Population genetics

Contributed by: C. C. Li

Publication year: 2014

The study of both experimental and theoretical consequences of mendelian heredity on the population level, in contradistinction to classical genetics which deals with the offspring of specified parents on the familial level. The genetics of populations studies the frequencies of genes, genotypes, and phenotypes, and the mating systems. It also studies the forces that may alter the genetic composition of a population in time, such as recurrent mutation, migration, and intermixture between groups, selection resulting from genotypic differential fertility, and the random changes incurred by the sampling process in reproduction from generation to generation. This type of study contributes to an understanding of the elementary step in biological evolution. The principles of population genetics may be applied to plants and to other animals as well as humans. *See also:* MENDELISM.

Mendelian populations

A mendelian population is a group of individuals who interbreed among themselves according to a certain system of mating and form more or less a breeding community. These individuals share a common gene pool which is the total genic content of the group. A mendelian population is the unit of study in population genetics. The population may be very large or very small, and is to be distinguished from species or varieties, which may consist of numerous isolated or partially isolated mendelian populations. Mendelian population is a genetic rather than a taxonomic term. Mendelian populations differ from each other in their genic content or chromosomal organization, not necessarily in their taxonomic features. The term deme, originally defined as an assemblage of taxonomically closely related individuals, has been used as a synonym for mendelian population. Gamodeme, a deme forming a more or less isolated local intrabreeding community, would be a better substitute.

Mutation pressure

Gene mutation arises from time to time in nature. The causes for mutation are not fully known, and thus it can be said that mutations arise "spontaneously." The effect of a new mutant gene is unpredictable and the gene is therefore said to mutate "at random." One property of mutation has been established: It is recurrent. Each type of gene mutates at a certain rate per generation. The rate is usually low—about 1 mutant in 10^5–10^8 genes of a given sort, varying from locus to locus on the chromosomes, even under uniform conditions. Ionizing radiation, certain chemicals, heat, and some other agents increase the rate of mutation. *See also:* MUTATION.

Let μ be the rate of mutation from an allele A to another form a per generation. If a fraction p of the genes of a population is A in one generation, then in the next generation the frequency of A will be diminished by the amount $p\ \mu$, so that the new frequency of A will be $p(1 - \mu)$. The amount of change, $p\ \mu$, is said to be due to the mutation pressure. If this pressure is unopposed generation after generation, the gene A will gradually disappear from the population, as $p_n = p_0\ (1 - \mu)^n = p_0\ e^{-n\mu}$, where p_0 is the initial gene frequency and p_n is the frequency after n generations. Therefore, for all existing genes there must be some kind of compensating mechanism which supports its continuing presence in nature. One important problem in population genetics is the mechanism of maintenance of a gene in a population or its change in frequency from generation to generation.

If, in addition to the mutation from A to a, there is reverse mutation from a to A at the rate ν per generation, then the net amount of change in the frequency of a is $\Delta q = p\ \mu - q\nu$. At the time when these opposing changes cancel each other, there will be no change in gene frequency despite the recurrent mutations. This state of affairs is said to be in equilibrium and is obtained when $\Delta q = 0$; that is, $\hat{p} = \nu/(\mu + \nu)$ and $\hat{q} = \mu/(\mu + \nu)$, where q is a frequency of a, $\hat{q}$ is the equilibrium point for a, and $\hat{p}$ is the equilibrium point for A. The equilibrium gene frequencies are determined by the opposing rates of mutation only and are independent of the initial frequencies of the genes in the population. The amount of change in gene frequency per generation is larger when the current q is far away from the equilibrium $\hat{q}$ than when q is close to $\hat{q}$. Substitution gives $\Delta q = -(\mu + \nu)(q - \hat{q})$, indicating that the amount of change per generation is proportional to the deviation $(q - \hat{q})$. It also shows that if $q > \hat{q}$, q decreases, and if $q < \hat{q}$, q increases, or that q will approach $\hat{q}$ from either side. Such an equilibrium is said to be stable. The changes in q described above are independent of the mating system practiced in the population.

In nature, and under artificial conditions, the mutation rates may not remain constant in all generations but may fluctuate within a certain range from time to time. In such cases, instead of a single fixed equilibrium point $\hat{q}$, there will be an equilibrium distribution of q within a certain range, and the apparent change in gene frequency from one generation to the next may be purely a stochastic phenomenon without necessarily having long-term significance. The same remark applies to all equilibria to be established in subsequent paragraphs. *See also:* STOCHASTIC PROCESS.

Migration and intermixture

If a fraction m of a population with a gene frequency q consists of immigrants from outside and the immigrant group has a gene frequency $\overline{q}$, the new gene frequency of the population will be $q_1 = (1 - m)q + m\ \overline{q} = q - m(q - \overline{q})$. The amount of change in gene frequency in one generation is thus $\Delta q = q_1 - q = -m(q - \overline{q})$, showing that the change is proportional to the deviation $(q - \overline{q})$. This expression for Δq is of the same form as that for mutation. If the immigrants have the same gene frequency as the population, there will be no change in gene frequency in spite of the migrations. The continued intermixture of neighboring populations will eventually make them homogeneous in terms of gene frequencies. Thus, if a large population is divided into a number of partially isolated subpopulations, migrations between the groups will eventually make all subpopulations have the same gene frequency $\bar{\mathbf{q}}$, which then denotes the average for the entire population in the absence of other

TABLE 1. Decrease in heterozygosis with systematic inbreeding

Inbreeding system	Manner of heterozygosis (*H*) decrease	Limiting situation
Self-fertilization	$H = {}^1/_2 H'$	$H = 0.500H'$
Same purebred sire ×successive daughters	$H = {}^1/_2 H'$	$H = 0.500H'$
Brother × sister	$H = {}^1/_2 H' + {}^1/_4 H''$	$H = 0.809H'$
Younger parent × offspring	$H = {}^1/_2 H' + {}^1/_4 H''$	$H = 0.809H'$
Half brother × half sisters	$H = {}^3/_4 H' + {}^1/_8 H''$	$H = 0.890H'$
Half brother × full sisters	$H = {}^1/_2 H' + {}^1/_4 H'' + {}^1/_{16} H'''$	$H = 0.870H'$
Double first cousins	$H = {}^1/_2 H' + {}^1/_4 H'' + {}^1/_8 H'''$	$H = 0.920H'$

disturbing factors. If the local populations are differentiated genetically, there must be some mechanism (for example, local selection) to counteract the pooling effect of migrations so that an equilibrium condition may be reached. The change in gene frequency due to migration is independent of the mating system practiced in the population.

Mating systems

In a gene pool with respect to one locus, if a proportion p of the genes is A and a proportion q of the genes is a, the genotypic proportions in the population are still unknown until the mating pattern is specified. The mating pattern is a system by which the genes are associated into pairs to form the diploid genotypes. The mating systems vary widely in nature for different organisms and populations. Thus, wheat may have 1% cross-pollination and 99% self-fertilization, whereas maize practices just the reverse. One of the simplest and most extensively studied systems is random mating, also known as panmixis.

Panmixis. Random mating between individuals is equivalent to a random union of gametes. Thus, if the (pA, ga) gametes of one sex unite at random with the (pA, qa) gametes of the opposite sex, the resulting genotypic array will be p^2 AA, $2pqAa$, q^2 aa. This principle was first discovered by W. E. Castle in 1903. These genotypic proportions will be realized only in very large populations. *See also:* HARDY-WEINBERG FORMULA; HUMAN GENETICS.

Inbreeding. Inbreeding refers to mating between genetically related individuals; the frequency with which two A gametes unite will be greater than p^2; and a similar situation is true for gene a. Consequently, inbreeding leads to an increase of homozygosis at the expense of heterozygosis. Let H be the heterozygosis proportion in a population, H' that in the preceding generation, H'' that two generations ago, and so on. On continued systematic inbreeding, the manner in which the value of H decreases is shown in Table 1.

Continued close inbreeding, such as those degrees indicated in Table 1 and many others, eventually leads to complete homozygosis. The population will then consist of *pAA* and *q aa*; that is, an *A* gamete will always unite with another *A* gamete, and an *a* gamete with another *a.* Inbreeding between remote relatives does not necessarily lead to complete homozygosis but only decreases the heterozygosis below the random mating level to a certain extent.

The inbreeding coefficient is an index intended to measure the amount or degree of inbreeding that has been accomplished in a population. Various indices may be constructed. One that has been proved highly useful in both theoretical investigation and practical breeding work is the inbreeding coefficient *F* defined as the correlation coefficient between the uniting gametes. The value of *F* ranges from 0 for random mating to 1 for inbreeding in a homozygous populations, as shown below.

Random mating:

	A	*a*	
A	p^2	pq	p
a	pq	q^2	q
	p	q	1

Inbred population:

	A	a	
A	$p^2 + Fpq$	$pq - Fpq$	p
a	$pq - Fpq$	$q^2 + Fpq$	q
	p	q	1

Correlation = F

Inbreeding in homozygous population:

	A	a	
A	p	0	p
a	0	q	q
	p	q	1

Correlation = 1

In an inbred population where the correlation between the uniting gametes is F, the genotypic array in the population will be as in Eqs. (1)–(3).

$$AAp^2 + Fpq = (1 - F)p^2 + Fp \tag{1}$$

$$Aa2pq - 2Fpq = 2(1 - F)pq \tag{2}$$

$$aaq^2 + Fpq = (1 - F)q^2 + Fq \tag{3}$$

These equations show that the population may be mathematically considered as having two separate components, $(1 - F)$ panmictic and F fixed. If the mating system is such that F remains constant (instead of increasing) from generation to generation, the population will reach an equilibrium state with the genotypic array shown above.

The correlation between uniting gametes is due to the correlation between mating individuals. In an equilibrium population, if M denotes the correlation between mates, then $M = 2F/(1 + F)$ or $F = M/(2 - M)$.

Genotype selective values

Within a large population not all individuals produce the same number of offspring. In the situation to be considered, the average number of living offspring born to each of the genotypes in the population is studied, while the random fluctuation in the number of offspring from family to family is ignored. Furthermore, it is assumed that the population is so large that only the relative frequencies of the various genotypes and genes in the population are of interest. Suppose that the average number of offspring for each genotype is as follows: *AA*: 2.00; *Aa*: 2.50; *aa*: 1.50. Given these differential rates of reproduction, the new gene frequency of the next generation may be calculated. Inasmuch as it is only their relative magnitude that matters, these reproductive rates may be simplified into the ratio $W_{11}:W_{12}:W_{22} = 1:1.25:0.75$ or alternatively into 0.80:1:0.60. In order to standardize the description, it is convenient to take one of the three reproductive values as unity. In the previous example, depending upon whether the reproductive value of *AA* or *Aa* is taken as unity (the standard), that of *aa* is $0.75 = 1 - 0.25$ or $0.60 = 1 - 0.40 = 1 - s$ in general. The value of s is known as the selection coefficient against the genotype *aa*. When a selection coefficient is used, it should always be stated which genotype has been employed as the standard.

Natural selection. The doctrine of the survival of the fittest needs clarification from the genetic viewpoint. The relative genotypic reproductive values (W_{11}, W_{12}, W_{22} of the preceding paragraph) simply given an ex post facto description, by which the genetic composition of the offspring generation may be related to that of the parent generation. These W values include all causes for differential reproduction such as fecundity, fertility, sexual maturity and capacity, survival and viability, length of reproductive life, and many others depending on the details of the life cycle of the organism. The W value, sometimes briefly referred to as relative fitness, is not necessarily correlated with any observable morphological characteristics, no matter how desirable they may seem to humans. From the genetic viewpoint, only those who reproduce count. Thus, natural selection has no particular purpose except to perpetuate those who are fit to reproduce under the given conditions. Only when a characteristic lowers the organism's reproductive capacity does it have a genetic effect on the subsequent generations.

Selection pressure and equilibrium. The effect of selection may be described in terms of changes in gene frequency. In a random mating population with respect to one gene locus, the situation is as shown in Table 2.

The population after selection is the parental population of the next generation through random mating. The value $\bar{W}$ is the total of the selected parental population, but may also be regarded as the average fitness of the original unselected population. The new frequency of gene a among the selected is $q' = (pqW_{12} + q^2 W_{22})\ \bar{W}$

TABLE 2. Effect of selection in a random mating population

Genotype	Frequency, *f*	Fitness, *W*	Frequency after selection, *fW*
AA	p^2	W_{11}	$p^2 W_{11}$
Aa	$2pq$	W_{12}	$2pq W_{12}$
aa	q^2	W_{22}	$q^2 W_{22}$
Total	1.00		$\bar{W}$

and therefore the amount of change per generations is $\Delta q = q' - q$, or more explicitly, Eq. (4).

$$\Delta q = \frac{pq}{2\overline{W}} - \frac{d\overline{W}}{dq} \tag{4}$$

This represents the effect of selection pressure on gene frequency. When p or q is zero, there is no change in gene frequency; there can be no selection in the absence of alternatives. Therefore, all selection effects involve the factor pq. Further, when p or q is very small, the selection is ineffective whether it is for or against a gene. Besides these terminal conditions, if there exists a q value such that $\Delta q = 0$, it is called the equilibrium value of gene frequency, because a population with that particular gene frequency will remain unchanged in spite of the selection pressure. When such a q value exists, it must be the solution of the relations $d\bar{W}/dq = 0$ shown in Eq. (5).

$$\hat{q} = \frac{W_{11} - W_{12}}{(W_{11} - W_{12}) + (W_{22} - W_{12})} \tag{5}$$

In order that $\hat{q}$ be a positive fraction, the differences $W_{11} - W_{12}$ and $W_{22} - W_{12}$ must be both positive or both negative; that is, the selective value of the heterozygote must be lower or higher than those of both homozygotes. For all other cases, there will be no equilibrium except when $q = 0$ or 1.

Stability of an equilibrium. The value of $\bar{W} = p^2 W_{11} + 2pqW_{12} + q^2 W_{22}$ may be plotted against the value of q or p. When W_{12} is greater than W_{11} and W_{22}, The $\bar{W}$ curve has a maximum point (**Fig. 1**). The q value corresponding to the maximum value of $\bar{W}$ is the stable equilibrium point. This means that whether it is smaller or larger than $\hat{q}$, the q value will approach $\hat{q}$ as selection proceeds from generation to generation. A stable equilibrium of this type leads to balanced genetic polymorphism, that is, to the coexistence of alleles in a population. Conversely, if q_{11} is lower than both W_{11} and W_{22}, the $\bar{W}$ curve has a minimum point yielding an unstable equilibrium; the selection pressure will make the q value move away from the equilibrium value toward either 0 or 1, depending upon which side of the equilibrium the q happens to be. Consequently, selection against the heterozygote leads to the

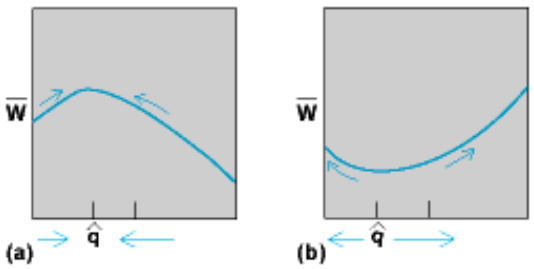

Fig. 1 Diagram of the relationship between the average fitness value $\bar{W}$ and $\hat{q}$, the equilibrium point for *a.* (*a*) Stable equilibrium. (*b*) Unstable equilibrium.

elimination of one of the alleles. In more complicated situations, there could be more than one stable or unstable equilibrium value in a population, or both.

The genotype selective values W_{11}, W_{12}, W_{22} have been assumed to be fixed for each genotype, but in nature they may vary in a number of ways. In addition to the omnipresent random fluctuations, the selective values may vary with the gene frequency itself. For instance, a genotype favored by selection when it is rare in the population may suffer a disadvantage when it is too common. For such cases, there will be an equilibrium yielding genetic polymorphism. Let U_{11}, a function of q, be the varying selective value of genotype AA, and so on. Then the equilibrium value of gene frequency is given by the appropriate solution of Eq. (6).

$$q = \frac{U_{11} - U_{12}}{(U_{11} - U_{12}) + (U_{22} - U_{12})} \tag{6}$$

The study of selection effects may be extended to cases with multiple alleles, sex-linked alleles, autopolyploids, and inbreeding populations.

Gamete selection. The effective rate at which the A and a gametes function may not be the same; that is, selection may operate in the gametic stage instead of in the diploid genotypic stage. If the selective actions for the genotype aa, for example, and gamete a are in opposite directions, an equilibrium may result.

Balance between selection and mutation

There are many different types of genotypic selection. Two simple cases will illustrate the principle of balance between selection and mutation pressures.

Selection against recessives. Suppose that the selective values of AA, Aa, and aa are 1, 1, and $1 - s$, where s is a positive fraction known as the selection coefficient. Then the new gene frequency will be $q' = (q - sq^2)/(1 - q^2)$ in the next generation, so that the amount of change per generation is $q' - q = -sq^2(1 - q)/(1 - sq^2)$. At the

same time, if μ is the mutation rate from A to a, the value of q will be increased by the amount $\mu(1 - q)$ per generation. At equilibrium, the forces to increase and to decrease the gene frequency must cancel each other; that is, $\mu(1 - q) = sq^2(1 - q)/(1 - sq^2)$. By solving, $sq^2 = \mu/(1 + \mu)$ is obtained, which closely approximates μ. Hence, $a^2 = \mu/s$ and $q = \sqrt{\mu/s}$, which usually is a small quantity. When aa is lethal or unable to reproduce, $s = 1$ and $q = \sqrt{\mu}$. This explains the persistence of deleterious recessive genes in a population in spite of continuous selection.

Selection against dominants. If the selection is against homozygous dominants only, the situation is the same as in the previous instance except for substitution of p for q and ν (mutation rate from a to A) for μ. To bring out the distinction between selection against dominants and that against recessives, take the extreme case in which AA is lethal, the selective value of Aa is $1 - s$, and aa is the norm. The value of p will then be so low that the usual genotypic proportions p^2, $2pq$, q^2 will take the limiting form 0, $2p$, $1 - 2p$, as q is very close to unity. The increase in q through selection is approximately sp, whereas the loss through mutation from a to A is $q\nu = \nu$. Hence, at equilibrium, $p = \nu/s$. This value is much lower than $q = \sqrt{\mu/s}$ for selection against recessives. Thus, selection against dominant alleles is more effective than selection of the same intensity against recessives. Selection against heterozygotes will eventually lead to the same limiting situation. All the equilibrium values supported by mutation pressure are low but stable. Mutations prevent complete extinction of an allele.

Random drift

The random drift of gene frequencies in finite populations is often called the Sewall Wright effect because of his analysis of its significance. The gene frequency of any generation is determined by the uniting gametes produced by the parents of the preceding generation. If the number of parents is limited and constitutes a random sample of the entire population, the gene frequency of the next generation will not remain exactly the same as that of the previous generation but will be subject to a random fluctuation on account of the sampling process. In a random mating population of N individuals, one-half of whom are males and one-half females, and maintaining the same population size, the variance of the gene frequency based on $2N$ gametes is $q(1- q)/2N$. The gene frequency may become a little higher or a little lower in the following generation. The smaller the population, the greater is the variance. This random process will continue to operate in all generations. In a sufficiently long time, the value of q will reach either the terminal value 0 or 1. Hence the random drift leads eventually to complete homozygosis for small populations. It can be shown that the limiting rate of reaching the state 0 or 1 is each $^1/_4$ per generation, so that the total rate of "decay" of genetic variability is $^1/_2N$ per generation. Naturalists have found numerous small isolated colonies (for example, snails in mountain valleys) with characteristics uncorrelated with the environmental conditions to substantiate the theory of random (nonadaptive) fixation.

The effective size of a population is the actual number of individuals producing offspring and thereby responsible for the genetic constitution of the next generation. The random mating population with one-half males and one-half females, and producing the same number of offspring, is an idealized model. Any deviation from the ideal situation will have a different sampling variance and a different rate of decay. Equating these to the

"standard" variance $q(1 - q)/2N$ or the ideal decay rate $^1/_2N$, an equivalent N is obtained for the ideal population. The latter number is known as the effective size of a population. It is convenient to use in mathematical descriptions of the genetic behavior of a population. Some of the factors that tend to make the effective size smaller than the actual breeding size are given below.

1. Unequal number of males and females. If M and F are the respective numbers, the effective size N_e is not simply $M + F$ but is defined by $1/N_e = {}^1/_4M + {}^1/_4F$ and is equal to $N_e = 4MF/(M + F)$. The larger the difference between M and F, the smaller the number N_e as compared with $M + F$.

2. Unequal size of families. If the gametes are drawn wholly at random from the parents, the number of gametes k contributed by a parent will form a Poisson distribution. In such a case, the effective size is the same as the actual breeding size. However, without perfect random sampling, if the mean number of gametes per parent is $\bar{k} = 2$, the effective size is equal to N_e as expressed in Eq. (7),

$$N_e = \frac{4N - 2}{\sigma_k^2 + 2} \tag{7}$$

where σ is presumably larger than 2.

3. Inbreeding. If F is the inbreeding coefficient of a population, then the effective size is $N_e = N/(1 + F)$.

Periodic change in population size. If $N_1, N_2, \ldots, N_t$ are the respective sizes of the t generations, the average effective size for the period is approximately equal to the harmonic mean of the t sizes. The harmonic mean is much closer to the smallest number of a series than to the largest one.

Distribution

A stationary distribution of gene frequencies results from two opposing forces: the systematic pressures (mutation, migration, selection) which tend to make the gene frequency attain a certain fixed value, and the random variation due to sampling which tends to make the gene frequency drift away from any fixed value. The result of these opposing tendencies is not a single equilibrium value of gene frequency but a stationary distribution of gene frequencies. This distribution may be viewed in three different ways: as the distribution of q for a particular locus in a population in a long period of time; as the distribution of the allelic frequencies of all loci subject to the same pressures in one population at any given time; and finally as the distribution of q of one locus among a large number of populations of the same size and with the same pressures at a given time. *See also:* ALLELE.

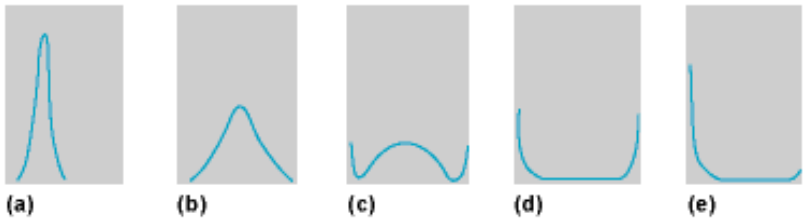

Fig. 2 **Distribution of gene frequency in various populations under differing conditions. (*a*) Large population, distribution clustered close to equilibrium value. (*b*) Intermediate population, distribution of wider spread. (*c*) Intermediate small population, distribution with some terminal fixation. (*d*) Small population, most genes close to fixation or extinction in the absence of systematic pressures. (*e*) Small population with strong selection pressure.**

Under selection pressure and the sampling variation, the distribution function $\varphi(q)$ is $\varphi(q) = C\,\bar{W}^{2N}/q(1-q)$, where C is a constant, $\bar{W}$ the average fitness of the population, and N the effective size of the population. The exact form of the distribution depends upon the value of $\bar{W}$, which is a function of the selection coefficients as well as the gene frequency.

If there is mutation, or migration pressure, or both, the distribution of gene frequency is simply a β distribution: $\varphi(q) = Cq^{U-1}(1-q)^{V-1}$, where $U = 4N\mu$, $V = 4N\nu$ if there is only mutation pressure; $U = 4Nm\,\bar{\mathbf{q}}$, $V = 4Nm\,\bar{p}$ if there is only migration pressure; and $U = 4N(\mu + m\,\bar{\mathbf{q}})$, $V = 4N(\nu + m\,\bar{p})$ if there are both. The exact form of the β distribution depends upon the values of U and V. When they are smaller than unity, the population is considered small; when they are close to unity, the population is intermediate in size; when they are much larger than unity, the population is considered large. When the effects of mutation, migration, and selection are combined, the distribution function is as shown in Eq. (8).

$$\phi(q) = C\overline{W}^{2N} q^{U-1}(1-q)^{V-1} \tag{8}$$

Figure 2 shows some of the forms of the distribution under various conditions. The joint distribution of more than one locus is naturally very complicated. Furthermore, if the mutation rates and selection coefficients also vary instead of being constant, the mathematical description of the stochastic process becomes very laborious and the forms shown in **Fig. 2** give only a first approximation to the real situation.

The distribution forms of the gene frequencies depend upon the relative magnitudes of the various factors which bring about populational changes. In large random-mating populations all gene frequencies remain close to their stable equilibrium values, which are determined by the counteracting but systematic pressures of mutation, selection, and migration. There will be no further genetic change unless the environmental conditions change so as to define new equilibrium points. Evolution is such large populations is guided essentially by intragroup selection, and progress is very slow.

In small and completely isolated populations, most of the gene frequencies are close to 0 or 1 because of the random drift process which dominates the situation. Selection is ineffective. The loci are prevented from being completely fixed only by occasional mutations or immigrants. The ultimate fate of such small homozygous populations is probably extinction because they are nonadaptive and unable to respond to new conditions.

In populations of intermediate size, all factors, both random and systematic, come into play and the population is more responsive to evolutionary change. If a large population is subdivided into many partially isolated groups with migrations between them, there will be some differentiation among the groups, some of it adaptive and some nonadaptive, but there is very little fixation. The selection effect, varying from one locality to another, then operates largely on an intergroup basis which is more efficient than the intragroup selection within one single large population. If the groups are small, some of them will be eliminated by selection while others flourish. This provides the most favorable condition for evolutionary success for the species as a whole. The conclusion is that there is no one all-important factor in evolution. Evolutionary advance depends upon the interplay and balance of all factors.

Two-locus selection

All of the properties described above apply strictly to one locus, or to independently distributed loci. When selection acts upon a two-locus genotype directly, the genes of two loci, linked or unlinked, are in general not independently distributed, even in a large random mating population. In selection involving more than one locus, it is inadequate to consider the gene frequencies at the various loci separately and it is necessary to consider the gamete frequencies. As the simplest example, consider two unlinked loci (*A*,*a* and *B*,*b*). Let the four gamete frequencies (*x*) be as follows:

	B	b	
A	x_1	x_2	p
a	x_3	x_4	q
	u	v	1

Random union of these gametes yields the following offspring genotype frequencies:

	BB	*Bb*	*bb*	
AA	x_1^2	$2x_1x_2$	x_2^2	p_2
Aa	$2x_1x_3$	$2(x_1x_4 + x_2x_3)$	$2x_2x_4$	$2pq$
aa	x_3^2	$2x_3x_4$	x_4^2	q_2
	u^2	$2uv$	v^2	1

Suppose that the nine genotypes have the relative fitness values (*W*) indicated below:

		BB	*Bb*	*bb*
	AA	1	1	0
W:	*Aa*	1	2	1
	aa	0	1	1

The symmetry of the selection pattern ensures the gene frequencies to be $p = q = {}^1/_2$ at the (*A*,*a*) locus and $u = v = {}^1/_2$ at the (*B*,*b*) locus. But it may be verified that the equilibrium gamete frequencies are and not

	B	*b*	
A	.28493	.21508	.50
a	.21508	.28492	.50
	.50000	.50000	1.00

	B	*b*	
A	.25	.25	.50
a	.25	.25	.50
	.50	.50	1.00

The alleles of the two loci are positively correlated. The deviation from random distribution is $D = .28492 - .25000 = .03492$. If these two loci were linked, the deviation D would be even more pronounced. Thus, selection leads to association (positive or negative) among the alleles of different loci. Also, the ultimate result of selection is not necessarily the maximization of the average fitness ($\bar{W}$) of a random mating population as is the case with only one locus (**Fig. 1**). The lack of independence of genes at different loci renders the problem of natural selection and its evolutionary consequences far more complicated than is depicted by simple mathematical models. *See also:* BIOMETRICS; GENETICS; MUTATION.

C. C. Li

Bibliography

G. DeJong, *Population Genetics and Evolution*, 1988

J. S. Gale, *Theoretical Population Genetics*, 1989

J. B. S. Haldane, *The Causes of Evolution*, 1966, reprint 1990

N. D. Henderson and R. Benno (eds.), *Developmental Behavior Genetics: Neural, Biometrical and Evolutionary Approaches*, 1990

T. Nagylaki, *Introduction to Theoretical Population Genetics*, 1992

Additional Readings

A. Altland et al., Rare events in population genetics: Stochastic tunneling in a two-locus model with recombination, *Phys. Rev. Lett.*, 106(8):088101, 2011 DOI: http://doi.org/10.1103/PhysRevLett.106.088101

M. T. Hamblin, E. S. Buckler, and J. L. Jannink, Population genetics of genomics-based crop improvement methods, *Trends Genet.*, 27(3):98–106, 2011 DOI: http://doi.org/10.1016/j.tig.2010.12.003

J. H. Relethford, *Human Population Genetics*, Wiley-Blackwell, Hoboken, NJ, 2012

D. A. Robinson, E. J. Feil, and D. Falush (eds.), *Bacterial Population Genetics in Infectious Disease*, John Wiley & Sons, Hoboken, NJ, 2010

J. Xu, *Microbial Population Genetics*, Caister Academic Press, Norfolk, UK, 2010