generations. The other two populations were still polymorphic after 150 generations but will ultimately reach fixation (allele frequency = 0% or 100%) if the experiment is continued long enough. (Bodmer WF and Cavalli-Sforza (1976) *Genetics, Evolution, and Man*. San Francisco: Freeman.)

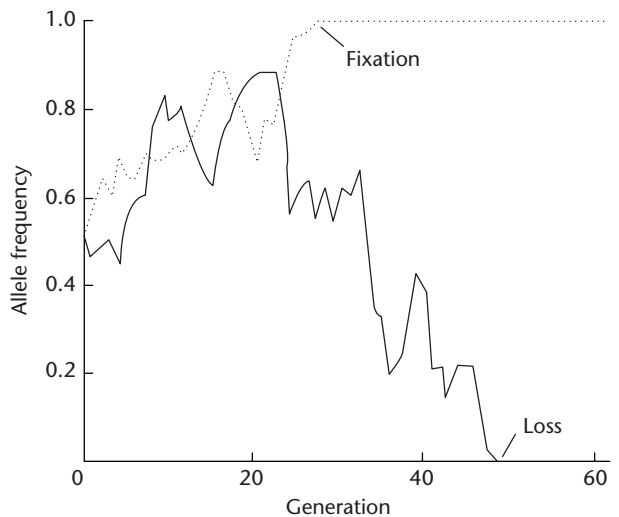


Figure 5 Two possible outcomes of random genetic drift in populations of size 25 and $p_0 = 0.5$.

gametes. For example, stochastic changes in selection intensity can also bring about random changes in allele frequencies.

In **Figure 4**, we illustrate the effects of random sampling on the frequencies of alleles in populations of different sizes. The allele frequencies change from generation to generation, but the direction of the change is random at any point in time. The most obvious feature of random genetic drift is that fluctuations in allele frequencies are much more pronounced in small populations than in larger ones. In **Figure 5** we show two possible outcomes of random genetic drift in a population of size $N = 25$. In one replicate, allele A_1 becomes fixed in generation 27; in the other replicate, allele A_1 is lost in generation 49.

Let us follow the dynamics of change in allele frequencies due to the process of random genetic drift in succeeding generations. The frequencies of allele A_1 are written as $p_0, p_1, p_2, \dots, p_t$, where the subscripts denote the generation number. The initial frequency of allele A_1 is p_0 . In the absence of selection, we expect p_1 to be equal to p_0 , and so on for all subsequent generations. The fact that the population is finite, however, means that p_1 will be equal to p_0 only on the average (i.e. when repeating the sampling process an infinite number of times). In reality, sampling occurs only once in each generation, and p_1 is usually different from p_0 . In the second generation, the frequency p_2 will no longer depend on p_0 but only on p_1 . Similarly, in the third generation, the frequency p_3 will depend on neither p_0 nor p_1 but only on p_2 . Thus, the most important property of random genetic drift is its cumulative behavior; that is, from generation to generation, the frequency of an allele will tend to deviate more and more from its initial frequency.

In mathematical terms, the mean and variance of the frequency of allele A_1 at generation t , denoted by \bar{p}_t and $V(P_t)$, respectively, are given by

$$\bar{p}_t = p_0 \quad (12)$$

and

$$V(P_t) = p_0(1-p_0) \left[1 - \left(1 - \frac{1}{2N} \right)^t \right] \approx p_0(1-p_0) \left[1 - e^{-t/(2N)} \right] \quad (13)$$

where p_0 denotes the initial frequency of A_1 . Note that although the mean frequency does not change with time, the variance increases with time, that is, with each passing generation the allele frequencies will tend to deviate further and further from their initial values; however, the change in allele frequencies will not be systematic in its direction.

Once the frequency of an allele reaches 0 or 1, its frequency will not change in subsequent generations. The first case is referred to as extinction and the second, fixation. If the process of sampling continues for long periods of time, the probability of such an eventuality reaches certainty. Thus, the ultimate result of random genetic drift is the fixation of one allele and the loss of all the others. This will happen unless there is a constant input of alleles into the population by mutation, or unless polymorphism is actively maintained by balancing selection.

Effective Population Size

A basic parameter in population biology is the population size, N , defined as the total number of individuals in a population. From the point of view of

population genetics and evolution, however, the relevant number of individuals to be considered consists of only those individuals that actively participate in the reproductive process. Since not all individuals take part in reproduction, the population size that matters in the evolutionary process is different from the census size. This part is called the effective population size and is denoted by N_e . This concept, introduced by Wright in 1931, was rigorously defined as the size of an idealized population that would have the same effect of random sampling on allele frequencies as that of the actual population.

Consider, for instance, a population with a census size N , and assume that the frequency of allele A_1 at generation t is p . If the number of individuals taking part in reproduction is N , then the variance of the frequency of allele A_1 in the next generation, p_{t+1} , may be obtained from eqn [13] by setting $t = 1$:

$$V(p_{t+1}) = \frac{p_t(1-p_t)}{2N} \quad (14)$$

In practice, since not all individuals in the population take part in the reproductive process, the variance will be larger than that obtained from eqn [14]. The effective population size is the value that is substituted for N in order to satisfy eqn [14],

$$V(p_{t+1}) = \frac{p_t(1-p_t)}{2N_e} \quad (15)$$

In general, N_e is smaller, sometimes much smaller, than N .

Gene Substitution

Gene substitution is defined as the process whereby a mutant allele completely replaces the predominant or wild-type allele in a population. In this process, a mutant allele arises in a population as a single copy, and becomes fixed after a certain number of generations. Not all mutants, however, reach fixation. In fact, the majority of them are lost after a few generations. Thus, we need to address the issue of fixation probability and discuss the factors affecting the chance that a new mutant allele will reach fixation in a population. The time it takes for a new allele to become fixed is called the fixation time. In the following we shall identify the factors that affect the rapidity with which a new mutant allele replaces an old one in a population. New mutations arise continuously within populations. Consequently, gene substitutions occur in succession, with one allele replacing another and being itself replaced in time by a newer allele. Thus, we can speak of the rate of gene substitution, that is, the number of fixations of new alleles per unit time.

Fixation probability

The probability that a particular allele will become fixed in a population depends on its frequency, its selective advantage or disadvantage (s), and the effective population size (N_e). In the following, we shall consider codominance and assume that the relative fitnesses of the three genotypes A_1A_1 , A_1A_2 and A_2A_2 are 1, $1 + s$ and $1 + 2s$ respectively.

The probability of fixation of A_2 is given by

$$P = \frac{1 - e^{-4N_e s q}}{1 - e^{-4N_e s}} \quad (16)$$

where q is the initial frequency of allele A_2 . Equation [16] reduces to $P \approx q$ as s approaches 0. Thus, for a neutral allele, the fixation probability equals its frequency in the population.

We note that a new mutant arising as a single copy in a diploid population of size N has an initial frequency of $1/(2N)$. The probability of fixation of an individual mutant allele, P , is thus obtained by replacing q with $1/(2N)$ in eqn [16]. When $s \neq 0$,

$$P = \frac{1 - e^{-(2N_e s)/N}}{1 - e^{-4N_e s}} \quad (17)$$

For a neutral mutation, that is, $s = 0$, eqn [17] becomes

$$P = \frac{1}{2N} \quad (18)$$

If the population size is equal to the effective population size, eqn [17] reduces to

$$P = \frac{1 - e^{-2s}}{1 - e^{-4N_e s}} \quad (19)$$

If the absolute value of s is small, we obtain

$$P = \frac{2s}{1 - e^{-4N_e s}} \quad (20)$$

For positive values of s and large values of N , eqn [20] reduces to

$$P \approx 2s \quad (21)$$

Thus, if an advantageous mutation arises in a large population and its selective advantage over the rest of the alleles is small, the probability of its fixation is approximately twice its selective advantage. This result is quite noteworthy, since it essentially indicates that an advantageous mutation does not always become fixed in the population. Thus, the perception of adaptive evolution as a process in which advantageous mutations arise in populations and invariably take over the population in subsequent generations is a naïve concept. Moreover, even deleterious mutations have a finite probability of becoming fixed in a population, albeit a small one. However, the mere fact that a deleterious allele may become fixed in a

population at the expense of better alleles illustrates in a powerful way the importance of chance events in determining the fate of mutations during evolution.

As the population size becomes larger, the chance effects become smaller.

Fixation time

The time required for the fixation or loss of an allele depends on the frequency of the allele, its selective advantage or disadvantage, and the size of the population. The mean time to fixation or loss becomes shorter as the frequency of the allele approaches 1 or 0 respectively.

When dealing with new mutations, it is more convenient to treat fixation and loss separately. In the following, we shall deal with the mean fixation time of those mutants that will eventually become fixed in the population. This variable is called the conditional fixation time. In the case of a new mutation whose initial frequency in a diploid population is by definition $q = 1/(2N)$, the mean conditional fixation time was calculated. For a neutral mutation, it is approximated by

$$\bar{t} = 4N \text{ generations} \quad (22)$$

and for a mutation with a selective advantage of s or a selective disadvantage $-s$, it is approximated by

$$\bar{t} = (2/s) \ln(2N) \text{ generations} \quad (23)$$

Figure 6 presents in a schematic manner the dynamics of gene substitution for advantageous and neutral mutations. We note that advantageous mutations are either rapidly lost or rapidly fixed in the population. In contrast, the frequency changes for neutral alleles are slow, and the fixation time is much longer than that for advantageous mutants.

Rate of gene substitution

The rate of gene substitution is defined as the number of mutants reaching fixation per unit time. We shall first consider neutral mutations. If neutral mutations occur at a rate of u per gene per generation, then the number of mutants arising at a locus in a diploid population of size N is $2Nu$ per generation. Since the probability of fixation for each of these mutations is $1/(2N)$, we obtain the rate of substitution of neutral alleles by multiplying the total number of mutations by the probability of their fixation:

$$K = 2Nu \left(\frac{1}{2N} \right) = u \quad (24)$$

Thus, for neutral mutations, the rate of substitution is equal to the rate of mutation, a remarkably simple and important result. This result can be intuitively

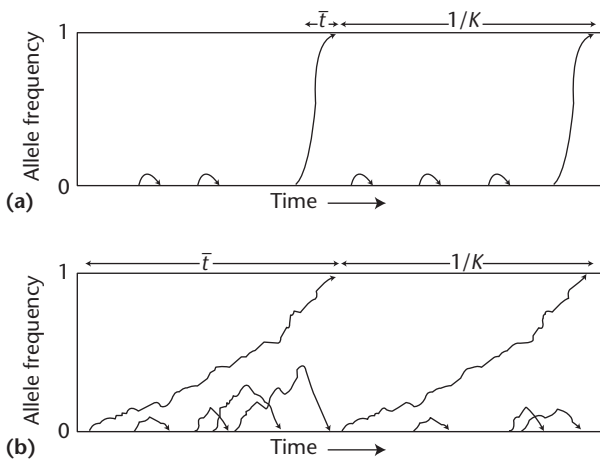


Figure 6 Schematic representation of the dynamics of gene substitution for (a) advantageous and (b) neutral mutations. Advantageous mutations are either quickly lost from the population or quickly fixed, so that their contribution to genetic polymorphism is small. The frequency of neutral alleles, on the other hand, changes very slowly in comparison, so that a large amount of transient polymorphism is generated, \bar{t} is the conditional fixation time, and $1/K$ is the mean time between two consecutive fixation events. (Nei M (1987) *Molecular Evolutionary Genetics*. New York: Columbia University Press.)

understood by noting that, in a large population, the number of mutations arising every generation is high but the fixation probability of each mutation is low. In comparison, in a small population, the number of mutations arising every generation is low, but the fixation probability of each mutation is high. As a consequence, the rate of substitution for neutral mutations is independent of population size.

For advantageous mutations, the rate of substitution can also be obtained by multiplying the rate of mutation by the probability of fixation for advantageous alleles as given in eqn [21]. For codominance with $s > 0$, we obtain

$$K = 4Nsu \quad (25)$$

In other words, the rate of substitution for the case of codominance depends on the population size (N) and the selective advantage (s), as well as on the rate of mutation (u).

Driving Forces in Evolution

Evolutionary explanations can be broadly classified into three types according to the relative importance assigned to random genetic drift versus the various forms of selection in determining a particular evolutionary outcome. Mutationist hypotheses are those

theories in which an evolutionary phenomenon is mainly explained by the effects of mutational input and random genetic drift. Neutralist hypotheses explain evolutionary phenomena by stressing the effects of mutation, random genetic drift and purifying selection. Selectionist explanations emphasize the effects of the advantageous and balancing modes of selection as the main driving forces in the evolutionary process. The above distinction is a useful framework for understanding some of the most important controversies in the history of molecular evolution.

According to the selectionist perception of the evolutionary process, gene substitutions occur as a consequence of selection for advantageous mutations. Polymorphism, on the other hand, is maintained by balancing selection. Thus, selectionists regard substitution and polymorphism as two separate phenomena driven by different evolutionary forces. Gene substitution is the end result of a positive adaptive process whereby a new allele takes over future generations of the population if and only if it improves the fitness of the organism, while polymorphism is maintained when the coexistence of two or more alleles at a locus is advantageous for the organism or the population (e.g. overdominant selection).

In contrast to selectionism, the neutral theory of molecular evolution contends that at the molecular level the majority of evolutionary changes and much of the variability within species are caused neither by positive selection of advantageous alleles nor by balancing selection, but by random genetic drift of mutant alleles that are selectively neutral or nearly so. Neutrality, in the sense of the theory, does not imply strict equality in fitness for all alleles. It only means that the fate of alleles is determined largely by random genetic drift. In other words, selection may operate, but its intensity is too weak to offset the influences of chance effects. For this to be true, the absolute value of the selective advantage or disadvantage of an allele, $|s|$, must be smaller than $1/(2N_e)$, where N_e is the effective population size.

According to the neutral theory, the frequency of alleles is determined by purely stochastic rules, and the picture that we obtain at any given time is merely a transient state representing a temporary frame from an ongoing dynamic process. Consequently, polymorphic loci consist of alleles that are either on their way to fixation or are about to become extinct. Viewed from this perspective, all molecular manifestations that are relevant to the evolutionary process should be regarded as the result of a continuous process of mutational input and a concomitant random extinction or fixation of alleles. Thus, the neutral theory regards substitution and polymorphism as two facets of the same phenomenon. Substitution is a long and gradual process whereby the frequencies of mutant

alleles increase or decrease randomly, until the alleles are ultimately fixed or lost by chance. At any given time, some loci will possess alleles at frequencies that are neither 0% nor 100%. These are the polymorphic loci. According to the neutral theory, most genetic polymorphism in populations is unstable and transient, that is, allele frequencies fluctuate with time and the alleles are replaced continuously.

Interestingly, the neutral theory, even in its strictest form, does not preclude adaptation. According to some, a population that is free from selection can accumulate many polymorphic neutral alleles. Then, if a change in ecological circumstances occurs, some of the neutral alleles will no longer be neutral but deleterious, against which purifying selection may operate. After these alleles are removed, the population will become more adapted to its new circumstances than before. Thus, theoretically at least, adaptive evolution may occur without adaptive selection.

The essence of the dispute between neutralists and selectionists basically concerns the distribution of fitness values of mutant alleles. Both theories agree that most new mutations are deleterious, and that these mutations are quickly removed from the population so that they contribute neither to the rate of substitution nor to the amount of polymorphism within populations. The difference concerns the relative proportion of neutral mutations among nondeleterious mutations. While selectionists claim that very few mutations are selectively neutral, neutralists maintain that the vast majority of nondeleterious mutations are effectively neutral.

See also

Darwin and the Idea of Natural Selection
Dominance and Recessivity

Effective Population Size
Evolution: Neutralist View
Evolution: Selectionist View
Fitness and Selection
Fixation Probabilities and Times
Genetic Drift
Genetic Variation: Polymorphisms and Mutations
Kimura, Motoo
Mutation Rate
Population Genetics
Pseudogenes: Patterns of Mutation
Purifying Selection: Action on Silent Sites
Single-base Mutation
Stationary Allele Frequency Distributions

Further Reading

- Gillespie JH (1991) *The Causes of Molecular Evolution*. New York, NY: Oxford University Press.
- Gillespie JH (1998) *Population Genetics: A Concise Guide*. Baltimore, MD: Johns Hopkins University Press.
- Graur D and Li W-H (2000) *Fundamentals of Molecular Evolution*, 2nd edn. Sunderland, MA: Sinauer Associates.
- Hartl DL and Clark AG (1997) *Principles of Population Genetics*, 3rd edn. Sunderland, MA: Sinauer Associates.
- Hedrick PW (1983) *Genetics of Populations*. Portola Valley, California, CA: Science Books International.
- Hey J (1999) The neutralist, the fly and the selectionist. *Trends in Ecological Evolution* **14**: 35–38.
- Kimura M (1962) On the probability of fixation of mutant genes in populations. *Genetics* **47**: 713–719.
- Kimura M (1968) Evolutionary rate at the molecular level. *Nature* **217**: 624–626.
- Kimura M (1968) Genetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles. *Genetic Research* **11**: 247–269.
- Kimura M (1983) *The Neutral Theory of Molecular Evolution*. Cambridge, UK: Cambridge University Press.
- Kimura M and T Ohta (1969) The average number of generations until fixation of a mutant gene in a finite population. *Genetics* **61**: 763–771.
- King JL and Jukes TH (1969) Non-Darwinian evolution. *Science* **164**: 788–798.
- Wright S (1931) Evolution in Mendelian populations. *Genetics* **16**: 97–159.