

STATISTICAL GENETICS AND EVOLUTION

SEWALL WRIGHT

Introduction. When Darwin developed the theory of evolution by natural selection, practically nothing was known of hereditary differences beyond their existence. Since 1900, a body of knowledge on the mechanism of heredity and on mutation has been built up by experiment that challenges any field in the biological sciences in the extent and precision of its results. The implications for evolution are not, however, immediately obvious. It is necessary to work out the statistical consequences.

Studies in the field of statistical genetics began shortly after the rediscovery of Mendelian heredity in 1900. Those of J. B. S. Haldane [7] and R. A. Fisher [4] have been especially important with respect to the application to evolution. My own approach to the subject came through experimental studies conducted in the U. S. Bureau of Animal Industry on the effects of inbreeding, crossbreeding and selection on populations of guinea pigs [21, 22, 23, 37] and through the attempt to formulate principles applicable to livestock breeding [19, 20, 24, 25, 13, 34]. On moving into the more academic atmosphere of the University of Chicago, I have become more directly concerned with the problem of evolution.

I should note that the deductive approach, to which I shall confine myself here, involves many questions that can only be settled by observation and experimental work on natural populations and that a remarkable resurgence of interest in such work is in progress [2, 9].

Postulates. It will be desirable to begin with a brief review of the more important factors of which account must be taken.

The basic fact of modern genetics is that heredity can be analyzed into separable units, "genes," whose most essential property is that of duplicating themselves with extraordinary precision, irrespective of the characteristics of the organism in whose cells they are carried. We shall restrict consideration to changes in the system of genes and aggregates of genes (chromosomes). There are relatively rare and obscure hereditary changes which must be attributed to other cell components but our knowledge of these does not warrant the elaboration of a statistical theory.

Fortunately the same theory applies to a large extent to gene muta-

The sixteenth Josiah Willard Gibbs Lecture, delivered at Chicago, Illinois, September 3, 1941, under the auspices of the American Mathematical Society; received by the editors October 20, 1941.

tions and to most classes of grosser chromosomal changes (duplications, deficiencies, inversions, translocations, and so on). It will be assumed here that a given kind of mutation occurs at a constant rate per generation. Observed rates in organisms as remote as corn plants, vinegar flies and man are of the order of 10^{-5} or less per generation. Reverse mutation may occur at measurable rates.

It is simplest to deal with mere pairs of alternative conditions (alleles) but the theory remains seriously inadequate unless capable of extension to multiple alleles.

In general I shall assume that the reproductive cells are haploid (that is, contain just one representative from each set of alleles) and that their union results in diploid individuals (with two such representatives in all cells, until reduction occurs in the formation of the germ cells). This is the usual case but there are species in which other situations prevail (tetraploids, hexaploids, aneuploids, and so on). The group of sex linked genes constitutes an important special case in many otherwise completely diploid organisms (including man). I shall not go far into the extension to these cases.

It is simplest to assume that the members of different series of alleles are distributed at random in the reduction division by which the reproductive cells receive a half sample of the genes of the individuals producing them (that, for example, individual $AaBb$ produces germ cells AB , Ab , aB and ab in equal numbers). The phenomenon of partial linkage, exhibited by genes carried in the same chromosome should, however, be taken into account. These are the principal postulates as far as the mechanism of heredity is concerned though others are required in special cases.

The relations of genes to observed characteristics are important. In general, any measurable character is affected by genes at many loci and a single gene often has multiple apparently unrelated effects. The effects of genes in combinations are often roughly cumulative but marked exceptions are also very common. Account must be taken of noncumulative effects within series of alleles (dominance) and between series (gene interaction).

The breeding structure of the population is important. The situation in nature is so complex that models must be chosen that are compromises between mathematical simplicity and biological adequacy [35]. I shall introduce only the simplest models in the course of the present discussion.

Natural selection is an exceedingly complex affair. Selection may occur at various biological levels—between members of the same brood, between individuals of the same local population, between

such populations (as through differential increase and migration) and finally between different species, a subject that carries us outside the field of genetics and which has been discussed mathematically by Lotka [12], Volterra [17] and Nicholson and Bailie [14]. Selection among individuals may relate to the mating activities of one or both sexes, to differences in rate of attainment of maturity, to differential fecundity and to differential mortality at all ages. Selection may act steadily or may vary both in intensity and direction in different regions and at different times. Again I can only deal here with the simplest models.

Gene frequency. In such a complex situation, verbal discussion tends toward a championing of one or another factor. We need a means of considering all factors at once in a quantitative fashion. For this we need a common measure for such diverse factors as mutation, crossbreeding, natural selection and isolation. At first sight these seem to be incommensurables but if we fix attention on their effects on populations, rather than on their own natures, the situation is simplified. Such a measure may be found in the effects on *gene frequency* in each series of alleles.

Because of the complete symmetry of the Mendelian mechanism, gene frequency has no tendency to change in an indefinitely large closed population not subject to mutation or selection. Each homozygote (for example, A_1A_1 , A_2A_2 or A_3A_3) produces only one kind of germ cell. Each heterozygote (for example, A_1A_2 , A_1A_3 , A_2A_3) produces two kinds in *equal* numbers. In a population in which the array of gene frequencies is $(q_1A_1 + q_2A_2 + \dots + q_mA_m)$ (letting the q 's represent the frequencies, and the A 's the genes) the frequencies of genotypes come to equilibrium according to the terms in the expansion of $(q_1A_1 + q_2A_2 + \dots + q_mA_m)^2$ in the first generation of random mating after attainment of equality of gene frequencies in the sexes [8]. Under sex linkage [10, 15] and in polyploids [6] equilibrium is not reached at once but is rapidly approached. Inbreeding and assortative mating change the relative frequencies of homozygotes and heterozygotes but not the gene frequencies.

One immediate consequence of this persistence of gene frequencies is that variability tends to persist. But the slightest continuing unbalanced pressure on gene frequency tends to cause cumulative change. It is obvious that recurrent mutation, immigration, selection, and the accidents of sampling in an isolated population of small size are all factors that can bring about such change.

The frequencies of combinations of different series of alleles (for

example, $A, a; B, b$) reach equilibrium in a random breeding population in which gene frequencies are constant only when the genes are combined at random (terms of $[(1-q_A)a+q_AA]^2[(1-q_B)b+q_BB]^2$). Equilibrium is not reached immediately, however. The departure from equilibrium is halved in each generation of random mating in the case of two pairs of alleles in different chromosomes. In general, the departure is reduced by the proportion \bar{c} , where \bar{c} is the mean chance of recombination [18, 11, 16].

Systematic changes of gene frequencies. The rate at which gene frequency changes under recurrent mutation is obvious [27]. Let q be the frequency of the gene and u the rate at which it mutates to its alleles as a group and Δq the rate of change of q per generation $\Delta q = -uq$.

If reverse mutation occurs at the rate v per generation, the net rate of change of q is

$$(1) \quad \Delta q = v(1 - q) - uq.$$

In the case of multiple alleles, v is the weighted average for the various alleles of the gene in question and is thus a function of their relative frequencies. It is, however, independent of q .

The effect of crossbreeding is similar if we adopt the simplest model [27]. If a population with gene frequency q exchanges the proportion m each generation with a random sample of immigrants from the whole species (gene frequency q_i) the rate of change in gene frequency is

$$(2) \quad \Delta q = -m(q - q_i).$$

In actual cases the immigrants are not likely to be a random sample from the whole species but to come largely from neighboring populations. Effective m is thus, in general, smaller than the apparent amount of immigration and is not necessarily the same for all loci. There may also be selective migration. The simplest model must suffice here. It permits identification of the theories of mutation and immigration by substituting mq_i for v , and $m(1-q_i)$ for u .

The effects of selection have been considered extensively by Haldane [7] in terms of the frequency ratio $(q/1-q)$ and by Fisher [4]. As there can be no selection pressure without at least two alternatives, any expression for it, applicable to all values of q , must include the factor $q(1-q)$, excluding certain limiting cases. Thus the form $aq(1-q)$ has been used by Fisher as the basis for general discussion. For the present purpose somewhat less general forms are more useful. Consider first the case of a random breeding population of

diploid individuals in which the combinations of paired alleles may be assigned constant relative selective values [30].

<i>Genotype</i>	<i>Frequency (f)</i>	<i>Value (W)</i>
<i>AA</i>	q^2	W_{AA}
<i>AA'</i>	$2q(1-q)$	$W_{AA'}$
<i>A'A'</i>	$(1-q)^2$	$W_{A'A'}$

For the frequency of *A* after a generation

$$q_1 = [W_{AA}q^2 + W_{AA'}q(1-q)]/\bar{W}$$

where

$$\begin{aligned} \bar{W} &= \sum fW = W_{AA}q^2 + 2W_{AA'}q(1-q) + W_{A'A'}(1-q)^2, \\ (3) \Delta q &= q_1 - q = q(1-q)[W_{AA}q + W_{AA'}(1-2q) - W_{A'A'}(1-q)]/\bar{W}, \\ \Delta q &= \frac{q(1-q)}{2\bar{W}} \frac{d\bar{W}}{dq}. \end{aligned}$$

Selection, however, really applies to the organism as a whole not to single series of alleles. If the population is heterallelic in n pairs of pertinent alleles, the number of possible combinations is 3^n . Each of these has a certain frequency and a certain relative selective value, the latter of which we here assume to be constant. If the three phases in the *A* series of alleles are combined at random with the combinations of the other series, the average selective values of *AA*, *AA'* and *A'A'* are independent of q_A although functions of the other q 's. Thus

$$(4) \quad \Delta q_A = \frac{q_A(1-q_A)}{2\bar{W}} \frac{\partial \bar{W}}{\partial q_A}.$$

We have assumed only pairs of alleles, but as any group of alleles may be treated formally as one, this formula may be applied to multiple allelic series. The selective values W_{AA} , $W_{AA'}$ and $W_{A'A'}$ are then functions of the relative frequencies within the group of alleles of the gene under consideration, but not of q_A .

In previous general discussions (for example [35, 36]) I have restricted myself to this convenient model of selection pressure. As this has given rise to misapprehension [5], it should be emphasized that it applies only under the conditions implied in its derivation.

If there are selective differences between the sexes, as is very likely to be the case, there are departures from random combination within series of alleles. These are, however, unimportant for most purposes unless there is rather strong selection.

Selection itself tends to bring about departures from random combination among different series of alleles. Again the effects are unimportant in most cases, especially if all relative selective differences are slight.

The formula must be written in a more generalized form to include polyploidy [32] and sex linkage. For small selective differences in a $2k$ -ploid

$$(5) \quad \Delta q_A = \frac{q_A(1 - q_A)}{2k\bar{W}} \frac{\partial \bar{W}}{\partial q_A}.$$

This applies (approximately) under sex linkage if $k = 3/4$ and $\bar{W} = (\bar{W}_m \bar{W}_f)^{1/2}$ where \bar{W}_m and \bar{W}_f are the mean selective values in males and females, respectively.

There may be departures from random mating because of a constant tendency toward mating of relatives, giving the following genotypic frequencies within a series of alleles [23, 3].

<i>Genotype</i>	<i>Frequency</i>
AA	$(1 - F)q^2 + Fq$
AA'	$2(1 - F)q(1 - q)$
$A'A'$	$(1 - F)(1 - q)^2 + F(1 - q)$

Random combination between series of alleles is not disturbed appreciably if the selective differences are small or if the inbreeding coefficient F is small, giving the following formula in which \bar{W}_R and \bar{W}_I are mean selective values of the random bred and inbred components of the frequencies relative to the A series [3]

$$(6) \quad \begin{aligned} \Delta q_A &= \frac{q_A(1 - q_A)}{\bar{W}} \left[\frac{(1 - F)}{2} \frac{\partial \bar{W}_R}{\partial q_A} + F \frac{\partial \bar{W}_I}{\partial q_A} \right] \\ &= \frac{q_A(1 - q_A)}{2\bar{W}} \left[\frac{\partial \bar{W}}{\partial q_A} + F \frac{\partial \bar{W}_I}{\partial q_A} \right]. \end{aligned}$$

Inbreeding that leads to subdivision into partially isolated groups is best dealt with by a different mathematical model.

Under assortative mating based on similarity in characteristics there are very great departures from random combination of different series of alleles [20]. Again we may best consider such a mating system as one leading to subdivision of the population into partially isolated groups.

Returning to consideration of random breeding populations, it may easily be seen that if the W 's are functions of q_A

$$(7) \quad \Delta q_A = \frac{q_A(1 - q_A)}{2\bar{W}} \left[\frac{\partial \bar{W}}{\partial q_A} - \sum \left(f \frac{\partial W}{\partial q_A} \right) \right].$$

Intra-brood selection is an example of a case in which $\partial \bar{W} / \partial q_A = 0$ but gene frequency nevertheless changes. As in our model case, the process does not necessarily lead to fixation of the most favorable of the genotypes possible from the genes present in the population.

Finally, if the genes under consideration affect the system of mating itself, as is the case of the self-sterility alleles of many plants [33] or of genes that determine self fertilization [5], the changes in gene frequency can only be found from the composition of the population in successive generations.

We may note here that while, in principle, selection must be considered to apply to the organism as a whole, one may analyze the organism into character complexes which evolve largely independently through changes in largely independent systems of genes, the components of which are distributed at random among the chromosomes. The case which we have chosen as a model of selection pressure (4) and its generalization (5) should apply sufficiently well to most reaction systems in freely interbreeding populations and gives an insight into certain aspects of the effects of selection which cannot be obtained as easily from the more complex special cases.

If a character complex is affected by n loci and m_i alleles at a particular locus, it requires

$$\sum_{i=1}^n (m_i - 1)$$

dimensions to represent the system of gene frequencies and

$$\prod_{i=1}^n [m_i(m_i + 1)/2]$$

kinds of genotypes are possible. Assuming that the relative selective values of these genotypes are independent of their frequencies, the mean selective values (\bar{W}) of possible random breeding populations form a surface relative to this multi-dimensional system of gene frequencies, the gradient of which determines the way in which the population tends to change under the influence of selection. The number of loci that may affect even the simplest characters are known in certain cases to be great and many, if not all such loci are probably represented at all times by multiple alleles. Thus the number $\prod_{i=1}^n m_i$ of homozygous types possible from genes actually present in

a species and affecting a particular character complex may often run into astronomical figures. Under these conditions it is to be expected that in general the surface \bar{W} for any character complex will have numerous peaks, corresponding not only to different combinations of genes that give the same character [29] but also to different harmonious combinations of elementary characters that permit the organism to overcome the same conditions in different ways.

Mutation, immigration and selection may all be occurring simultaneously. The net rate of change of gene frequency may be obtained by simply adding the contributions of these factors (1), (2), (4), if these are small. In our ideal case of a random breeding population of diploid individuals subject to reversible mutation, immigration and constant selective differences between genotypes [27, 30]

$$(8) \quad \Delta q = v(1 - q) - uq - m(q - q_i) + \frac{q(1 - q)}{2\bar{W}} \frac{\partial \bar{W}}{\partial q}.$$

There is equilibrium, stable or unstable, if $\Delta q = 0$. With reversible mutation, there must be at least one gene frequency other than 0 or 1 that is in stable equilibrium. There may also be stable equilibrium as a result of opposing selection pressures alone.

Accidents of sampling. There is another possibility of change of gene frequency to be considered. In a population that is not indefinitely large, gene frequency may be expected to change from generation to generation merely from the accidents of sampling. The composition of a population of N diploid individuals depends on that of $2N$ gametes produced by the preceding generation. If these are a random sample from the array $[(1 - q)A' + qA]$, the probability array for values of q in the next generation is $[(1 - q)A' + qA]^{2N}$ with the standard deviation $(q(1 - q)/2N)^{1/2}$. We will call a random deviation of q of this sort δq in contrast with the systematic deviation Δq produced by mutation, migration or selection

$$(9) \quad \sigma_{\delta q}^2 = \frac{q(1 - q)}{2N}.$$

It might seem that these random deviations would be negligible in any reasonably large population but in the absence of any systematic pressure toward equilibrium, the squared standard deviation for later generations increases approximately with the number of generations until there is an approach to the limiting value $q(1 - q)$ of complete fixation, $[(1 - q)A'A' + qAA]$. The exact value for the n th generation

is $q(1-q)[1-(1-1/2N)^n]$. The rate of fixation of heterallelic genes approaches $1/2N$ per generation.

The effective value of N should often be much smaller than its apparent value [27, 35]. It obviously applies only to individuals that reach maturity. If there is cyclic variation in population size, N is more closely related to the minimum than to the maximum number. It is also reduced if there is excessive variability in the number of mature offspring from different parents.

In a $2k$ -ploid population [32],

$$(10) \quad \sigma_{\delta q}^2 = \frac{q(1-q)}{2kN} \text{ approximately.}$$

For sex linked genes it is approximately $q(1-q)[2/9N_f + 1/9N_m]$ where N_f and N_m are the effective numbers of females and males, respectively, and thus is $2q(1-q)/3N$ if these are equal.

The distribution of gene frequencies in the case of equilibrium.

The tendency toward a stable equilibrium in the value of q , found when there are opposing systematic pressures, and the tendency to drift away from this point, due to the sampling variance, should result in a probability distribution which one might expect to find realized by the values taken by the frequency of a particular gene over a long period of generations in the ideal case of a population in which all conditions remain constant. It would also be the distribution of values of q taken by this gene at a given time in an array of completely isolated populations, all of which are subject to the same conditions. Finally, all genes subject to systematic pressures of the same magnitude should exhibit such a distribution at one time in a single population. While these are ideal cases, not likely to be approached in actual cases, it is of primary importance in the genetics of populations to be able to reach conclusions on the nature of such distributions.

The distribution of gene frequencies in the case of equilibrium must satisfy the conditions of stability of the mean

$$\overline{(q + \Delta q + \delta q)} = \bar{q}$$

and stability of the variance ($\sigma_{(q+\Delta q+\delta q)}^2 = \sigma_q^2$). The possible values of q must range from 0 to 1. It is convenient to use integration for summation in expanding these expressions. Let $\phi(q)$ be the ordinate of the required distribution. The formula of the distribution may be derived as follows [30, 31]

$$\int_0^1 (q + \Delta q + \delta q)\phi(q) dq = \int_0^1 q\phi(q) dq,$$

$$\int_0^1 (q + \Delta q + \delta q - \bar{q})^2 \phi(q) dq = \int_0^1 (q - \bar{q})^2 \phi(q) dq.$$

Since the mean value of δq is 0 and since δq is not correlated with $(q + \Delta q)$, these reduce to the following, omitting a term involving $(\Delta q)^2$, negligible if Δq is small

$$\int_0^1 \Delta q \phi(q) dq = 0,$$

$$2 \int_0^1 (q - \bar{q}) \Delta q \phi(q) dq + \int_0^1 \sigma_{\delta q}^2 \phi(q) dq = 0.$$

Let $\int \Delta q \phi(q) dq = \chi(q)$ and integrate the first term of the preceding equation by parts

$$\chi(1) - \chi(0) = 0,$$

$$\int_0^1 \chi(q) dq - [\bar{q}\chi(0) + (1 - \bar{q})\chi(1)] - (1/2) \int_0^1 \sigma_{\delta q}^2 \phi(q) dq = 0.$$

It may be found by trial that both of these conditions are satisfied by the following equation if $\phi(0)$ and $\phi(1)$ are finite. Note that $\sigma_{\delta q}^2 = 0$ if $q = 0$ or if $q = 1$, since there can be no sampling variance unless there are alternatives

$$\chi(q) - \chi(1) = (1/2) \sigma_{\delta q}^2 \phi(q),$$

$$d \log [\chi(q) - \chi(1)] = \frac{d\chi(q)}{\chi(q) - \chi(1)} = \frac{2\Delta q dq}{\sigma_{\delta q}^2},$$

$$[\chi(q) - \chi(1)] = \frac{C}{2} e^{2\int (\Delta q / \sigma_{\delta q}^2) dq},$$

$$(11) \quad \phi(q) = (C / \sigma_{\delta q}^2) e^{2\int (\Delta q / \sigma_{\delta q}^2) dq},$$

where C is a constant such that $\int_0^1 \phi(q) dq = 1$.

The frequency of a particular value of q is approximately $f(q) = \phi(q) / 2N$. The amount of exchange between the subterminal and terminal classes is approximately half the frequency of the former from consideration of the Poisson distributions of the classes that are close to fixation [27].

For the model case in which $\Delta q = v(1 - q) - uq - m(q - q_t) + (q(1 - q) / 2\bar{W})(\partial \bar{W} / \partial q)$ and $\sigma_{\delta q}^2 = q(1 - q) / 2N$ equation (11) reduces to the following

$$(12) \quad \phi(q) = C\bar{W}^{2N}q^{4N(mq_t+v)-1}(1-q)^{4N[m(1-q_t)+u]-1},$$

$$(13) \quad \begin{aligned} f(0) &= f(1/2N)/4N[mq_t + v], \\ f(1) &= f(1 - 1/2N)/4N[m(1 - q_t) + u]. \end{aligned}$$

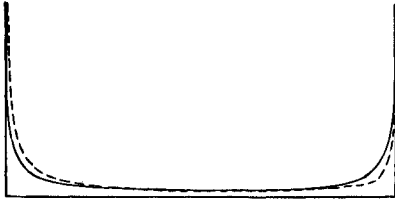


FIG. 1

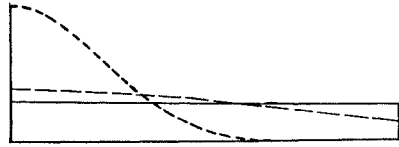


FIG. 2

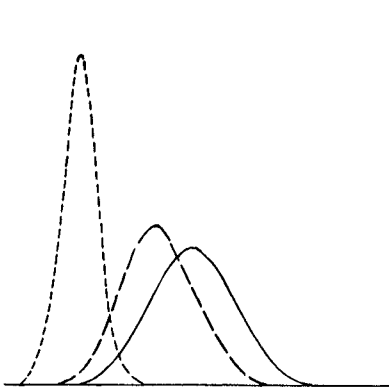


FIG. 3

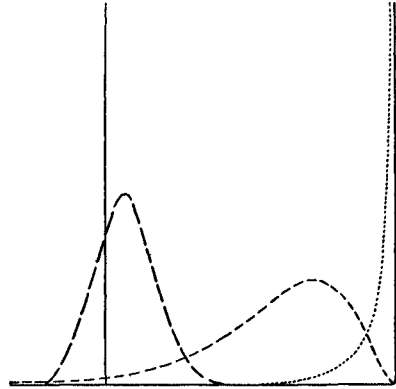


FIG. 4

Figures 1 to 3. Some of the forms taken by the distribution of frequencies of a completely recessive gene. Mutation rates are assumed equal in both directions ($u=v$). $N=1/40v$ in Figure 1, $10/40v$ in Figure 2 and $100/40v$ in Figure 3. In each case the solid line represents the least selection ($t=-v/5$), the long dashes represent selection 10 times as severe (omitted in Figure 1 since practically indistinguishable from the preceding) and the short dashes represent selection 100 times as severe $\phi(q) = Ce^{2Ntq^2}q^{4Nv-1}(1-q)^{4Nv-1}$.

Figure 4. Frequency distribution of a semidominant gene in subgroups of a large population in which the varying conditions of selection among subgroups has lead to a mean gene frequency, $q_t = .25$. The subgroups represented are assumed to be of the same size ($N=1000$) and subject to the same selection pressure ($W_{AA}=1, W_{AA'}=.9975, W_{A'A'}=.995$) but to different degrees of isolation (long dashes: $m = .01$, short dashes: $m = .001$, dots $m = .0001$), $\phi(q) = Ce^{10q}q^{1000m-1}(1-q)^{3000m-1}$.

This brings the effects of reversible mutation, crossbreeding, selection and size of population into a single formula. Figures 1 to 4 show

the forms taken by this distribution in certain special cases. The *U*-shaped distributions in small populations (Figure 1) may be compared with the *I*-shaped ones in large populations (Figure 3). The figures bring out the relatively slight effects of selection in small populations.

The joint frequency distribution for multiple pairs of alleles may be written as follows for the model case that we have been considering [30]

$$(14) \quad \phi(q_1, q_2, \dots, q_n) = C \bar{W}^{2N} \prod_{i=1}^n q_i^{4N(mi q_i + v_i) - 1} (1 - q_i)^{4N[m_i(1 - q_i) + u_i] - 1}.$$

This applies to $2k$ -ploids if $4Nk$ is substituted for $4N$ in the exponents of q_i and $(1 - q_i)$. In the case of sex linkage and equal numbers of the sexes, $3N$ is to be substituted for $4N$ in these exponents. As the exponent of \bar{W} is not affected in these cases, the formula applies to joint distributions including different degrees of ploidy (aneuploids) and both autosomal and sex-linked genes.

Figure 5 illustrates the frequencies along two diagonals of the joint distribution for two pairs of alleles which act cumulatively on the

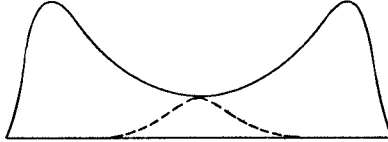


FIG. 5

The frequencies along the diagonals of the joint distribution for two series of alleles with equal and additive effects on a character on which adverse selection acts according to the square of the deviation from the mean ($W_{AAbb} = W_{AaBb} = W_{aaBB} = 1$, $W_{AABb} = W_{AaBB} = W_{Aabb} = W_{aaBb} = 1 - s$, $W_{AABB} = W_{aabb} = 1 - 4s$). The solid line shows the frequencies along the line connecting the two favorable types, *AAbb* and *aaBB*. The dashes show the frequencies along the line connecting the extreme types *AABB* and *aabb*. In the case shown, $u_a = v_a = u_b = v_b$, $N = 1/2v_a$, $s = 5v_a$.

same character of which the midgrade is optimum. There are two peak frequencies corresponding to approximate fixation of two different genotypes that give the midgrade of the character. In cases involving large numbers of genes there may be an indefinitely large number of peak frequencies.

The distribution of gene frequencies under irreversible mutation.

It is also important to determine the form taken by the distribution of gene frequencies when fixation of one of the alleles is an irreversible process. The distribution curve should reach constancy of form, but

all class frequencies (except that in which fixation is occurring) should fall off at a uniform rate K . The conditions may be expressed as follows [31]

$$\begin{aligned} \overline{q + \Delta q + \delta q} &= (1 - K)\bar{q} + K, \\ \sigma_{(q+\Delta q+\delta q)}^2 &= (1 - K)\sigma_q^2 + K(1 - \bar{q})^2, \\ K &= (1/2)f(1 - 1/2N). \end{aligned}$$

It can be found by trial that with mutation at rate v , no selection, and $\sigma_{\delta q}^2 = q(1 - q)/2N$ those conditions are satisfied by the following equation, with decay at rate $K = v$ per generation [31]

(15)
$$f(q) = 2vq^{4Nv-1}.$$

I have not been able to obtain a general solution comparable to equation (11) but formulae have been obtained for an important class of cases by another method [27, 30, 31]. Random breeding diploid populations with frequency array $[(1 - q)A' + qA]$ are distributed in the following generation according to the expression $[(1 - q - \Delta q)A' + (q + \Delta q)A]^{2N}$. Letting $p = 1 - q$, the contribution to the frequency $f(q_c)$ of populations characterized by gene frequency q_c , is thus $[(2N)! / (2Np_c)!(2Nq_c)!] (p - \Delta q)^{2Np_c} (q + \Delta q)^{2Nq_c} f(q)$. The condition that this frequency be reconstructed after a generation except for a reduction by the amount K can be represented sufficiently accurately as follows:

$$(1 - K)\phi(q_c) = \frac{\Gamma(2N)}{p_c q_c \Gamma(2Np_c) \Gamma(2Nq_c)} \int_0^1 (p - \Delta q)^{2Np_c} (q + \Delta q)^{2Nq_c} \phi(q) dq.$$

If $K = 0, \Delta q = 0$, this equation is satisfied by $\phi(q) = Cq^{-1} + D(1 - q)^{-1}$ for any values of C and D . For irreversible mutation at rate v , it yields equation (15) and for migration and reversible mutation, but no selection, the same result as obtained from equation (12) [27, 31].

Selection pressure gives more difficulty. An important case that can be solved is that for very rare mutations ($4Nv$ negligibly small) subject to selection pressure of the fairly general form $\Delta q = (s + tq)q(1 - q)$. Here K may be taken as 0

$$2Nf(q_c) = A \int_0^1 p^{2Np_c} q^{2Nq_c} [1 - q(s + tq)]^{2Np_c} [1 + p(s + tq)]^{2Nq_c} \phi(q) dq,$$

where

$$A = \frac{\Gamma(2N)}{p_c q_c \Gamma(2Np_c) \Gamma(2Nq_c)}.$$

The following approximate relations may be used

$$\begin{aligned} [(1 - q(s + tq)]^{2Np_c} &= e^{-2Np_cq(s+tq)} [1 - Np_cq^2(s + tq)^2], \\ [(1 + p(s + tq)]^{2Nq_c} &= e^{2Nq_cp(s+tq)} [1 - Nq_cp^2(s + tq)^2]. \end{aligned}$$

Let $\phi(q) = e^{2Nsq+2Ntq^2}(C_0 + C_1q + C_2q^2 + \dots)/q(1-q)$ and use the approximate relation

$$\begin{aligned} \frac{\Gamma(2N)}{\Gamma(2Np_c)\Gamma(2Nq_c)} \int_0^1 p^{2Np_c-1} q^{2Nq_c-1+x} dq \\ = q^x + [q_c^{x-1} - q_c^x] [x(x-1)/4N]. \end{aligned}$$

It may be found that $C_m = [(4N^2s^2 + 2Nt)C_{m-2} + 8N^2stC_{m-3} + 4N^2t^2C_{m-4}]/m(m+1)$ ignoring terms in which the exponent of N is less than the sum of those of s and t . After further reduction

$$(16) \quad f(q) = [Ce^{4Nsq+2Ntq^2} + Dqe^{2Nsq+2Ntq^2}\psi(2Nsq, 2Ntq^2)]/q(1-q)$$

where C and D are any constants and $\psi(a, b)$ is as follows

$$\begin{aligned} \psi(a, b) = 1 + \frac{a^2}{3!} + \frac{a^4}{5!} + \frac{a^6}{7!} + \frac{a^8}{9!} + \dots \\ + b(1 + a) \left[\frac{1}{3!} + \frac{2a^2}{5!} + \frac{3a^4}{7!} + \frac{4a^6}{9!} + \dots \right] \\ + b^2 \left[\frac{7}{5!} + \frac{2a}{5!} + \frac{69a^2}{7!} + \frac{6a^3}{7!} + \frac{282a^4}{9!} + \frac{12a^5}{9!} + \dots \right] \\ + b^3 \left[\frac{27}{7!} + \frac{27a}{7!} + \frac{348a^2}{9!} + \frac{204a^3}{9!} + \dots \right] \\ + b^4 \left[\frac{321}{9!} + \frac{132a}{9!} + \dots \right] \\ + b^5 \left[\frac{2265}{11!} + \dots \right] + \dots \end{aligned}$$

The special cases of most interest are those of irreversible mutation in one direction or the other and of equilibrium. Consider a population in which $2Ns$ and $2Nt$ have given values but N is indefinitely large and hence s and t are indefinitely small. If mutation is occurring from the class $q = 0$ at an exceedingly minute rate v with irreversible fixation in the class $q = 1$, the frequency of the subterminal class, $q = 1/2N$, must be approximately $4Nvf(0)$ while that of the other subterminal class ($q = 1 - 1/2N$) must be so much smaller as to be negligible. The following are sufficiently accurate, letting $f(0) = 1$

$$f(1/2N) = 2N \left[C + \frac{D}{2N} \right] = 4Nv,$$

$$f(1 - 1/2N) = 2N [C e^{4Ns+2Nt} + D e^{2Ns+Nt} \psi(2Ns, 2Nt)] = 0.$$

As $\psi(2Ns, 2Nt)$ is of the order of e^{2Ns+Nt} , $D/2N$ is negligible compared with C

$$C = 2v,$$

$$D = - \frac{2v e^{2Ns+Nt}}{\psi(2Ns, 2Nt)},$$

$$(17) \quad f(q) = \frac{2v}{q(1-q)} \left[e^{4Ns q + 2N t q^2} - q e^{2Ns(1+q) + N t(1+q^2)} \frac{\psi(2Ns q, 2N t q^2)}{\psi(2Ns, 2Nt)} \right].$$

In the subterminal regions selection is practically inoperative, $\Delta q = s/2N$, approximately for $q = 1/2N$; and $\Delta q = (s+t)/2N$ for $q = (1-1/2N)$. The formula of the curve in the neighborhood of $q = 1/2N$ is therefore approximately $2v/q$. Thus with given v but different values of N , $f(1/2N) (=4Nv)$ always falls very nearly on the same smooth curve. The relation between the selection pressure and the sampling variance (which measures the additive effect of sampling) is constant for given values of $2Ns$, $2Nt$ and q , $(\Delta q/\sigma_{\Delta q}^2) = (2Ns + 2Ntq)$. Thus the smoothed probability curve, $f(q)$, should be the same throughout with given $2Ns$ and $2Nt$ irrespective of the values of N and of s , t separately. Thus equation (17) is the general formula for the case of irreversible mutation from the class at $q = 0$.

The determination of $\psi(a, b)$ in specific cases is a rather formidable task since it involves two variables and converges slowly. For the case of semidominance however, $t = 0$ and $\psi(2Ns q, 0)$ reduces to $(e^{2Ns q} - e^{-2Ns q})/4Ns q$

$$(18) \quad f(q) = \frac{2v}{q(1-q)} \frac{(1 - e^{-4Ns(1-q)})}{(1 - e^{-4Ns})}.$$

This agrees with a result obtained by Fisher [4] by a different method, involving a transformation of scale ($\theta = \cos^{-1}(1-2q)$) designed to make sampling variance uniform, and expression of the conditions in the form of a differential equation. The chance of fixation of a mutation in this case is given by the ratio of the subterminal classes ($f(1-1/2N)/f(1/2N)$) and is thus $2s/(1-e^{-4Ns})$ where s is the selection favoring the heterozygote. This is practically constant ($2s$) in large populations. There is a small chance of fixation of even unfavorable mutations ($2s/(e^{4Ns} - 1)$). Figure 6 illustrates the distribu-

tion curve for various intensities of selection (but with the selection favoring heterozygotes represented by $(1/2)s$ instead of s).

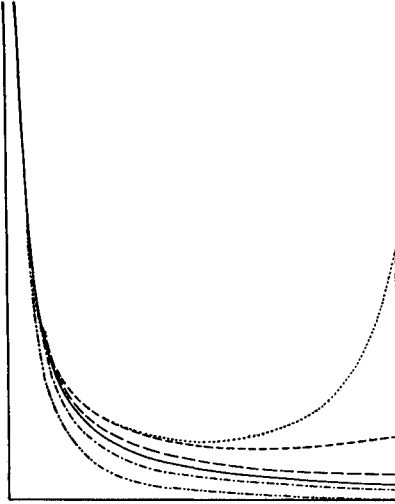


FIG. 6

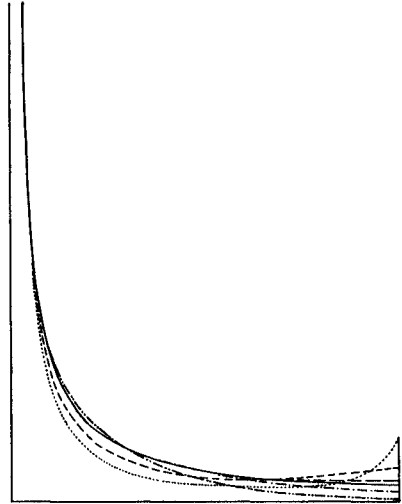


FIG. 7

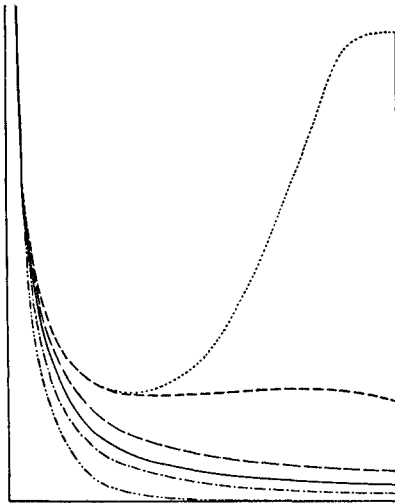


FIG. 8

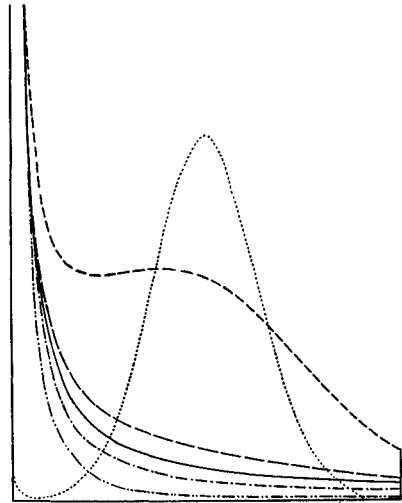


FIG. 9

Figures 6 to 9. Distributions of gene frequencies under irreversible mutation at indefinitely low rates (v). Six cases are shown in each figure. Dash and two dots, $s = -4/2N$; dash and dot, $s = -1/2N$; solid line, $s = 0$ ($f(q) = 2v/q$ in all figures); long

In the case of recessive mutations, $s = 0$ and $\Delta q = tq^2(1 - q)$. The law of the series for $\psi(0, b)$ is that $E_m = (E_{m-1} + E_{m-2})/2m(2m - 1)$ where E_m is the coefficient of b^m . The distribution curve for a number of values of $2Nt$ are shown in Figure 7 (but with s instead of t as the selection coefficient). The chance of fixation was determined empirically from the ratio of subterminal frequencies up to $2Nt = 12$. For larger values (up to $2Nt = 64$) it was more convenient to calculate it from the ratio of the flux ($2N \Delta q f(q)$) in the region of maximum Δq , ($q = 2/3$) to the proportion of mutations ($2Nv$) (usable only for large $\Delta q/\sigma_{sq}^2$). For values of t ranging from $4/2N$ to $64/2N$ the average chance of fixation came out $1.1(t/2N)^{1/2}$, apparently approaching $(t/2N)^{1/2}$ and thus a function of N even in large populations contrary to the case of semidominance.

In the case of dominant mutations, $\Delta q = (s - sq)q(1 - q)$ if mutations are taken as occurring from the class $q = 0$. It is more convenient to assume that they are occurring from the class $q = 1$ ($\Delta q = -sq^2(1 - q)$) since this merely requires evaluation of $\psi(0, -2Nsq^2)$ instead of the two-dimensional series $\psi(2Nsq, -2Nsq^2)$. From considerations analogous to those discussed above, it appears that for $\Delta q = (s + tq)q(1 - q)$ but irreversible mutation from the class at $q = 1$, $C = 0$, $D = 2ve^{-2Ns - Nt}/\psi(2Ns, 2Nt)$ to a sufficient approximation,

$$(19) \quad f(q) = \frac{2v}{q(1 - q)} \left[\frac{qe^{-2Ns(1-q) - Nt(1-q^2)} \psi(2Nsq, 2Ntq^2)}{\psi(2Ns, 2Nt)} \right].$$

dashes, $s = 1/2N$; short dashes, $s = 4/2N$; dots, $s = 16/2N$. The ordinate at $q = 1/2N$ is the same for all curves ($f(q) = 4Nv$) and far above the range of the figures (except in the case $s = 16/2N$, Figure 9, in which all ordinates are greatly reduced).

Figure 6. The case of a semidominant mutation A' ($W_{AA} = 1$, $W_{AA'} = 1 + s/2$, $W_{A'A'} = 1 + s$). The probabilities (P) of fixation of a mutation are as follows: For $s = -4/2N$, $P = 0.075/2N$; for $s = -1/2N$, $P = 0.58/2N$; for $s = 0$, $P = 1/2N$; for $s = 1/2N$, $P = 1.6/2N$; for $s = 4/2N$, $P = 4.1/2N$; for $s = 16/2N$, $P = 16/2N$; for large s , $P = s$.

Figure 7. The case of a recessive mutation a , ($W_{AA} = W_{Aa} = 1$, $W_{aa} = 1 + s$). For $s = -4/2N$, $P = 0.12/2N$; for $s = -1/2N$, $P = 0.70/2N$; for $s = 0$, $P = 1/2N$; for $s = 1/2N$, $P = 1.3/2N$; for $s = 4/2N$, $P = 2.3/2N$; for $s = 16/2N$, $P = 4.3/2N$; for large s , $P = (s/2N)^{1/2}$.

Figure 8. The case of a dominant mutation A ($W_{aa} = 1$, $W_{Aa} = W_{AA} = 1 + s$). For $s = -4/2N$, $P = 0.042/2N$; for $s = -1/2N$, $P = 0.49/2N$; for $s = 0$, $P = 1/2N$; for $s = 1/2N$, $P = 1.9/2N$; for $s = 4/2N$, $P = 6.6/2N$; for $s = 16/2N$, $P = 31/2N$; for large s , $P = 2s$.

Figure 9. The case of a mutation A' , selected only in heterozygotes ($W_{AA} = W_{A'A'} = 1$, $W_{AA'} = 1 + s$). For $s = -1/4N$, $P = 0.23/2N$; for $s = -1/2N$, $P = 0.71/2N$; for $s = 0$, $P = 1/2N$; for $s = 1/2N$, $P = 1.4/2N$; for $s = 4/2N$, $P = 3.2/2N$; for large s , P approaches 0 and all loci tend to become heterallelic.

Figure 8 shows the form taken by the distribution in special cases (with s as the selection favoring the dominant mutation and mutation taken as occurring from the class $q=0$). The chance of fixation of favorable dominant mutation with large N is approximately $2s$ which is the same as for favorable semidominant mutations provided that s is the selection favoring the heterozygotes in both cases. This is to be expected since homozygotes are relatively rare until mutation has passed through the point of maximum selection pressure.

Figure 9 shows the distribution curve in the case in which there is no selective difference between the homozygotes but selection favors or opposes the heterozygotes. In this case in which $\Delta q = s(1-2q)q(1-q)$, formula (17) was used for values of $2Ns$ up to 4. For large values of $2Ns$, there is so little fixation that the distribution under equilibrium may be used.

By combining the formulae for irreversible mutation in each direction in such a ratio that the amounts of fixation are equal, we obtain a distribution identical with that of reversible mutation occurring at rates at which $4Nv$ and $4Nu$ are negligible. The result agrees with the limiting value obtained from equation (12)

$$(20) \quad f(q) = \frac{C}{q(1-q)} [e^{4Ns q + 2Nt q^2}].$$

The evolutionary process. I can go only briefly into the implications for evolution. We must distinguish two processes (a) the transformation of a single population until it has become so different that a new species or higher category must be recognized and (b) the cleavage of species.

Consider first the possibilities of transformation in a very large, closed, freely interbreeding population, living under conditions that are the same on the average for thousands of generations [28]. In such a population, random changes in gene frequency are negligible. Gene frequencies can change only according to the systematic pressures of mutation and selection, a process which must stop when all Δq 's become zero, unless there is a flow of untried mutations that are favorable from the first. We have a theory of the stability of species in spite of variability due to continually occurring mutations and in spite of continuous action of selection.

A stable state of this sort may be far from being the most adaptive of the systems possible from the genes actually present. Consider here the situation with respect to characters to which our model selection pressure applies. As noted, many distinct peak values are to be expected in the surface of selective values, \bar{W} , relative to the multi-

dimensional system of gene frequencies. In a species located at a particular point in the system, each gene frequency will change until all Δq 's are zero. These changes will be such that the mean selective value of the populations changes approximately by the amount $\Delta \bar{W} = \Sigma(\Delta q \partial \bar{W} / \partial q)$, the species moving up the steepest gradient in the surface \bar{W} except as affected by mutation pressures [29]. It stops when $\Delta \bar{W} = 0$, a point in the neighborhood of the peak toward which selection has been directed, but not at the highest point because of the mutation terms in the Δq 's. In general there will be other peaks on the surface \bar{W} that are higher but the species cannot reach them.

Perhaps, however, we have been too hasty in assuming that all Δq 's would ever reach zero simultaneously. It is probable that the potential alleles at each locus form an indefinitely extended series in which any one allele can give rise to certain others, these to ones at two removes from the first, and so on. A continual flow of untried favorable mutations may keep the population in a state of flux. In general, however, it would seem probable that the rate of the movement toward the equilibrium point indicated by the existing genes would be of a higher order of magnitude than the rate of elevation of this peak by the occurrence of mutations of this very unusual sort.

In a population in approximate equilibrium, the variability due to the balance between mutation and selection is not likely to be great. If, for example, $\Delta q = sq(1 - q) - uq$, $q = 1 - u/s$ at equilibrium. As mutation rates are typically of the order 10^{-5} or less, q is close to 1 if the gene in question has an appreciable advantage. Loci in which there are opposing selection pressures would contribute more to variability. This may occur where a heterozygote combines favorable effects of two genes (Δq of form $q(1 - q)[s_2 - (s_1 + s_2)q]$ with stable equilibrium at $q = s_2/(s_1 + s_2)$ if both s_1 and s_2 are positive). It may also occur where different homozygotes have advantages in different ecological niches occupied by the species in such a way that the net selective values are related inversely to the frequencies (a case to which our model applies only approximately). With only one or the same few alleles maintained at high frequencies by the population the chance for the occurrence of fixation of untried favorable mutations at several removes from those present is small. The extreme improbability of such mutations justifies use of the formulae for irreversible mutation at very low rates ($4Nv$ much less than 1) in spite of the fact that we are considering large populations (17), (19) (Figures 6 to 9). There is the possibility of an indefinitely continuing evolutionary process but the rate is restricted by the low probability of the necessary mutations and the incomplete utilization of the potentialities for adaptation pro-

vided by the genes actually present (cf. however Fisher [4] and critique [26]).

Conditions are, however, continually changing. Selection of increased severity but unchanged direction merely carries the location of the system of gene frequencies closer to the peak and increases somewhat the chance for a novel favorable mutation to reach fixation. With secular changes in the *direction* of selection, on the other hand, peaks in the surface \bar{W} may become depressed and low places elevated. In species which are sufficiently labile to avoid extinction, the system of gene frequencies is kept continually on the move. It is likely to be shuffled into regions of \bar{W} that are in general the higher ones. This process is undoubtedly of great importance for evolutionary change.

In sufficiently small populations, the random divergences of gene frequencies from their equilibrium values become important. In very small populations, these tend to bring about approximate fixation of some random (and therefore, in general, non-adaptive) combination of genes (Figure 1). Moreover, while selection pressure is less effective in small populations than in large ones, mutation pressure remains the same. Random mutations are more likely to be degenerative than adaptive. Long continued reduction in the size of a population is thus likely to lead to extinction. On the other hand, the less extreme random variations found in populations of intermediate size ($4Nv$ or $4Ns$ of the order 1 to 10) (cf. Figures 2, 3, 5) act somewhat like changes in the direction of the selection. The system of gene frequencies is kept continually on the move and this gives a trial and error process which at times may lead to adaptive combinations which would not have been reached by direct selection. The rate of change of this sort is slow under the required conditions.

Consider next a large population that is divided into many small partially isolated races. These may differ in size and degree of isolation and in the direction and severity of the selection to which they are subjected. The conditions are present for an extensive testing by trial and error of a relatively large number of alleles at each locus and of different combinations of these.

Local differences in direction of selection are effective if the selection coefficient s (writing selection pressure in the form $\Delta q = sq(1 - q)$) exceeds the crossbreeding coefficient m (cf. Figure 4). While these differences are primarily of merely local significance, there is the possibility of acquirement of an adaptation that turns out to be of general value which then may spread throughout the species. Moreover, if different alleles come to be characteristic of different races, the store

of variability of the species as a whole is increased. This is also true of each local race as a consequence of crossbreeding.

In races in which $4Nm$ is of the order 1 to 10 there is considerable random differentiation without the approach to fixation in equally small completely isolated groups. Such changes may occur with considerable rapidity in this case and while non-adaptive are not necessarily anti-adaptive to an appreciable extent. It is merely that the location of the system of gene frequencies on the surface of selective values in our model case is not constrained to move up the *steepest* gradient but may move up gradients that are not the steepest and occasionally even down hill. Among the many local races exploring the neighborhood of a peak in the surface of selective values, one or more may reach a gradient leading to a higher peak (cf. Figure 10).

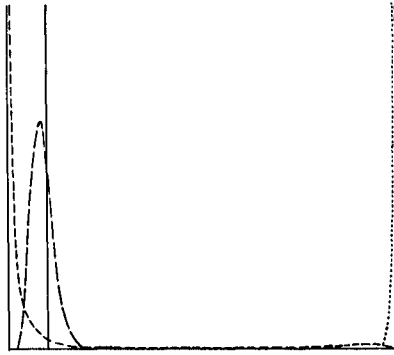


FIG. 10

The frequencies along the diagonal (0, 0) to (1, 1) of the joint distribution of frequencies of recessive genes, a and b in subgroups of the same size ($N=1000$) but different degrees of isolation (long dashes: $m=.01$; short dashes: $m=.001$; dots: $m=.0001$) in a large population in which the conditions are everywhere the same (each recessive with selective disadvantage of .001 relative to the type $A-B-$ but the double recessive with an advantage .01 over type). With mutation rates $u_a=v_a=u_b=v_b=10^{-5}$ there would be two positions of stable equilibrium, one at about (0.1, 0.1) which is assumed to hold for the major portion of the species for historical reasons, and one at about (.999, .999) which is more adaptive

$$\phi(q_a, q_b) = C[1 - .002(q_a + q_b) + .012q_aq_b]^{2000} (q_aq_b)^{400m-1} [(1 - q_a)(1 - q_b)]^{3600m-1}.$$

Local populations for which $m=.01$ vary only slightly from the frequencies characteristic of the species. Most of those for which $m=.001$ show a close approach to fixation of the type genes ($AABB$) but occasionally there is an approach to fixation of the more adaptive double recessive. Those with $m=.0001$ are largely homallelic in $aabb$. In the long run such superior subgroups might be expected to pull the whole species to this position by intergroup selection. The scales are not the same in this figure.

By expansion of numbers and excess migration such races tend to bring the species as a whole under control of this peak. Intergroup selection of this sort, with respect to racial differentiation that has jointly adaptive and non-adaptive aspects, seems to provide the most effective mechanism for testing many alleles at each locus and many combinations of these and is thus the most effective mechanism for a continuing evolutionary process [26, 27, 28, 29, 35, 36].

It should be emphasized that we are not concerned here with local races as incipient species. As long as isolation is incomplete the races are bound together by crossbreeding and thus are carried along by the evolution of the species as a whole although subject to the minor kaleidoscopic changes in character which according to this theory play a major role in the evolution of the whole.

The cleavage of species depends on virtually complete isolation of portions of the species from each other. Even if there are no significant character differences at the time of separation and even if conditions remain substantially the same for the two portions, the process described above will insure that they drift apart. Each continues to be adapted to the conditions but in somewhat different ways. They may be expected to move to increasingly more remote peaks on the surface \bar{W} . In the course of time genic and chromosomal differences may be expected to accumulate that prevent crossing and so clinch the specific distinction. Before this point is reached, the occasional occurrence of hybridization may transfer blocks of genes from one species to the other or lead to the origin of a completely hybrid species, presenting a mechanism of reticulate evolution, analogous to that described above but on a coarser scale (cf. [1]).

Under certain conditions the multiplication and diversification of species may be a very rapid process. These include a relaxation of the general selection pressure on the species permitting great increase in numbers and great variability; the opportunity for the occupation of widely distinct ecological niches associated with almost complete isolation of the groups seizing these opportunities and with subdivision of these groups into partially isolated local populations. A species that is the first of its general kind to reach unoccupied territory finds most at least of these conditions realized. This is also the case with a species that by any means acquires an adaptation of first rate general significance which gives its subgroups an advantage over species already established in various ecological niches, that more than compensates for the initial lack of special adaptations for these niches.

The most general conclusion that can be drawn from the attempt to develop a mathematical theory of the simultaneous effects of all sta-

tistical processes that affect the genetic composition of populations is that in general the most favorable conditions for evolutionary advance are found when these are balanced against each other in certain ways, rather than where any one completely dominates the situation. Finally it may be said that the more detailed knowledge of heredity and mutation that is now available confirms Darwin's general contention that evolution is a process of statistical transformation of populations.

LITERATURE CITED

1. E. Anderson, *The species problem in iris*, Annals of the Missouri Botanical Garden, vol. 23 (1936), pp. 457-509.
2. T. Dobzhansky, *Genetics and the Origin of Species*, Columbia University Press, New York, 1941, 445 pp.
3. T. Dobzhansky and S. Wright, *Genetics of natural populations. V. Relations between mutation rate and accumulation of lethals in populations of *Drosophila pseudoobscura**, Genetics, vol. 26 (1941), pp. 23-51.
4. R. A. Fisher, *The Genetical Theory of Natural Selection*, Clarendon Press, Oxford, 1930. 272 pp. (Much earlier work is here summarized.)
5. ———, *Average excess and average effect of a gene substitution*, Annals of Eugenics, vol. 9 (1941), pp. 53-63.
6. J. B. S. Haldane, *Theoretical genetics of autopolyploids*, Journal of Genetics, vol. 22 (1930), pp. 359-372.
7. ———, *The Causes of Evolution*, Harper, London, 1932, 235 pp. (Much earlier work is here summarized.)
8. G. H. Hardy, *Mendelian proportions in a mixed population*, Science, vol. 28 (1908), pp. 49-50.
9. J. S. Huxley (editor), *The New Systematics*, Clarendon Press, Oxford, 1940, 583 pp.
10. H. S. Jennings, *The numerical results of diverse systems of breeding*, Genetics, vol. 1 (1916), pp. 53-89.
11. ———, *The numerical results of diverse systems of breeding with respect to two pairs of characters, linked or independent, with special relation to the effects of linkage*, Genetics, vol. 2 (1917), pp. 97-154.
12. A. J. Lotka, *Elements of Physical Biology*, Williams and Wilkins, Baltimore, 1925, 460 pp.
13. H. C. McPhee and S. Wright, *Mendelian analysis of the pure breeds of live stock. III. The shorthorns*, Journal of Heredity, vol. 16 (1925), pp. 205-215.
14. A. J. Nicholson and V. A. Baillie, *The balance of animal populations*, Proceedings of the Zoological Society of London, 1935, pp. 551-598.
15. R. B. Robbins, *Application of mathematics to breeding problems. II*, Genetics, vol. 3 (1918), pp. 73-92.
16. ———, *Some applications of mathematics to breeding problems. III*, Genetics, vol. 3 (1918), pp. 375-389.
17. V. Volterra, *Leçons sur la théorie mathématique de la lutte pour la vie*, Paris, 1931.
18. W. Weinberg, *Über Vererbungsgesetze beim Menschen*, Zeitschrift für induktive Abstammungs und Vererbungslehre, vol. 1 (1909), pp. 277-330.

19. S. Wright, *Principles of livestock breeding*, Bulletin no. 905, U. S. Department of Agriculture, 1920, 68 pp.
20. ———, *Systems of mating*, Genetics, vol. 6 (1921), pp. 111–178.
21. ———, *The effects of inbreeding and crossbreeding on guinea pigs*. I. *Decline in vigor*, Bulletin no. 1090, U. S. Department of Agriculture, 1922, pp. 1–36.
22. ———, *The effects of inbreeding and crossbreeding on guinea pigs*. II. *Differentiation among inbred families*, *ibid.*, pp. 37–63.
23. ———, *The effects of inbreeding and crossbreeding on guinea pigs*. III. *Crosses among highly inbred families*, Bulletin no. 1121, U. S. Department of Agriculture, 1922, 59 pp.
24. ———, *Mendelian analysis of the pure breeds of livestock*. I. *The measurement of inbreeding and relationship*, Journal of Heredity, vol. 14 (1923), pp. 339–348.
25. ———, *Mendelian analysis of the pure breeds of livestock*. II. *The Duchess family of shorthorns as bred by Thomas Bates*, *ibid.*, pp. 405–422.
26. ———, *The genetical theory of natural selection* (review of book by R. A. Fisher), Journal of Heredity, vol. 21 (1930), pp. 349–356.
27. ———, *Evolution in Mendelian populations*, Genetics, vol. 16 (1931), pp. 97–159
28. ———, *The roles of mutation, inbreeding, crossbreeding and selection in evolution*, Proceedings of the 6th International Congress of Genetics, vol. 1 (1932), pp. 356–366.
29. ———, *Evolution in populations in approximate equilibrium*, Journal of Genetics, vol. 30 (1935), pp. 257–266.
30. ———, *The distribution of gene frequencies in populations*, Proceedings of the National Academy of Sciences, vol. 23 (1937), pp. 307–320.
31. ———, *The distribution of gene frequencies under irreversible mutation*, Proceedings of the National Academy of Sciences, vol. 24 (1938), pp. 253–259.
32. ———, *The distribution of gene frequencies in populations of polyploids*, *ibid.* pp. 372–377.
33. ———, *The distribution of self-sterility alleles in populations*, Genetics, vol. 24 (1939), pp. 538–552.
34. ———, *Genetic principles governing the rate of progress of livestock breeding*, Proceedings of the American Society of Animal Production, 1939, pp. 18–26.
35. ———, *Breeding structure of populations in relations to speciation*, American Naturalist, vol. 74 (1940), pp. 232–248.
36. ———, *The statistical consequences of Mendelian heredity in relation to speciation* (chapter in *The New Systematics*, edited by J. S. Huxley), Clarendon Press, Oxford, 1940.
37. S. Wright and O. N. Eaton, *The persistence of differentiation among inbred families of guinea pigs*, Technical Bulletin no. 103, U. S. Department of Agriculture, 1929, 45 pp.

UNIVERSITY OF CHICAGO