

# EVOLUTION IN MENDELIAN POPULATIONS

SEWALL WRIGHT

*University of Chicago, Chicago, Illinois*

Received January 20, 1930

## TABLE OF CONTENTS

	PAGE
Theories of evolution . . . . .	97
Variation of gene frequency . . . . .	100
Simple Mendelian equilibrium . . . . .	100
Mutation pressure . . . . .	100
Migration pressure . . . . .	100
Selection pressure . . . . .	101
Equilibrium under selection . . . . .	102
Multiple allelomorphs . . . . .	104
Random variation of gene frequency . . . . .	106
Rate of decrease in heterozygosis . . . . .	107
The population number . . . . .	110
The distribution of gene frequencies and its immediate consequences . . . . .	111
No mutation, migration or selection . . . . .	111
Nonrecurrent mutation . . . . .	116
Reversible recurrent mutation . . . . .	121
Migration . . . . .	126
Irreversible recurrent mutation . . . . .	128
Selection . . . . .	129
General formula . . . . .	133
The distribution curves . . . . .	134
Dominance ratio . . . . .	137
The mean and variability of characters . . . . .	139
The evolution of Mendelian systems . . . . .	142
Classification of the factors of evolution . . . . .	142
Lability as the condition for evolution . . . . .	147
Control of evolution . . . . .	151
Agreement with data of evolution . . . . .	153
"Creative" and "emergent" evolution . . . . .	154
SUMMARY . . . . .	155
LITERATURE CITED . . . . .	158

## THEORIES OF EVOLUTION

One of the major incentives in the pioneer studies of heredity and variation which led to modern genetics was the hope of obtaining a deeper insight into the evolutionary process. Following the rediscovery of the Mendelian mechanism, there came a feeling that the solution of problems of evolution and of the control of the process, in animal and plant breeding

and in the human species, was at last well within reach. There has been no halt in the expansion of knowledge of heredity but the advances in the field of evolution have, perhaps, seemed disappointingly small. One finds the subject still frequently presented in essentially the same form as before 1900, with merely what seems a rather irrelevant addendum on Mendelian heredity.

The difficulty seems to be the tendency to overlook the fact that the evolutionary process is concerned, not with individuals, but with the species, an intricate network of living matter, physically continuous in space-time, and with modes of response to external conditions which it appears can be related to the genetics of individuals only as statistical consequences of the latter. From a still broader viewpoint (compare LOTKA 1925) the species itself is merely an element in a much more extensive evolving pattern but this is a phase of the matter which need not concern us here.

The earlier evolutionists, especially LAMARCK, assumed that the somatic effects of physiological responses of individuals to their environments were transmissible to later generations, and thus brought about a directed evolution of the species as a whole. The theory remains an attractive one to certain schools of biologists but the experimental evidence from genetics is so overwhelmingly against it as a general phenomenon as to render it unavailable in present thought on the subject.

DARWIN was the first to present effectively the view of evolution as primarily a statistical process in which random hereditary variation merely furnishes the raw material. He emphasized differential survival and fecundity as the major statistical factors of evolution. A few years later, the importance of another aspect of group biology, the effect of isolation, was brought to the fore by WAGNER. Systematic biologists have continued to insist that isolation is the major species forming factor. As with natural selection, a connection with the genetics of individuals can be based on statistical considerations.

There were many attempts in the latter part of the nineteenth century to develop theories of direct evolution in opposition to the statistical viewpoint. Most of the theories of orthogenesis (for example, those of EIMER and of COPE) implied the inheritance of "acquired characters." NÄGELI postulated a slow but self contained developmental process within protoplasm; practically a denial of the possibility of a scientific treatment of the problem. Differing from these in its appeal to experimental evidence and from the statistical theories in its directness, was DE VRIES' theory of the abrupt origin of species by "mutations." A statistical process, selec-

tion or isolation, was indeed necessary to bring the new species into predominance, but the center of interest, as with Lamarckism, was in the physiology of the mutation process.

The rediscovery of Mendelian heredity in 1900 came as a direct consequence of DE VRIES' investigations. Major Mendelian differences were naturally the first to attract attention. It is not therefore surprising that the phenomena of Mendelian heredity were looked upon as confirming DE VRIES' theory. They supplemented the latter by revealing the possibilities of hybridization as a factor bringing about an extensive recombination of mutant changes and thus a multiplication of incipient species, a phase emphasized especially by LOTSY. JOHANNSEN'S study of pure lines was interpreted as meaning that DARWIN'S selection of small random variations was not a true evolutionary factor.

A reaction from this viewpoint was led by CASTLE, who demonstrated the effectiveness of selection of small variations in carrying the average of a stock beyond the original limits of variation. This effectiveness turned out to depend not so much on variability of the principal genes concerned as on residual heredity. As genetic studies continued, ever smaller differences were found to mendelize, and any character, sufficiently investigated, turned out to be affected by many factors. The work of NILSSON-EHLE, EAST, SHULL, and others established on a firm basis the multiple factor hypothesis in cases of apparent blending inheritance of quantitative variation.

The work of MORGAN and his school securely identified Mendelian heredity with chromosomal behavior and made possible researches which further strengthened the view that the Mendelian mechanism is the general mechanism of heredity in sexually reproducing organisms. The only exceptions so far discovered have been a few plastid characters of plants. That differences between species, as well as within them, are Mendelian, in the broad sense of chromosomal, has been indicated by the close parallelism between the frequently irregular chromosome behavior and the genetic phenomena of species crosses (FEDERLEY, GOODSPEED and CLAUSEN, etc. ). Most of DE VRIES' mutations have turned out to be chromosome aberrations, of occasional evolutionary significance, no doubt, in increasing the number of genes and in leading to sterility of hybrids and thus isolation, but of secondary importance to gene mutation as regards character changes. As to gene mutation, observation of those which have occurred naturally as well as of those which MULLER, STADLER, and others have recently been able to produce wholesale by X-rays, reveals characteristics which seem as far as possible from those required for a directly adaptive evolutionary process. The conclusion nevertheless seems warranted by

the present status of genetics that any theory of evolution must be based on the properties of Mendelian factors, and beyond this, must be concerned largely with the statistical situation in the species.

#### VARIATION OF GENE FREQUENCY

##### *Simple Mendelian equilibrium*

The starting point for any discussion of the statistical situation in Mendelian populations is the rather obvious consideration that in an indefinitely large population the relative frequencies of allelomorphs remain constant if unaffected by disturbing factors such as mutation, migration, or selection. If  $[(1-q)a + qA]$  represents the frequencies of two allelomorphs, (a, A) the frequencies of the zygotes reach equilibrium according to the expansion of  $[(1-q)a + qA]^2$  within at least two generations,<sup>1</sup> whatever the initial composition of the population (HARDY 1908). Combinations of different series are in equilibrium when these are combined at random, but as WEINBERG (1909) and later, in more detail, ROBBINS (1918) have shown, equilibrium is not reached at once but is approached asymptotically through an infinite number of generations. Linkage slows down the approach to equilibrium but has no effect on the ultimate frequencies.

##### *Mutation pressure*

The effects of different simple types of evolutionary pressure on gene frequencies are easily determined. Irreversible mutation of a gene at the rate  $u$  per generation changes gene frequency ( $q$ ) at the rate  $\Delta q = -uq$ . With reverse mutation at rate  $v$  the change in gene frequency is  $\Delta q = -uq + v(1-q)$ . In the absence of other pressures, an equilibrium is reached between the two mutation rates when  $\Delta q = 0$ , giving  $q = v/(u+v)$ .

##### *Migration pressure*

The frequency of a gene in a given population may be modified by migration as well as by mutation. As an ideal case, suppose that a large population with average frequency  $q_m$  for a particular gene, is composed of subgroups each exchanging the proportion  $m$  of its population with a random sample of the whole population. For such a subgroup,  $\Delta q = -m(q - q_m)$ .

The conditions postulated above are rather artificial since, in an actual species, subgroups exchange individuals with neighboring subgroups rather

<sup>1</sup> This statement assumes that there is no overlapping of generations which may bring about some delay in the attainment of equilibrium.

than with a random sample of the whole species and the change in  $q$  will be only a fraction of that given above. The fraction is the average degree of departure of the neighboring subgroups toward the population average. The formula may be retained by letting  $q_m$  stand for the gene frequency of immigrants rather than of the whole species.

### *Selection pressure*

Selection, whether in mortality, mating or fecundity, applies to the organism as a whole and thus to the effects of the entire gene system rather than to single genes. A gene which is more favorable than its allelomorph in one combination may be less favorable in another. Even in the case of cumulative effects, there is generally an optimum grade of development of the character and a given plus gene will be favorably selected in combinations below the optimum but selected against in combinations above the optimum. Again the greater the number of unfixed genes in a population, the smaller must be the average effectiveness of selection for each one of them. The more intense the selection in one respect, the less effective it can be in others. The selection coefficient for a gene is thus in general a function of the entire system of gene frequencies. As a first approximation, relating to a given population at a given moment, one may, however, assume a constant net selection coefficient for each gene. Assume that the genes  $a$  and  $A$  tend to be reproduced in the ratio  $(1-s):1$  per generation. The gene array  $[(1-q)a + qA]$  becomes  $[(1-s)(1-q)a + qA]/[1-s(1-q)]$ . The change in the frequency of  $A$  is  $\Delta q = [sq(1-q)]/[1-s(1-q)]$  or with sufficiently close approximation  $\Delta q = sq(1-q)$  if the selection coefficient is small.

A second approximation may be obtained by considering the zygotic frequencies. Assume that the types  $aa$ ,  $Aa$ , and  $AA$  reproduce in the ratio  $(1-s') : (1-hs') : 1$  per generation. The change in the frequency of  $A$  to a sufficiently close approximation is  $\Delta q = s'q(1-q)[1-q+h(2q-1)]$  (WRIGHT 1929). In the case of selection for or against a complete recessive ( $h=0$ ,  $s'$  negative or positive respectively),  $\Delta q = s'q(1-q)^2$ .

The case of no dominance ( $h = \frac{1}{2}$ ) is the same as the case of genic selection except that the selection against the gene is  $s'/2$  instead of  $s$ .

The two factor case in which the phenotypes  $aabb$ ,  $aaB-$ ,  $A-bb$  and  $A-B-$  reproduce at the rates  $(1-s_{ab}) : (1-s_a) : (1-s_b) : 1$  respectively yields (for low values of the selection coefficients):

$$\Delta q_A = q_A(1 - q_A)^2 [s_a + (s_{ab} - s_a - s_b)(1 - q_B)^2].$$

The frequency of  $A$  depends on the frequency and selection of  $B$ , becom-

ing independent only if  $s_{ab} = s_a + s_b$ , that is, if the two series of genes are cumulative with respect to selection. It does not seem profitable to pursue this subject further for the purpose of the present paper, since in the general case, each selection coefficient is a complicated function of the entire system of gene frequencies and can only be dealt with qualitatively. Attention may, however, be called to HALDANE'S (1924-1927) studies of selection rates and of the consequent number of generations required for unopposed selection to bring about any required change in gene frequency under various assumptions with respect to mode of inheritance and system of mating.

### *Equilibrium under selection*

There may be equilibrium between allelomorphs as a result wholly of selection, namely, selection against both homozygotes in favor of the heterozygous type. Putting  $\Delta q = s'q(1-q)[1-q+h(2q-1)] = 0$  gives  $q = [1-h]/[1-2h]$  as the condition.<sup>2</sup> This includes the case of selection against both homozygotes and also that in favor of them, but examination of the signs of  $\Delta q$  above and below the equilibrium point shows that only the former is in stable equilibrium in agreement with FISHER (1922). The linkage of a favorable dominant with an unfavorable recessive of another series is a case in which selection would be against both homozygotes as JONES (1917) has pointed out, and stressed as a factor in the vigor of heterozygosis. In a population produced by the intermingling of types in which different deleterious recessives have become fixed, there will be a temporary selection in favor of the heterozygotes even without any linkage at all. Unless the linkage is very strong, however, this effect does not persist long enough to have much effect on gene equilibrium. The extreme case of equilibrium of the sort discussed here is, of course, that of balanced lethals, found in nature in *Oenothera*.

In the two factor case, discussed in the preceding section,

$$\Delta q_A = 0 \text{ if } q_B = 1 - \sqrt{\frac{s_a}{s_a + s_b - s_{ab}}}$$

$$\text{and } \Delta q_B = 0 \text{ if } q_A = 1 - \sqrt{\frac{s_b}{s_a + s_b - s_{ab}}}$$

<sup>2</sup> The condition can be expressed in a more symmetrical form by using a different form of statement of the selection coefficients. Assume that the rates of reproduction of the three types aa, Aa and AA are as  $(1-s_a):1:(1-s_A)$ . The value of  $q$  at equilibrium comes out  $q = \frac{s_a}{s_A + s_a}$  with stable equilibrium only for positive values of the two selection coefficients.

There may be equilibrium here, if  $s_a$  and  $s_b$  are alike in sign, and  $s_{ab}$  is either opposite in sign or of the same sign and smaller, but it is an unstable equilibrium. Of more general importance, perhaps, is the equilibrium reached by a deleterious mutant gene. For mutation opposed by genic selection  $\Delta q = -uq + sq(1 - q) = 0$ ,  $q = 1 - u/s$ . For mutation opposed by zygotic selection (aa, Aa and AA reproducing at rates  $(1 - s') : (1 - hs') : 1$  it is easily shown that  $q = 1 - u/hs'$  (WRIGHT 1929), unless  $h$  approaches 0. Thus with no dominance,  $q = 1 - 2u/s'$ , and for selection against a dominant mutation,  $q = 1 - u/s'$ . The important case of selection against a recessive is that in which  $h = 0$ . The formula becomes  $q = 1 - \sqrt{u/s'}$ . All of these cases are illustrated in figure 1 in which the ordinates show the selection pressure as related to factor frequency, under the different conditions of selection. The intersections with the straight line representing mutation pressure give the points of equilibrium. If deleterious dominant and recessive mutations occur with equal frequency and are subject to the same degree of selection, the frequency of the recessive mutant genes will be greater than that of the dominant ones in the ratio  $\sqrt{u/s'}$  to  $u/s'$ . The corresponding figure for factors lacking dominance is  $2u/s'$ , where  $s'$  is the selection against the homozygote. These considerations alone should lead to a marked correlation in nature between recessiveness and deleterious effect. This correlation is further increased by the greater frequency of recessive mutation which seems to be a general phenomenon. It is this correlation which gives the theoretical basis for the immediate degeneration which usually accompanies inbreeding, a process which increases the proportion of recessive phenotypes.

The amount to which gene frequency in a subgroup may depart from the species average as a result of local selection held in check by population interchange with other regions may be calculated by solving the quadratic  $\Delta q = sq(1 - q) - m(q - q_m) = 0$ . If the local selection coefficient is much greater than the proportion of migration ( $s > m$ ),  $q = 1 - \frac{m}{s}(1 - q_m)$  or  $-mq_m/s$  depending on the direction of selection, formulae analogous to those for the equilibrium between mutation and selection. If, on the other hand, selection is weak compared with migration ( $s < m$ ), the departure from  $q_m$  is small and  $q = q_m [1 + \frac{s}{m}(1 - q_m)]$ . This case is doubtless the more important in nature. Large subgroups living under different selection pressures should show gene frequencies clustering about the average according to this expression. The effect of small size of the subgroups in bringing about random deviation in this and other cases is not here con-

sidered. The case in which  $s$  and  $m$  are of the same order of magnitude may be illustrated by the case of exact equality. Here  $q = \sqrt{q_m}$  or  $1 - \sqrt{1 - q_m}$  depending on the direction of selection.

*Multiple allelomorphs*

The foregoing discussion has dealt formally only with pairs of allelomorphs, a wholly inadequate basis for consideration of the evolutionary pro-

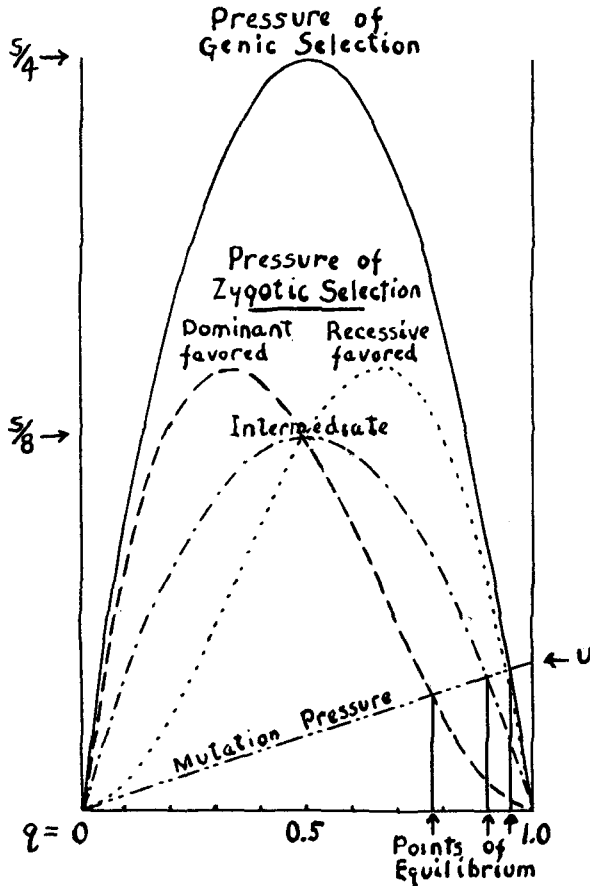


FIGURE 1.—Rate of change of gene frequency under selection or mutation. Genic selection (A,  $a$  reproducing at rates  $1:1-s$ ); Zygotic selection: dominant (B-,  $bb$  at rates  $1:1-s$ ), recessive ( $cc$ ,  $C-$  at rates  $1:1-s$ ), intermediate (DD,  $Dd$ ,  $dd$  at rates  $1:1-\frac{1}{2}s:1-s$ ), Mutation pressure such that  $u = -0.05 s$ . Intersections of line of mutation pressure with those for selection pressure determine the equilibrium frequencies.

cess unless extension can be made to multiple allelomorphs. Among the laboratory rodents some 40 percent of the known series of factors affect-



ing coat color are already known to be multiple. The number of multiple series is large in other organisms, for example, *Drosophila* (MORGAN, STURTEVANT and BRIDGES 1925). It is not unlikely that further study will indicate that all series are potentially multiple. In this case, each gene has a history which is not a mere oscillation between approximate fixation of two conditions but a real evolutionary process in its internal structure. Presumably any particular gene of such an indefinitely extended series can arise at a step from only a few of the others<sup>3</sup> and in turn mutate to only a few. Since genes as a rule have multiple effects and change in one effect need not involve others, it is probable that in time a gene may come to produce its major effects on wholly different characters than at first. Continuing this line of thought, it indeed seems possible that all genes of all organisms may ultimately be traced to a common source, mitotic irregularities furnishing the basis for multiplication of genes.

The relative frequencies of all allelomorphs in a series tend, of course, to remain constant in the absence of disturbing forces. The zygotes reach the equilibrium of random combination of the genes in pairs by the second generation from any initial constitution of the population. The effects of the various kinds of evolutionary pressure on the frequency of each gene may be treated as before by contrasting each gene with the totality of its allelomorphs. In the binomial expression  $[(1-q)a + qA]$ ,  $A$  may be understood as representing any gene, and  $a$  as including all others of its series. Such treatment, however, requires further qualification with regard to the constancy of the various coefficients. It may still be assumed that the rate ( $u$ ) of mutational breakdown of the gene in question ( $A$ ) is reasonably constant, but its rate ( $v$ ) of mutational origin from allelomorphs must be expected to change. This may be expected to rise to a maximum, as genes closely allied to  $A$  in structure become frequent, and to fall off to zero as changes accumulate in the locus. Even at its maximum, however, the rate of formation should in general be of the second order compared with the rate of change to something else, simply because it is one and its alternatives many. Moreover, there is an indication that the genes which become more or less established in a population are not a random sample of the types of mutations which occur. It has been the common experience that mutations are usually recessive. Recessiveness is most simply interpreted physiologically as due to inactivation which may well be the commonest type of mutational change. But the evolutionary process presumably involves in-

<sup>3</sup> Those most closely related genetically, however, need not always be closest in effect. The complete inactivation of a gene in a particular respect may for example occur more freely than a partial inactivation.

crease in activity of genes at least as frequently as inactivation with the consequence that the rate of formation ( $v$ ) of genes of evolutionary significance becomes negligibly small in comparison with rate of breakdown ( $u$ ) of such genes. It should be said that FISHER has advanced an alternative hypothesis according to which genes originally without dominance become dominant through a process of selection of modifiers (FISHER 1928, 1929, WRIGHT 1929, 1929a).

The selection coefficient,  $s$ , relating to a gene A cannot be expected to be constant if the alternative term  $a$  includes more than one gene. The coefficient should rise to a maximum positive value as A replaces less useful genes but should fall off and ultimately become negative as the group of allelomorphs comes to include still more useful genes. But as already discussed, even if A has only one allelomorph, the dependence of the selection coefficient on the frequencies and selection coefficients of non-allelomorphs keeps it from being constant. The existence of multiple allelomorphs merely adds another cause of variation.

#### *Random variation of gene frequency*

There remains one factor of the greatest importance in understanding the evolution of a Mendelian system. This is the size of the population. The constancy of gene frequencies in the absence of selection, mutation or migration cannot for example be expected to be absolute in populations of limited size. Merely by chance one or the other of the allelomorphs may be expected to increase its frequency in a given generation and in time the proportions may drift a long way from the original values. The decrease in heterozygosis following inbreeding is a well known statistical consequence of such chance variation. The extreme case is that of a line propagating by self fertilization which may be looked upon as a self contained population of one. In this case, 50 percent of the factors with equal representation of two allelomorphs (that is, in which the individual is heterozygous) shift to exclusive representation of one of the allelomorphs in the following generation merely as a result of random sampling among the gametes. From the series of fractions given by JENNINGS (1916) for the change in heterozygosis under brother-sister mating (population of two) it may be deduced that the rate of loss in this case is a little less than 20 percent per generation. A general method for determining the decrease in heterozygosis under inbreeding has been presented in a previous paper (WRIGHT 1921). It can be shown from this that there is a rate of loss of about  $1/2N$  in the case of a breeding population of  $N$  individuals whether equally divided between males and females or composed of monoecious individ-

uals, assuming pairs of allelomorphs. HAGEDOORN (1921) has urged the importance of such random fixation as a factor in evolution.

Another phase of this question was opened by FISHER (1922) who attempted to discover the distribution of gene frequencies ultimately reached in a population as a result of the above process. He studied a number of conditions relative to mutation and selection. He does not state the rate of decrease in heterozygosis (where any) which would follow from the solutions which he reached but this can be deduced very directly from them. It comes out  $1/4N$  for a population of  $N$  breeding individuals in the absence of selection or mutation. This is just half the rate indicated by the method referred to above.

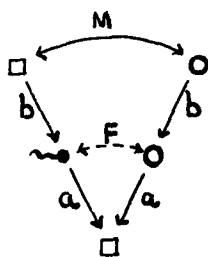


FIGURE 2.

### *Rate of decrease in heterozygosis*

The following symbols and formulae were used in the previous paper in determining the consequences of systems of inbreeding. Primes were used to indicate the number of generations preceding the one in question. Only pairs of allelomorphs are considered here.

$M$  correlation between genotypes of mates

$b$  ( $=\sqrt{\frac{1}{2}(1+F')}$ ) path coefficient, zygote to gamete

$a$  ( $=\sqrt{\frac{1}{2(1+F)}}$ ) path coefficient, gamete to fertilized egg

$F$  ( $=b^2M$ ) correlation between uniting egg and sperm, also, total proportional change in heterozygosis.

$P$  ( $=2q(1-q)(1-F)$ ) proportion of heterozygosis.

The general formula for the correlation between uniting gametes is easily deduced and has been used as a coefficient of inbreeding in dealing with complex livestock pedigrees (WRIGHT 1922, 1923, 1925, MCPHEE and WRIGHT 1925, 1926),

$$F = \Sigma \left[ \left( \frac{1}{2} \right)^{n_s + n_d + 1} (1 + F_A) \right].$$

Here  $F_A$  is the coefficient of inbreeding of any common ancestor that

makes the connecting link between a line of ancestry tracing back from the sire and one tracing back from the dam. The numbers of generations from sire and dam to such a common ancestor are designated  $n_s$  and  $n_d$  respectively. The contribution of a particular tie between the pedigrees of sire and dam is  $(\frac{1}{2})^{n_s+n_d+1} (1 + F_A)$  and the total coefficient is simply the sum of all such contributions. This formula makes it possible to compare quantitatively the statistical situation in actual populations with that in ideal populations.

In dealing with regular systems of mating the method of analysis consists in expressing the correlation between mated individuals in terms of path coefficients and correlations pertaining to the preceding generation ( $M = \phi(a, 'b, 'M')$ ) and from this obtaining expression for  $F$  in terms of the  $F$ 's of the preceding generations.

Consider a population in which there are  $N_m$  breeding males and  $N_f$  breeding females, and random mating. The proportion of matings between full brother and sister will be  $1/(N_m N_f)$ , that between half brother and sister  $(N_m + N_f - 2)/(N_m N_f)$ , and that between less closely related individuals  $(N_m - 1)(N_f - 1)/(N_m N_f)$ . The correlation between mated individuals may be written as follows, giving due weight to these three possibilities:

$$M = a'^2 b'^2 \left[ \frac{1}{N_m N_f} (2 + 2M') + \frac{N_m + N_f - 2}{N_m N_f} (1 + 3M') + \frac{(N_m - 1)(N_f - 1)}{N_m N_f} 4M' \right].$$

This leads to the following formula for proportional change in heterozygosis since the foundation period:

$$F = F' + \left( \frac{N_m + N_f}{8N_m N_f} \right) (1 - 2F' + F'').$$

The proportion of heterozygosis may be written, relative to that of preceding generations:

$$P = P' - \left( \frac{N_m + N_f}{8N_m N_f} \right) (2P' - P'').$$

It is to be expected that the proportional change per generation will reach approximate constancy. This rate may be found by equating  $P/P'$  to  $P/P''$

$$-\frac{\Delta P}{P'} = \frac{1}{2} \left( 1 + \frac{N_m + N_f}{4N_m N_f} \right) - \frac{1}{2} \sqrt{1 + \left( \frac{N_m + N_f}{4N_m N_f} \right)^2}.$$

This gives  $(1/8N_m + 1/8N_f)(1 - 1/8N_m - 1/8N_f)$  as a close approximation even for the smallest populations while for reasonably large ones the form  $1/8N_m + 1/8N_f$  is sufficiently accurate.

The simplest case is that of continued mating of brother with sister ( $N_m = N_f = 1$ ). The rate of loss of heterozygosis comes out  $\frac{1}{4}(3 - \sqrt{5})$  or 19.1 percent per generation. The formula for proportion of heterozygosis takes the form  $P = \frac{1}{2}P' + \frac{1}{4}P''$  as given in the previous paper, with results in exact agreement with those derived by JENNINGS (1916) by working out in detail the consequences of every possible mating from generation to generation.

Another simple case is that in which one male is mated with an indefinitely large number of half-sisters. This is approximately the system of breeding continuously within one herd, headed always by just one male. In this case  $N_m = 1$ ,  $N_f \doteq \infty$ , with rate of loss of heterozygosis of 11.0 percent per generation in agreement with previous results (WRIGHT 1921).

With a relatively limited number of males but unlimited number of females, the rate becomes approximately  $1/8N_m$ .

An especially important case is that in which the population is equally divided between males and females. Here  $N_m = N_f = \frac{1}{2}N$  and the rate of loss is approximately  $1/2N$  (or somewhat more closely  $1/(2N + 1)$ ) where  $N$  is the total size of the breeding population.

It is not, perhaps, clear at first sight that a population of  $N$  monoecious organisms, in which self fertilization is prevented, should show a decrease in heterozygosis exactly equal to that in a population of the same size equally divided between males and females. The chance that uniting gametes come from full sisters is  $2/[N(N - 1)]$ , the chance that they come from half sisters is  $4(N - 2)/[N(N - 1)]$  while the chance that they come from less closely related individuals is  $(N - 2)(N - 3)/[N(N - 1)]$  giving

$$M = \frac{a'^2 b'^2}{N(N - 1)} [2(2 + 2M') + 4(N - 2)(1 + 3M') + (N - 2)(N - 3)4M']$$

$$P = P' - \frac{1}{2N}(2P' - P'')$$

exactly as in the preceding case.

The somewhat arbitrary case in which the gametes produced by  $N$  monoecious individuals unite wholly at random is that which can be compared directly with FISHER'S results. The gametes have a chance  $1/N$  of coming from the same individual and of  $(N - 1)/N$  of coming from different individuals. The correlation between uniting gametes may thus be written

$$F = \frac{1}{N}b^2 + \left(\frac{N-1}{N}\right)4b^2a'^2F'$$

$$P = \frac{(2N-1)}{2N}P'$$

As might be expected, the result does not differ appreciably from that of the preceding case. The rate of loss of heterozygosis is exactly  $1/2N$  instead of merely approximately this figure. The simplest special case is, of course, continued self fertilization in which  $N=1$  and the formula gives the obviously correct result of 50 percent loss of heterozygosis per generation.

From the mode of analysis it might be thought that the loss in heterozygosis is wholly the consequence of the occasional matings between very close relatives. This, however, is not the case. If instead of random sampling of the gametes produced by the population it is assumed that all individuals reproduce equally and that inbreeding is consistently avoided as much as possible, the percentage of heterozygosis still falls off. The rate of loss is, however, only about half as rapid (approximately  $1/4N$ ) in a reasonably large population equally divided between males and females. The cases of  $N=2, 4, 8$  and  $16$  have been given previously (WRIGHT 1921).

In dealing with heterozygosis in the foregoing, it has been assumed for simplicity that each locus was represented by only two allelomorphs in the population in question and that either complete fixation or complete loss of a particular gene means homozygosis of all individuals with respect to the locus. But in any case beyond that of self fertilization, more than two allelomorphs may be present and complete loss of the gene no longer implies homozygosis of the locus. The initial rate of loss of heterozygosis in a large population may thus be only  $1/4N$  with gradual approach to  $1/2N$  as the number of loci with only two remaining allelomorphs increases. The rate of decay of the distribution of gene frequencies is  $1/2N$  regardless of number of allelomorphs.

#### *The population number*

It will be well to discuss more fully, before going on, what is to be understood by the symbol  $N$  used here for population number. The conception is that of two random samples of gametes,  $N$  sperms and  $N$  eggs, drawn from the total gametes produced by the generation in question ( $N/2$  males and  $N/2$  females each with a double representation from each series of allelomorphs). Obviously  $N$  applies only to the breeding population and not to the total number of individuals of all ages. If the population fluctu-

ates greatly, the effective  $N$  is much closer to the minimum number than to the maximum number. If there is a great difference between the number of mature males and females, it is closer to the smaller number than to the larger. In fact, as just shown, a population of  $N_m$  males and an indefinitely large number of females is approximately equivalent to a population of  $4N_m$  individuals equally divided between males and females.

The conditions of random sampling of gametes will seldom be closely approached. The number of surviving offspring left by different parents may vary tremendously either through selection or merely accidental causes, a condition which tends to reduce the effective  $N$  far below the actual number of parents or even of grandparents. How small the effective  $N$  of a population may be is indicated by recent studies of SMITH and CALDER (1927) on the Clydesdale breed of horses in Scotland, in which they find a steady increase in the degree of inbreeding (Coefficient  $F$ ) equivalent to that in population headed by only about a dozen stallions. Even more striking is the rapid increase in the coefficient of inbreeding in the early history of the Shorthorn breed of cattle (McPHEE and WRIGHT 1925).

#### THE DISTRIBUTION OF GENE FREQUENCIES AND ITS IMMEDIATE CONSEQUENCES

##### *No mutation, migration or selection*

On making a cross between two homozygous strains a population is produced in which the members of each pair of allelomorphs in which the strains differ are necessarily equally numerous. The proportional frequency of each allelomorph in unfixed series is  $q = 0.50$ . In an indefinitely large population, there should be no change in this frequency in later generations (except by recurrent mutation or selection). In any finite population, however, some genes will come to be more frequent than their allelomorphs merely by chance. This means a decrease in heterozygosis, since the proportion of heterozygosis under random mating is  $2q(1-q)$ , and this quantity is maximum when  $q = 0.50$ . As time goes on, divergences in the frequencies of factors may be expected to increase more and more until at last some are either completely fixed or completely lost from the population. The distribution curve of gene frequencies should, however, approach a definite form if the genes which have been wholly fixed or lost are left out of consideration. This can easily be seen by considering a case opposite in a sense to that considered above. Suppose that a large number of different mutations occur in a previously pure line. The frequency ratio of mutant to type allelomorphs is initially  $(1/2N):(2N-1)/2N$  where  $N$  is the number of individuals. The great majority of such muta-

tions will be lost, by the chances of sampling, as FISHER (1922) points out. Those which persist are largely those for which there has been a chance increase in frequency. The distribution curve of frequencies of persisting mutations will thus continually spread toward higher frequencies. There must be a position of equilibrium as far as form is concerned between this situation and that first considered, although a uniform decline in absolute numbers.

As noted above, decrease in heterozygosis takes place in the early generations following a cross without any appreciable fixation or loss of genes. But after equilibrium has been reached in the form of the distribution curve, further loss in heterozygosis must be identical in rate with fixation plus loss.

In simple cases, the equilibrium distribution of gene frequencies can easily be worked out directly. Under brother-sister mating, for example, the following relative frequencies of the 4 possible types of mating involving unfixd factors are in equilibrium although the absolute frequencies of all are falling off 19.1 percent ( $=\frac{1}{4}(3-\sqrt{5})$ ) each generation as new genes enter the fixed states, AA  $\times$  AA or aa  $\times$  aa.

<i>Mating</i>	<i>Relative Frequency</i>
	Percent
AA $\times$ Aa	$7 - 3\sqrt{5} = 29.2$
Aa $\times$ Aa	$-22 + 10\sqrt{5} = 36.1$
AA $\times$ aa	$9 - 4\sqrt{5} = 5.6$
Aa $\times$ aa	$7 - 3\sqrt{5} = 29.2$
	100.1

Similarly in populations of 2 and 3 monoecious individuals with random union of gametes, the following relative frequencies are in equilibrium although the absolute frequencies are decreasing in each generation by exactly 25 percent and  $16\frac{2}{3}$  percent respectively verifying the  $1/2N$  of theory.

<i>Gene Frequency</i>	<i>Percent</i>	<i>Class Frequency</i>
3A:1a	32	Case of 2 monoecious individuals per generation
2A:2a	36	
1A:3a	32	
<i>Gene Frequency</i>	<i>Percent</i>	<i>Class Frequency</i>
5A:1a	18.3	Case of 3 monoecious individuals per generation
4A:2a	21.0	
3A:3a	21.4	
2A:4a	21.0	
1A:5a	18.3	
	100.0	



In order to determine generally the distribution of gene frequencies, consider the way in which genes (A) with frequency  $q$  are distributed after one generation of random mating. In a population of  $N$  breeding individuals, each of the specified genes will have  $2Nq$  representatives among the zygotes and their allelomorphs  $2N(1-q)$ . A random sample of the same size will be distributed according to the expression  $[(1-q)a + qA]^{2N}$ . The contribution of this sample to the frequency class with an allelomorphic ratio of  $q_1:(1-q_1)$  will be in proportion to the  $2Nq_1$ 'th term of the above expression and to the number of genes included in the contributing class ( $f$ ). The sum of contributions from all such classes should give the  $2Nq_1$ 'th term an absolute frequency smaller than its value in the preceding generation ( $f_1$ ) by the amount  $1/(2N+1)$  deduced above. Following is the equation to be solved for  $f$  as a function of  $q$ .

$$f_1 \left( 1 - \frac{1}{2N+1} \right) = \frac{\frac{2N}{2Nq_1} \sum q^{2Nq_1} (1-q)^{2N(1-q_1)} f}{\frac{2N}{2N(1-q_1)}}$$

Replacing summation by integration and letting  $f = \phi(q)/2N = \phi(q)dq$  we have<sup>4</sup>

$$\frac{\phi(q_1)}{2N+1} = \frac{\frac{2N}{2Nq_1} \int_0^1 q^{2Nq_1} (1-q)^{2N(1-q_1)} \phi(q) dq}{\frac{2N}{2N(1-q_1)}}$$

The cases of 2 and 3 monoecious individuals as worked out by simple algebra suggest an approach to a uniform distribution. As a trial let  $\phi(q) = C$ . It will be found that this makes the right and left members of the equation identical and is thus a solution.

$$\frac{C}{2N+1} = \frac{C \frac{2N}{2Nq_1} \frac{\Gamma(2Nq_1+1)\Gamma(2N-2Nq_1+1)}{\Gamma(2N+2)}}{\frac{2N}{2N(1-q_1)}} = \frac{C}{2N+1}$$

The case of loss at rate  $1/2N$  should not differ appreciably from that at rate  $1/(2N+1)$ . It would appear that after a cross, the gene frequencies will spread out from 50 percent toward fixation and loss until a practically uniform distribution is reached. The frequencies of all classes will then slump at a rate of about  $1/2N$  as  $1/4N$  of the genes become fixed and the same number lost per generation. Figure 3 is intended to illustrate this situation.

<sup>4</sup>  $f$  must be equated to  $\phi(q)/2N$  here, rather than  $\phi(q)/(2N-1)$ , if the convenient limits 0 and 1 are to be used for integration in place of the limits  $1/2N$  and  $(2N-1)/2N$  of the summation with its  $2N-1$  terms.

Before finally accepting this solution, however, it will be well to examine the terminal conditions. The amount of fixation at the extremes if  $N$  is large can be found directly from the Poisson series according to which the chance of drawing 0 where  $m$  is the mean number in a sample is  $e^{-m}$ . The contribution to the 0 class will thus be  $(e^{-1} + e^{-2} + e^{-3} \dots)f = \frac{e^{-1}}{1 - e^{-1}} f, = 0.582f$ .

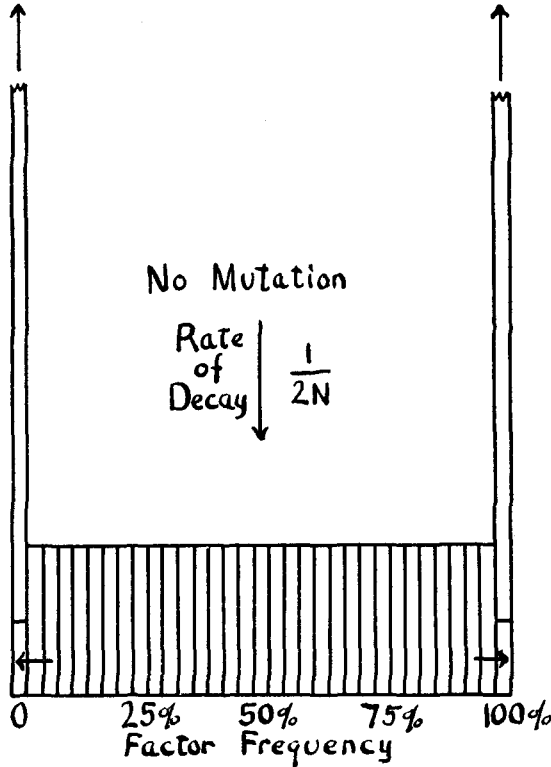


FIGURE 3.—Distribution of gene frequencies in an isolated population in which fixation and loss of genes each is proceeding at the rate  $1/4N$  in the absence of appreciable selection or mutation.  $y = L_0 e^{-T/2N}$ .

This is a little larger than the  $\frac{1}{2}f$  deduced above and indicates a small amount of distortion near the ends due to the element of approximation involved in substituting integration for summation. The nature and amount of this distortion are indicated by the exact distributions obtained in the extreme cases of only 2 and 3 monoecious individuals.

Letting  $L_0$  be the initial number of unfixed loci (pairs of allelomorphs) and  $T$  the number of generations we have approximately

Unfixed loci in the  $T$ 'th generation  $L_T = L_0 e^{-T/2N}$

An analogous formula holds for genes in multiple series, but in this case, as previously noted, the rate of fixation of loci is only half that given above.

The amount of genetic variation with respect to cumulative characters is easily calculated assuming for simplicity pairs of allelomorphs. The contribution of each factor pair to variance, in the case of no dominance, is  $2a^2q(1-q)$  where  $a$  is the average difference in effect between plus and minus allelomorphs. The general formula for variance in this or any other distribution is thus  $\sigma^2 = 2a^2 \int_0^1 q(1-q)\phi(q) dq$ . In the present case in which  $\phi(q) = L$  this reduces to  $\sigma^2 = \frac{1}{3}La^2$ . PEARSON'S  $\beta_2$  comes out with a value 2.8 a slightly platykurtic distribution. Since the percentage of heterozygosis for a given factor frequency,  $q$ , is  $2q(1-q)$ , the formula for heterozygosis is the same as that for variance except that  $a^2$  is to be omitted.

Similarly in the case of dominance, the contribution of a single factor pair to variance is  $4a^2(1-q)^2(2q-q^2)$  where  $a$  is half the average difference in effect between dominant and recessive zygotes. The total variance with perfect dominance is thus in general

$$\sigma^2 = 4a^2 \int_0^1 (1-q)^2(2q-q^2)\phi(q) dq.$$

In the case of a uniform distribution this gives  $\sigma^2 = 8/15 La^2$ .

FISHER (1918) has emphasized the importance of a characteristic of the population which he calls the dominance ratio. He analyzes the variance of characters into three portions, that due to genetic segregation ( $\tau^2$ ) that due to dominance, as something which causes deviations of the phenotype from the closest possible linear relation with the genotype ( $\epsilon^2$ ), and that due to environment. Assuming environment constant,  $\sigma^2 = \tau^2 + \epsilon^2$ . The simple formulae for the correlations between relatives, to be found if there is random mating and no dominance, must be modified, if dominance is present depending on the value of the dominance ratio defined as  $\epsilon^2/\sigma^2$ . Following are examples which he gives:<sup>5</sup>

<sup>5</sup>The author wishes here to correct an error in his 1921 paper which was written without knowledge of FISHER'S results cited above. In this paper it was assumed that the correlation with no dominance needed merely to be multiplied by the squared correlation between genotype and phenotype, the same as FISHER'S  $\tau^2/\sigma^2 = (1 - \epsilon^2/\sigma^2)$ , to obtain that with dominance. This gives correct results (if there is no assortative mating) in the case of offspring with parents, all other ancestors and also in the case of collaterals where one of the individuals is related to the other through only one parent but it is more or less in error in other cases, the fraternal correlation being that most affected. The reasoning followed was not exact because a correlation in the deviations due to dominance in the cases indicated was overlooked.

	Correlation	
	No dominance	Dominance
Parent and offspring	$\frac{1}{2}$	$\frac{1}{2}(1 - (\epsilon^2/\sigma^2))$
Brothers	$\frac{1}{2}$	$\frac{1}{2}(1 - (\epsilon^2/2\sigma^2))$
Uncle and nephew	$\frac{1}{4}$	$\frac{1}{4}(1 - (\epsilon^2/\sigma^2))$
Double first cousins	$\frac{1}{4}$	$\frac{1}{4}(1 - (3\epsilon^2/4\sigma^2))$

FISHER has shown that the contribution of a single factor to  $\epsilon^2$ , if there is complete dominance, may be written  $\delta^2 = 4q^2(1-q)^2a^2$  where  $q$  is the frequency for dominant allelomorphs. Whether a particular dominant gene has a plus or minus effect on the character under consideration is immaterial. The contribution due to genetic segregation he gives as  $\beta^2 = 8q(1-q)^3a^2$  thus

$$\frac{\epsilon^2}{\sigma^2} = \frac{\Sigma\delta^2}{\Sigma(\delta^2 + \beta^2)} = \frac{\int_0^1 q^2(1-q)^2\phi(q)dq}{2 \int_0^1 q(1-q)^3\phi(q)dq + \int_0^1 q^2(1-q)^2\phi(q)dq}$$

In the present case this reduces to  $\frac{1}{4}$  as given by FISHER who also obtains a uniform distribution of factor frequencies for the case of no mutation or selection, although a different rate of decay.

#### *Nonrecurrent mutation*

If mutation is occurring, however low the rate, the decline in heterozygosis, following isolation of a relatively small group from a large population, cannot go on indefinitely. There will come a time when the chance elimination of genes will be exactly balanced by new genes arising by mutation. The equation to be solved is obviously as follows:

$$\frac{\phi(q_1)}{2N} = \frac{|2N|}{|2Nq_1|} \frac{|2N|}{|2N(1-q_1)|} \int_0^1 q^{2Nq_1}(1-q)^{2N(1-q_1)}\phi(q)dq.$$

It may be found by trial that the expression  $\phi(q) = C_1q^{-1} + C_2(1-q)^{-1}$  is a solution. The terminal condition, reduction of the class of fixed genes ( $q=1$ ) by an occasional mutation (contributing to the class  $q=(2N-1)/2N$  necessarily involves the appearance of new genes (contributing to the class  $q=1/2N$ ) and therefore means that only the symmetrical solution  $\phi(q) = Cq^{-1}(1-q)^{-1}$  can be accepted as descriptive of the distribution of the entire array of genes at equilibrium (under the rather arbitrary postulated condition, no selection, no migration, no recurrence of the same mutations). Letting  $f = (C/2N) q^{-1}(1-q)^{-1}$  and making  $\Sigma f = 1$ ,

$$C = \frac{1}{2[(0.577 + \log(2N - 1))]}$$

or approximately  $C = 1/(2 \log 3.6N)$  (compare figure 5).

Before attainment of equilibrium with respect to heterozygosis the distribution will pass through phases of approximately the form  $\phi(q) = C_1q^{-1}(1-q)^{-1} + C_3$  in which the term  $C_1$  gradually displaces  $C_3$  as the number of temporarily fixed genes approaches equilibrium with mutation.

Each particular gene has a probability distribution for the future which spreads in time from the initial frequency in curves which are at first approximately normal in form but later (if the initial  $q$  was not too close to 1) become flat, the chances of complete fixation or complete loss each increasing by  $1/4N$  each generation. As the chances of complete fixation increase, the chance of mutation must be taken into account. The distribution passes through phases of the type  $C_2(1-q)^{-1} + C_3$ ,  $C_2$  gradually displacing  $C_3$ , relatively, but itself beginning to decline as the chance of complete loss increases. With initial  $q$  equal (or close) to 1, equilibrium with mutation, and hence the hyperbolic distribution, is reached directly. The ultimate result in any case is complete loss of the gene in question (still assuming no recurrence of the same mutation and hence mutation *of* the gene but not *to* it). If there is reverse mutation, but at very low rate, a term  $C_1q^{-1}$  must be added to the formula, and an equilibrium will be reached in the form  $Cq^{-1}(1-q)^{-1}$ . This last formula means that in the long run (assuming no disturbances from selection, migration, etc.) the gene will usually be found either completely fixed or completely absent from the population (with frequencies proportional to the mutation rates to and from the gene respectively) but that occasionally fixation or absence will not be quite complete and that at extremely rare intervals the gene will drift from one state to the other.

The turnover among genes in equilibrium in the distribution  $Cq^{-1}(1-q)^{-1}$  can be determined from consideration of the variance of  $q$ , and independently by application of the Poisson law.

Let  $\sigma_q^2 = \Sigma(q - \frac{1}{2})^2f / \Sigma f$  be the variance of  $q$ , excluding the terminal classes, the summation including  $2N - 1$  terms. This variance is increased in the following generation by the spreading out of each frequency class as a result of random sampling. The variance from the spreading of a single class is  $q(1-q)/2N$  and the average is thus

$$\Delta\sigma_q^2 = \frac{\Sigma q(1-q)f}{2N\Sigma f} = \frac{1}{2N} \left( \frac{1}{4} - \sigma_q^2 \right) = \frac{2N-1}{(2N)^2} C.$$

The sum  $\sigma_q^2 + \Delta\sigma_q^2$  includes the newly fixed factors whose contribution is  $\frac{1}{4}K$  where  $K$  is the rate of fixation, plus loss, but excludes mutation.

Digressing for a moment to the case of no mutation but equilibrium of form, we have at once

$$\sigma_q^2 + \Delta\sigma_q^2 = K\frac{1}{4} + (1 - K)\sigma_q^2$$

$$\left(K - \frac{1}{2N}\right) \left(\sigma_q^2 - \frac{1}{4}\right) = 0 \text{ giving an independent demonstration}$$

that the rate of decay is  $1/2N$  in this case.

Returning to the case of equilibrium under mutation, the contribution of new mutations to variance is  $K(N-1)^2/(2N)^2$ .

$$\sigma_q^2 + \Delta\sigma_q^2 - \frac{1}{4}K + K\left(\frac{N-1}{2N}\right)^2 = \sigma_q^2$$

$$K = C = \frac{1}{2 \log 3.6N}.$$

The proportion exchanged at each extreme is thus about half the corresponding subterminal class where  $N$  is large ( $f_1 = f_{2N-1} = 2NC/(2N-1)$  by this method. This compares fairly well with the proportion as determined by the Poisson law, which is 0.46 times the subterminal class instead of 0.50.

The equilibrium frequencies can be worked out algebraically in simple cases. The figures below give the results in the case of a population of 3 monoecious individuals for comparison with the theoretical values deduced above. The rate of exchange at each extreme is actually 10.8 percent in comparison with 11.0 percent as  $\frac{5}{12}\left(\frac{2N-1}{4N}\right)$  the subterminal class, or 11.4 percent from the formula  $\frac{1}{4(.577 + \log 5)}$ . The case of irreversible mutation is also given.

Gene frequency	Reversible Mutation		Irreversible Mutation	
	Exact equilibrium	$Cq^{-1}(1-q)^{-1}$	Exact	$C(t-q)^{-1}$
5A:1a	27.5	26.3	47.7	43.8
4A:2a	15.4	16.4	20.6	21.9
3A:3a	14.1	14.6	14.1	14.6
2A:4a	15.4	16.4	10.2	10.9
1A:5a	27.5	26.3	7.3	8.8
Totals	99.9	100.0	99.9	100.0
Terminal exchange	10.8	11.0	18.0	18.25
Loss			3.6	3.65

The number of unfixed loci ( $L$ ) which a given mutation rate per individual ( $\mu$ ) will support in a population is easily found, assuming only pairs of allelomorphs. The number of mutations is  $KL$  as well as  $N\mu$ . Therefore  $L = N\mu/K = 2N\mu \log 3.6N$ . The variance of cumulatively determined characters worked out as in the preceding case comes out  $2N\mu a^2$  in the case of no dominance and  $10/3 N\mu a^2$  in the case of dominance, in both cases, directly proportional to the size of population<sup>6</sup> and to the mutation rate. In view of the piling up of new mutations, one might perhaps, expect to find a leptokurtic distribution for characters. This, however, turns out not to be the case: PEARSON'S  $\beta_2$  comes out exactly 3 in the case of no dominance on substitution in the general formula

$$\beta_2 = 3 + \frac{\int_0^1 q(1-q)[1-6q(1-q)]\phi(q) dq}{\left[ \int_0^1 q(1-q)\phi(q) dq \right]^2}$$

FISHER'S dominance ratio comes out 1/5 in this case.

The preceding results differ somewhat from those presented by FISHER (1922). The latter's analysis was based on a transformation of the scale of factor frequencies designed to make the variance due to random sampling uniform at all points. The variance at a given value of  $q$  is  $q(1-q)/2N$ . FISHER assumes that if the ratio of small differences on the  $q$  scale to the corresponding differences on a new  $\theta$  scale be made proportional to the varying standard deviation of  $q$ , the standard deviation on the  $\theta$  scale will be uniform. Letting  $dq/d\theta = \sqrt{q(1-q)}$  leads to the transformation  $\theta = \cos^{-1}(1-2q)$  with uniform variance of factor frequencies of  $1/2N$ . Letting  $y = F(\theta)$  be the distribution of factor frequencies in one generation, he wrote that in the next as

$$y + \Delta y = \int_0^\pi \frac{1}{\sigma\sqrt{2\pi}} e^{-\delta\theta^2/2\sigma^2} \left( y + y'\delta\theta + \frac{\delta\theta^2}{2} y'' \dots \right)$$

and measuring time in generations ( $T$ ) he reached the equation

$$\frac{\partial y}{\partial T} = \frac{1}{4N} \frac{\partial^2 y}{\partial \theta^2}$$

After noting that the solution for the symmetrical stationary case is

<sup>6</sup> These estimates of number of unfixed loci and of variance depend, of course, on the validity of the conditions on which the formula of the distribution curve is based. How far the mutation rate per locus can be considered negligibly small as size of population increases is discussed later.

$y=L/\pi$ , he proceeded to derive the formulae for increasing and decreasing  $y$ . Considering the latter,  $dy/dT = -Ky$  where  $K$  is the rate of decay, giving  $1/4N \frac{d^2y}{d\theta^2} = -Ky$  as the equation to be solved. In the symmetrical case this yields  $y=C \cos [\sqrt{4NK}(\theta-\pi/2)]$  where  $C = \sqrt{4NK}/[2 \sin (\frac{1}{2}\pi\sqrt{4NK})]$  in order to give a total frequency of unity and is to be multiplied further by  $L_0e^{-KT}$  to show change from the initial frequency of  $L_0$ .

The maximum value which  $K$  can take without giving negative frequencies within the range is obviously  $1/4N$  and FISHER found reason for accepting this as the value in the case of no mutation. The formula for the distribution in this case reduces to  $y = \frac{1}{2} \sin \theta$ . FISHER transformed these equations to the scale  $Z = \log [q/(1-q)]$  in which the case of no mutation becomes  $y = \frac{1}{4} \operatorname{sech}^2 \frac{1}{2}Z$  and the case of loss balanced by mutation becomes  $y = 1/2\pi \operatorname{sech} \frac{1}{2}Z$ . This transformation brings the curves into an approach to the form of the normal probability curve. For our present purpose it is preferable to transform to the scale of actual factor frequencies. The case of steady decay becomes  $y = 1$  with which my results are in agreement, although in disagreement as to rate of decay. In the case of loss balanced by mutation, FISHER's formula transforms into  $y = 1/[\pi\sqrt{q(1-q)}]$  instead of  $1/[2(\log 3.6N)q(1-q)]$  as developed in the present paper. FISHER obtained  $\sqrt{\pi/2N^3/2}\mu$  for the number of unfixed factors, in contrast with  $2N\mu \log 3.6N$ ; and  $\sqrt{\frac{2}{\pi N}}$  for the factor turnover in contrast with  $1/[2 \log 3.6N]$ .

It will be seen that FISHER's solution leads to a smaller number of unfixed factors with more rapid turnover in very small populations (less than 81) but to a larger number of such factors with slower turnover in larger populations. In a breeding population of one million with one mutation per 1000 individuals, FISHER's formula gives 1,250,000 unfixed factors with a turnover of 0.08 percent while I get 30,000 unfixed factors with a turnover of 3.3 percent.

The exact harmonizing of the results of the two methods of attack has been a somewhat puzzling matter, but Doctor FISHER, on examination of the manuscript of the present paper, has written to me the following which I quote at his suggestion. "... I have now fully convinced myself that your solution is the right one. It may be of some interest that my original error lay in the differential equation

$$\frac{\partial y}{\partial T} = \frac{1}{4N} \frac{\partial^2 y}{\partial \theta^2}$$

which ought to have been

$$\frac{\partial v}{\partial T} = \frac{1}{4N} \frac{\partial}{\partial \theta} (y \cot \theta) + \frac{1}{4N} \frac{\partial^2 y}{\partial \theta^2}$$



the new term coming in from the fact that the mean value of  $\delta p$  in any generation from a group of factors with gene frequency  $p$  is exactly zero,<sup>7</sup> and consequently the mean value of  $\delta\theta$  is not exactly zero but involves a minute term  $-1/4N \cot \theta$ . With this correction, I find myself in entire agreement, with your value  $2N$ , for the time of relaxation and with your corrected distribution for factors in the absence of selection."

### *Reversible recurrent mutation*

It only requires a very moderate mutation rate in a large population for the number of unfixed loci to become enormous. This raises the question as to the effect of a limitation in the number of mutable loci, and recurrence of mutations.

Consider now the case of genes with uniform rates of recurrence of mutation and reverse mutation. Let  $u$  be the rate per generation for break down of the gene  $A$  and  $v$  that for origin from allelomorphs. A class of genes with frequency  $q$  (that of all allelomorphs,  $1-q$ ) will be distributed in the following generation under random sampling according to the expansion of the expression

$$\{[(1-q) - v(1-q) + uq]a + [q + v(1-q) - uq]A\}^{2Nf}.$$

Equating the total contribution to a given class, to the frequency of this class in the parent generation, reduced by the proportion  $K$ , if there is a uniform rate of decay, gives as the equation to be solved:

$$\phi(q_1) \frac{1-K}{2N} = \frac{|2N|}{|2Nq_1|} \frac{|2N|}{|2N(1-q_1)|} \int_0^1 [q(1-u-v) + v]^{2Nq_1} [1 - q(1-u-v) - v]^{2N(1-q_1)} \phi(q) dq.$$

It will be found by trial that the right and left members became identical in certain cases in which  $\phi(q)$  is of the form  $q^s(1-q)^t$

$$\text{Let } x = q(1-u-v) + v$$

$$q = \frac{x-v}{1-u-v} \qquad dq = \frac{dx}{1-u-v}$$

$$q^s = \frac{x^s - svx^{s-1} \dots}{(1-u-v)^s} \qquad (1-q)^t = \frac{(1-x)^t - tu(1-x)^{t-1}}{(1-u-v)^t}.$$

<sup>7</sup>  $p$  is the  $q$  of the present paper. Since the above was written, FISHER has published this revision of his results in *The genetical theory of natural selection*, 1930.

The small amount of spread from a given class will justify retention of the untransformed limits of integration.

Noting that  $\Gamma(c+s+1) = \lfloor c \rfloor c^s \left(1 + \frac{s(1+s)}{2c}\right)$  approximately when  $c$  is an integer and  $s$  is small compared with  $c$ , and making use of the following derived relation

$$\frac{2N \lfloor 2N \rfloor}{\lfloor 2Nq_1 \rfloor \lfloor 2N(1-q_1) \rfloor} \int_0^1 x^{2Nq_1+s}(1-x)^{2N(1-q_1)+t} dx \\ = \frac{4N + s(s+1)q_1^{-1} + t(t+1)(1-q_1)^{-1}}{4N + (s+t+1)(s+t+2)} q_1^s (1-q_1)^t$$

the equation may be written as follows for small values of  $s$  and  $t$  (compared with  $N$ ) and values of  $u$  and  $v$  of a still lower order of size.

$$(1-K)q_1^s(1-q_1)^t \\ = \frac{1}{1-(u+v)(s+t+1)} \left[ \frac{4N}{4N + (s+t+1)(s+t+2)} q_1^s(1-q_1)^t \right. \\ \left. + \left( \frac{s(s+1)}{4N + (s+t+1)(s+t+2)} - \frac{4Nsv}{4N + (s+t)(s+t+1)} \right) q_1^{s-1}(1-q_1)^t \right. \\ \left. + \left( \frac{t(t+1)}{4N + (s+t+1)(s+t+2)} - \frac{4Ntu}{4N + (s+t)(s+t+1)} \right) q_1^s(1-q_1)^{t-1} \right]$$

The coefficients of  $q_1^{s-1}(1-q_1)^t$  and of  $q_1^s(1-q_1)^{t-1}$  must equal 0 either under complete equilibrium or equilibrium merely in form of distribution. Neglecting small quantities:

$$s = 0 \quad \text{or} \quad s = 4Nv - 1 \\ t = 0 \quad \text{or} \quad t = 4Nu - 1.$$

In the case of complete equilibrium ( $K=0$ ), it turns out that the coefficients of  $q_1^s(1-q_1)^t$  in the left and right members are also satisfied to a first approximation by  $s=4Nv-1$ ,  $t=4Nu-1$ . They are also satisfied by letting  $s=0$ ,  $t=0$  provided that  $u+v=1/2N$ . The relation between the fixed terminal and the unfixed subterminal classes, however, requires that  $u=v=1/4N$  in this case, which thus becomes merely a special case of the first solution. Similarly, the solutions  $s=0$ ,  $t=4Nu-1$  and  $s=4Nv-1$ ,  $t=0$  require that  $v=1/4N$  and  $u=1/4N$  respectively and thus also reduce to special cases of the first solution. It appears then that the distribution of gene frequencies in equilibrium under mutation and reverse mutation may be represented approximately by curves of PEARSON'S Type I,

$$\phi(q) = \frac{\Gamma(4Nu + 4Nv)}{\Gamma(4Nu)\Gamma(4Nv)} q^{4Nv-1}(1-q)^{4Nu-1}.$$

The terminal conditions are of interest in this and other cases to be considered. The factor turnover at each extreme may be written

$$K_0 = \frac{1}{2}f_1 = 2Nvf_0$$

$$K_{2N} = \frac{1}{2}f_{2N-1} = 2Nuf_{2N}$$

where the subterminal classes have the frequencies

$$f_1 = \frac{1}{2N}\phi\left(\frac{1}{2N}\right)$$

$$f_{2N-1} = \frac{1}{2N}\phi\left(1 - \frac{1}{2N}\right).$$

In the present case, the terminal classes have the frequencies  $f_0 = C/[4Nv(2N)^{4Nv}]$  and  $f_{2N} = C/[4Nu(2N)^{4Nu}]$  where  $C$  is the coefficient in the expression for  $\phi(q)$ .

It will be seen that the form of the curve depends not only on the rates of mutation of the genes but also on the size of the breeding population. With small populations or rare recurrence of mutations, the distribution approaches the symmetrical form  $y = 1/(2 \log 3.6N) q^{-1}(1-q)^{-1}$  already discussed (figure 5). The ratio of the class of temporarily fixed genes ( $f_{2N}$ ) to the class of complete absence ( $f_0$ ) must be approximately  $v:u$  in this case in order that the number of mutations at each extreme of the symmetrical distribution of unfixed factors may be equal.

With increase in size of the population, the gene frequencies tend in general to be distributed in asymmetrical U- or even I- or J-shaped curves. For example, if the size of population reaches  $1/4u$  and  $v$  is much smaller than  $u$ , the distribution will be the hyperbola  $\phi(q) = Cq^{-1}$  with a piling up of factors with few or no plus representatives.

With sufficient increase in the size of population, the distribution at length takes a form approaching that of the normal probability curve; centered about the point  $\bar{q} = v/(u+v)$  which, indeed, is always the mean

$$\left(\bar{q} = \int_0^1 q\phi(q) dq = \frac{v}{u+v}\right).$$

The variance of gene frequencies,  $\sigma_q^2 = \int_0^1 (q-\bar{q})^2 \phi(q) dq$  is

$$\frac{\bar{q}(1-\bar{q})}{4N(u+v) + 1}$$

The amount of genetic variation of cumulative characters may be calculated as before. In the case of no dominance and paired allelomorphs

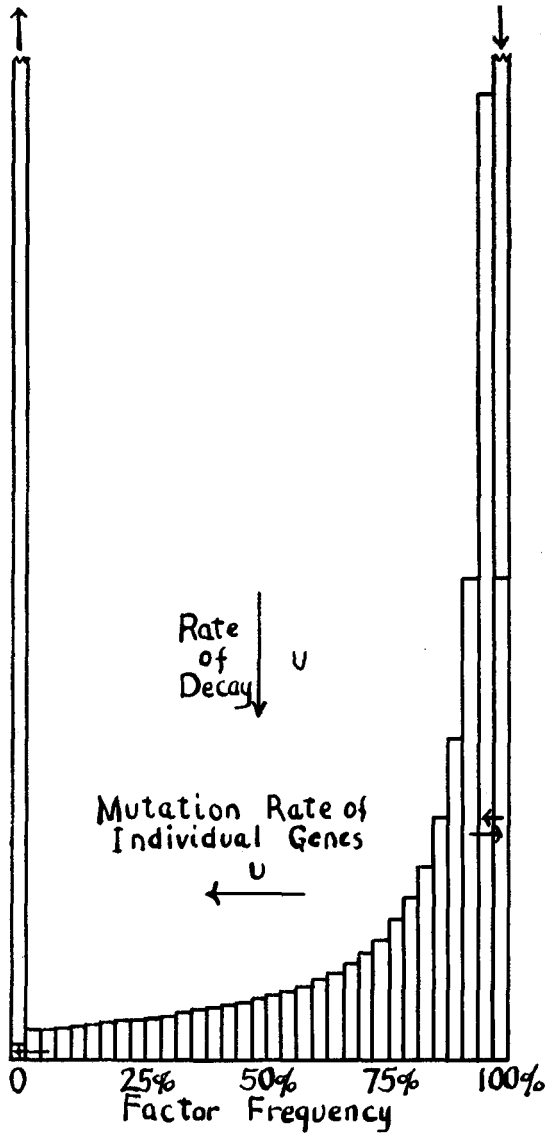


FIGURE 4.—Distribution of type genes in an isolated population in which equilibrium has been reached with destructive mutation but has not been approached with respect to formative mutation.  $y = 4NuL_0e^{-uT}(1-q)^{4Nu-1}$  with  $4Nu$  much smaller than 1 and the formula approxi-

mately  $\frac{L_0(1-q)^{-1}}{\log 3.6N}$ .

it is  $\sigma^2 = 2La^2 \frac{4Nuq}{(4Nu+4Nv+1)}$  or  $2La^2 [\bar{q}(1-q) - \sigma_q^2]$ . Where  $u$  is much greater than  $v$ , this can be written approximately  $\sigma^2 = 2La^2 \frac{4Nv}{(4Nu+1)}$  approaching  $2La^2 \bar{q}$  as a limit, as  $N$  increases and  $L$  comes to include all loci.

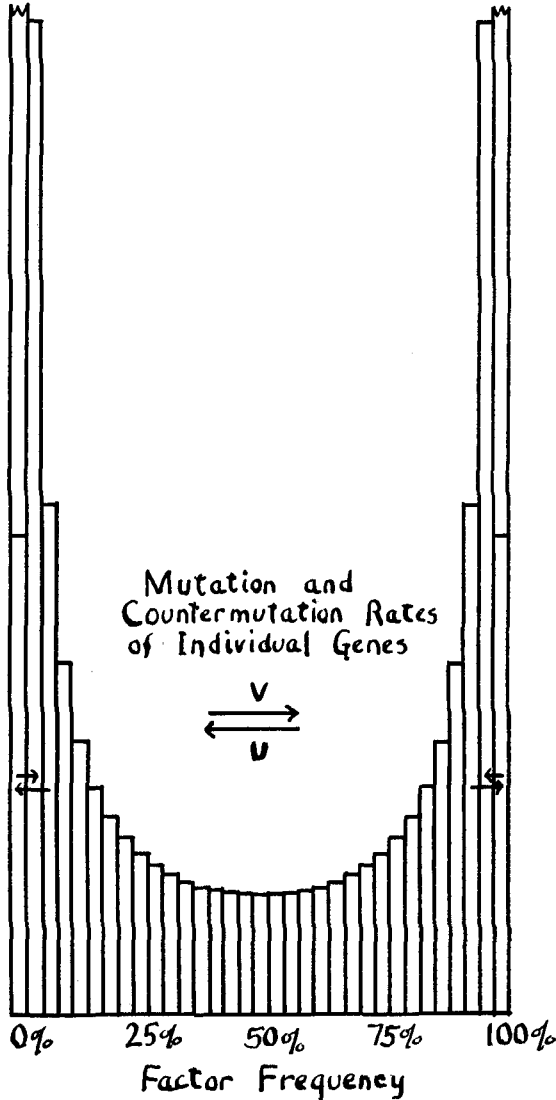


FIGURE 5.—Distribution of gene frequencies (or probability array of gene) where equilibrium with mutation has been attained. Population so small that the terms  $4Nu$  and  $4Nv$  are both much

smaller than 1.  $y = Cq^{4Nv-1}(1-q)^{4Nu-1}$ , approximately  $\frac{q^{-1}(1-q)^{-1}}{2 \log 3.6N}$ .

As the formula for this case was derived on the assumption of small values of  $u$  and  $v$ , it is desirable to obtain an independent test of its applicability to large values. This can be done as follows: the increase in variance of  $q$  due to random sampling is

$$\frac{1}{2N} \int_0^1 q(1-q)\phi(q) dq = \frac{\bar{q}(1-\bar{q}) - \sigma_q^2}{2N}.$$

Letting  $\Delta q = -uq + v(1-q)$  be, as before, the change in  $q$  due to mutation,  $q + \Delta q - \bar{q} = (q - \bar{q})(1 - u - v)$ . Thus the effect of mutation is as if all deviations from the mean were reduced in the proportion  $(1 - u - v)$ . The decrease in the variance of  $q$ , due to mutation is therefore  $\sigma_q^2 [1 - (1 - u - v)^2]$ . At equilibrium the increase in  $\sigma_q^2$  due to random sampling is exactly balanced by the decrease due to mutation yielding:

$$\sigma_q^2 = \frac{\bar{q}(1-\bar{q})}{4N(u+v) - 2N(u+v)^2 + 1}.$$

The term  $-2N(u+v)^2$  in the denominator is important only when  $(u+v)$  has a large absolute value. Omitting this, the formula is identical with that deduced by the first method and thus gives an independent demonstration of its validity. As mutation approaches its maximum value  $(u+v=1)$ , the variance of  $q$  approaches  $\bar{q}(1-\bar{q})/2N$ , that due to random sampling alone.

### *Migration*

The distribution of gene frequencies in an incompletely isolated subgroup of a large population can be obtained immediately from the preceding results. The change in gene frequency per generation under migration  $\Delta q = -m(q - q_m)$  can be written  $-m(1 - q_m)q + mq_m(1 - q)$  which is in the same form as the change of  $q$  under mutation,  $\Delta q = -uq + v(1 - q)$ . We may write at once for the distribution under negligible mutation rates:

$$\phi(q) = \frac{\Gamma(4Nm)}{\Gamma[4Nm q_m] \Gamma[4Nm(1 - q_m)]} q^{4Nm q_m - 1} (1 - q)^{4Nm(1 - q_m) - 1}.$$

The mutation terms  $4Nu$  and  $4Nv$  can be inserted, if mutation rates are not negligible.

Figure 6 shows how the form of the distribution changes with change in  $m$  or  $N$ . Where  $m$  is less than  $1/2N$  there is a tendency toward chance fixation of one or the other allelomorph. Greater migration prevents such fixation. How little interchange would appear necessary to hold a large population together may be seen from the consideration that  $m = 1/2N$

means an interchange of only one individual every other generation, regardless of the size of the subgroup. However, this estimate must be much qualified by the consideration that the effective  $N$  of the formula is in general much smaller than the actual size of the population or even than the breeding stock, and by the further consideration that  $q_m$  of the formula refers to the gene frequency of actual migrants and that a further factor must be included if  $q_m$  is to refer to the species as a whole. Taking both of these into account, it would appear that an interchange of the order of thousands of individuals per generation between neighboring subgroups of a widely distributed species might well be insufficient to prevent a considerable random drifting apart in their genetic compositions. Of course,

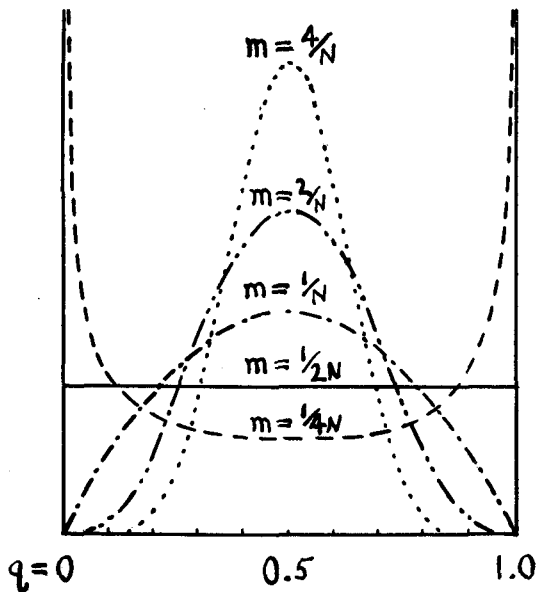


FIGURE 6.—Distribution of frequencies of a gene among subdivisions of a population in which  $q_m = 1/2$  (or probability array of gene within a subdivision) under various amounts of intermigration.  $y = Cq^{4Nm}q_m^{-1}(1-q)^{4N(1-q_m)-1}$ .

differences in the condition of selection among the subgroups may greatly increase this divergence. It appears, however, that the actual differences among natural geographical races and subspecies are to a large extent of the nonadaptive sort expected from random drifting apart. An interesting example, apparently nonadaptive, is the racial distribution of the 3 allelomorphs which determine human blood groups (BERNSTEIN 1925).

The variance of distribution of values of  $q$  among subgroups (in the ideal

case) is  $\sigma_q^2 = q_m(1 - q_m)/(4Nm + 1)$  by substitution in the formula for the preceding case.

The zygotic distribution  $[(1 - q)a + qA]^2$  cannot be expected to hold in a population made up of isolated groups among which gene frequency varies. WAHLUND (1928) has shown that the proportions in each homozygous class are increased at the expense of the heterozygotes by the amount of the variance of the gene frequencies among the subgroups,<sup>8</sup> the proportions becoming  $[(1 - q)^2 + \sigma_q^2]aa + [2q(1 - q) - 2\sigma_q^2] Aa + [q^2 + \sigma_q^2] AA$ . By substituting the expression for  $\sigma_q^2$ , given above, in WAHLUND'S formula one might determine empirically the effective value of  $4Nm$  for the population, except that it would be difficult to rule out the possibility that some of the variance of gene frequencies might be due to differences in the selection coefficients among the subgroups instead of merely to random drifting apart.

#### *Irreversible recurrent mutation*

The solution  $s = 0$ ,  $t = 4Nu - 1$  for the equation reached in the case of recurrent mutation satisfies the conditions for equilibrium of form under irreversible mutation ( $v = 0$ ), with decay at rate  $K = u$ .

$$\phi(q) = 4NuL_0e^{-uT}(1 - q)^{4Nu-1}.$$

The proportional frequency of the unfixed subterminal class which is not replenished by mutation is  $f_1/(L_0e^{-uT}) = 2u$ , twice the rate of decay and thus approximately satisfying the necessary terminal condition.

For values of  $u$  as small as  $1/(2N \log 3.6N)$  the coefficient in the expression for  $\phi(q)$  must be calculated to a closer approximation

$$\frac{4NuL_0e^{-uT}}{1 - \left(\frac{1}{2N}\right)^{4Nu}}$$

which approaches  $\frac{L}{\log 3.6N}$  as  $u$  approaches zero.

The distribution of gene frequencies under irreversible mutation is illustrated in figure 4.

This case is of most interest as representing for a long time the distribution of type genes in a small group isolated from a large one in which all type genes are close to fixation. The release of deleterious mutation pressure from equilibrium with selection will result in approximate equi-

<sup>8</sup> The percentage of heterozygotes is  $2\int_0^1 q(1 - q)\phi(q) dq$  where  $\phi(q)$  is the distribution of values of  $q$  among the subgroups. As shown above this reduces to  $2\bar{q}(1 - \bar{q}) - 2\sigma_q^2$ , thus demonstrating WAHLUND'S principle.



librium of the form described above. With decay at the rate  $u$ , it may be a very long time before effects of reverse mutation become appreciable and the final equilibrium  $y = Cq^{-1}(1-q)^{-1}$  approached. Assuming that type genes are dominant, the dominance ratio in this case is  $1/3$ .

### Selection<sup>9</sup>

Using  $\Delta q = sq(1-q)$  as the measure of the effect of genic selection, the class of genes with frequency  $(1-q)a : qA$  is distributed after one generation according to the expression:

$$\{(1-q)(1-sq)a + q[1 + s(1-q)]A\}^{2N}.$$

The distribution of gene frequencies which is in equilibrium may be obtained from the following equation which represents the total contribution to class  $q_1$  after one generation, as equal to its previous frequency.

$$\frac{\phi(q_1)}{2N} = \frac{\frac{2N}{2Nq_1} \frac{2N}{2N(1-q_1)}}{2N}$$

$$\int_0^1 q^{2Nq_1}(1-q)^{2N(1-q_1)}(1+s(1-q))^{2Nq_1}(1-sq)^{2N(1-q_1)}\phi(q) dq$$

To a first approximation, the selection terms approach the value  $e^{2Ns(q_1-q)}$ . The introduction of a factor  $e^{2Nsq}$  into the previously reached formula for  $\phi(q)$  gives a solution of the equation (for very small values of  $s$ ) since it cancels the new term  $e^{-2Nsq}$  in the integral, and leaves  $e^{2Nsq_1}$  as a factor in  $\phi(q_1)$ . This was the basis for the formula published (WRIGHT 1929a) as  $\phi(q) = Ce^{2Nsq}q^{4Nv-1}(1-q)^{4Nu-1}$  intended to exhibit in combination the effects of selection, mutation in both directions and size of population. Further consideration reveals that this solution is the correct one only for the case of irreversible mutation and then only when the selection coefficient is exceedingly small, less than  $1/2N$  in fact. FISHER (1930) in his recently published revision of the results of his method of attack on this problem has given a formula for a special case of selection, equilibrium of flux from an inexhaustible supply of mutating genes. This is given as accurate as long as  $Ns^2$  is small. Assuming one mutation per generation, he writes:

$$y = \frac{2dp(1 - e^{-4anq})}{pq(1 - e^{-4an})}.$$

<sup>9</sup>This and the following section have been rewritten since submission of the manuscript in order to take account of the correction of my formula, suggested by FISHER's results in *The genetical theory of natural selection*, 1930 as noted herein.

In this formula,  $a(= -s)$  is the selection coefficient,  $p(= 1 - q)$  is frequency of mutant genes and  $dp$  may be taken as  $1/2N$  numerically. This agrees with my previous formula for irreversible mutation,  $y = Ce^{2Ns}q(1 - q)^{-1}$  only when  $s$  is less than  $1/2N$ , above which value my formula rapidly leads to impossible results. On reexamination of my method, however, I find that the same degree of approximation can be reached by it. The expansion of  $[1 + s(1 - q)]^{2Nq_1} [1 - sq]^{2N(1 - q_1)}$  yields series of terms which condense into the expression  $e^{2Ns(q_1 - q)} \{1 - Ns^2[q_1(1 - q_1) + (q_1 - q)^2]\}$  taking into account terms in  $Ns^2, N^2s^3, N^3s^4, N^4s^5$  as well as those in which  $N$  and  $s$  have the same exponent. Since the random deviations of  $q$  have a variance of  $q_1(1 - q_1)/2N$  the term  $(q_1 - q)$  is of the order  $\sqrt{1/2N}$ . A second order approximation should be obtainable by retaining the term  $Ns^2q_1(1 - q_1)$  while that in  $Ns^2(q_1 - q)^2$  may be dropped. The equation to be solved can now be written.

$$\phi(q_1) = \frac{2N \sqrt{2N}}{\sqrt{2Nq_1} \sqrt{2N(1 - q_1)}} e^{2Ns q_1} [1 - Ns^2 q_1(1 - q_1)] \int_0^1 q^{2Nq_1} (1 - q)^{2N(1 - q_1)} e^{-2Ns q} \phi(q) dq.$$

Let  $\phi(q) = e^{2Ns q} q^{-1} (1 - q)^{-1} (a + bq + cq^2 + dq^3 \dots)$ .

The exponential term in the integral being cancelled, it becomes possible to carry out the integration by means of the approximate formula already used in the case of mutation (page 122).

$$\frac{2N \sqrt{2N}}{\sqrt{2Nq_1} \sqrt{2N(1 - q_1)}} \int_0^1 q^{2Nq_1 + z - 1} (1 - q)^{2N(1 - q_1) - 1} dq = \frac{4N + z(z - 1)q_1^{-1}}{4N + z(z - 1)} q_1^{z-1} (1 - q_1)^{-1}.$$

The resulting coefficients of the powers of  $q_1$  on the right side of the equation may now be equated separately to those of  $\phi(q_1)$ . To a sufficient approximation it turns out that  $c = \frac{(2Ns)^2}{\underline{2}} a$ ,  $d = \frac{(2Ns)^2}{\underline{3}} b$ ,  $e = \frac{(2Ns)^4}{\underline{4}} a$ ,

$$f = \frac{(2Ns)^4}{\underline{5}} b, \quad g = \frac{(2Ns)^6}{\underline{6}} a, \quad \text{etc.}$$

$$\text{Letting } C_1 = a/2 \text{ and } C_2 = \frac{b}{4Ns}$$

$$\phi(q) = 2e^{2Ns q} q^{-1} (1 - q)^{-1} [C_1 \cosh 2Ns q + C_2 \sinh 2Ns q].$$

From considerations of symmetry, it is obvious that another solution may be obtained by replacing  $q$  by  $(1-q)$  and  $s$  by  $-s$ . The full solution may be written in the form

$$\phi(q) = q^{-1}(1-q)^{-1} [C_1(e^{4Ns}q + 1) + C_2(e^{4Ns}q - 1) + C_3(1 + e^{-4Ns(1-q)}) + C_4(1 - e^{-4Ns(1-q)})].$$

The relative values of the coefficients in the case of equilibrium can be obtained by setting up the equation for the absence of flux. Each group of genes,  $f = \phi(q)dq$  tends to be shifted by the amount  $\Delta q = sq(1-q)$  in a generation. There is thus a total flux measured by  $\int_0^1 \phi(q)\Delta q dq$  unless there is counterbalancing mutation. The amount of mutation in each direction (assuming the rates of recurrence to be very small compared with  $1/4N$ ) is approximately half the respective subterminal classes, as demonstrated in the preceding cases.

$$f_1 = 2C_1 + 2sC_2 + (1 + e^{-4Ns})C_3 + (1 - e^{-4Ns})C_4$$

$$f_{2N-1} = (e^{4Ns} + 1)C_1 + (e^{4Ns} - 1)C_2 + 2C_3 + 2sC_4.$$

Since mutation moves genes from the fixed classes to the subterminal classes with gene frequencies of  $1/2N$  and  $\left(1 - \frac{1}{2N}\right)$  respectively, it creates a net flux of  $\frac{f_{2N-1}}{4N} - \frac{f_1}{4N}$  which at equilibrium should balance that due to selection

$$\int_0^1 \phi(q)\Delta q dq - \frac{f_{2N-1}}{4N} + \frac{f_1}{4N} = 0.$$

Substitution of the values given above leads to the condition  $C_1 - C_2 + C_3 + C_4 = 0$ . Under this condition the formula simplifies greatly, becoming for all values of  $s$  (of lower order than  $\frac{1}{\sqrt{N}}$ )

$$\phi(q) = Ce^{4Ns}q^{-1}(1-q)^{-1}.$$

The effect of selection in this case is perhaps best exhibited in the ratio of the classes of alternative fixed genes in the highly artificial case of equality in the rates of mutation in opposite directions. This ratio is  $e^{4Ns}$ .

More generally,  $f_0 = \frac{C}{4Nv}$  and  $f_{2N} = \frac{Ce^{4Ns}}{4Nu}$  where  $u$  and  $v$ , both assumed to

be very small compared with  $1/4N$ , are the opposing mutation rates. The number of unfixed loci (pairs of allelomorphs) takes the form

$$L = \frac{2N\mu}{f_1 + f_{2N-1}} = \frac{2N\mu}{C(e^{4Ns} + 1)}$$

where  $\mu$  is the mutation rate per individual and  $C$  is chosen so that  $\int_0^1 \phi(q) dq = 1$ . The effect of selection on the variance of cumulative characters (pairs of allelomorphs) may be seen by comparing the formula

$$\sigma^2 = 2\mu a^2 \left( \frac{e^{4Ns} - 1}{2s(e^{4Ns} + 1)} \right)$$

with the previously given form  $2N\mu a^2$  which it approaches as  $s$  approaches 0.

In the case treated by FISHER, there is assumed to be irreversible mutation at the rate of one per generation from an inexhaustible supply. As each new gene becomes fixed, it may be considered as transferred to the type class, ready to mutate to new allelomorphs in its series. Thus in place

of a return flux of  $\frac{f_1}{4N}$ , due to reversible mutation, we must write  $\frac{f_1}{2}$  (if  $v = 0$ )

$$\int_0^1 \phi(q) \Delta q dq - \frac{f_{2N-1}}{4N} + \frac{f_1}{2} = 0.$$

This is solved if  $C_1 = C_3 = C_4 = 0$  and

$$\phi(q) = C_2(e^{4Ns q} - 1)q^{-1}(1 - q)^{-1}.$$

In case the direction of mutation coincides with that of selection ( $u = 0$ ),

the mutational terms must be written  $\frac{f_{2N-1}}{2} - \frac{f_1}{4N}$  giving the solution

$$\phi(q) = C_4(1 - e^{-4Ns(1-q)})q^{-1}(1 - q)^{-1}.$$

These are identical with FISHER's results on proper choice of the coefficient.

An interesting question which FISHER has discussed, is the chance of fixation of a single mutation. This is given by the ratio of the subterminal classes in the formulae just considered. Where selection opposes

mutation,  $\frac{f_1}{f_{2N-1}} = \frac{2s}{e^{4Ns} - 1}$ , always less than  $1/2N$ . In the case of favor-

able mutations,  $\frac{f_{2N-1}}{f_1} = \frac{2s}{1 - e^{-4Ns}}$ , or approximately  $2s$ . FISHER also

gives an independent derivation of the last figure.

### *General formula*

It is of especial importance to assemble the effects of all evolutionary factors into a single formula. Unfortunately, the equation of equilibrium of class frequencies becomes rather complicated and has not yet been worked through. Presumably the form is given at least approximately by a formula of the type  $Ce^{sq}q^{4Nv-1}(1-q)^{4Nu-1}$  in the case of reversible mutation. In order that there may be no flux,  $\int_0^1 \phi(q)\Delta q dq = 0$ . It is not necessary to consider the terminal classes in this case. Thus

$$C \int_0^1 e^{sq} q^{4Nv-1} (1-q)^{4Nu-1} [-uq + v(1-q) + sq(1-q)] dq = 0.$$

Integration of the first term (that in  $-uq$ ) by parts gives an expression which is immediately solved by letting  $a=4N$ . Thus the selection term appears to be  $e^{4Nsq}$  regardless of the rates of mutation provided there is reversibility. It is approximately of this value in the case of irreversible mutation, discussed above, provided that  $s$  is considerably larger than  $1/4N$ . The conclusions based on the previously presented value  $e^{2Nsq}$  still hold,<sup>10</sup> except that they should be applied to selection intensities just half as great.

The position of the mode of the I-shaped distribution curve given when  $u$  and  $s$  are greater than  $1/4N$  can be found by equating the differential coefficient of the logarithm of the formula to zero.

$$4Ns + \frac{4Nv - 1}{q} - \frac{4Nu - 1}{1 - q} = 0.$$

When  $v$  is small but  $u$  and  $s$  are both large,  $q$  approaches the value  $1 - \frac{u}{s}$  already given as the equilibrium point. The mean would be some-

what below this point, as expected from the curvilinear relation of selection

<sup>10</sup> These conclusions were presented at the meeting of the AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE for 1929 and were summarized in the abstracts (WRIGHT 1929b).

pressure to gene frequency and in contrast with the case of equilibrium between opposing mutation pressures (but no selection) in which the mean is always the equilibrium point  $\frac{v}{u+v}$  and the mode is more extreme than

this figure,  $\frac{4Nv-1}{4N(u+v)-2}$ .

Migration pressure introduces no other complications. Combining all factors:

$$\phi(q) = Ce^{4Ns}q^{4N(mq_m+v)-1}(1-q)^{4N[m(1-q_m)+u]-1}.$$

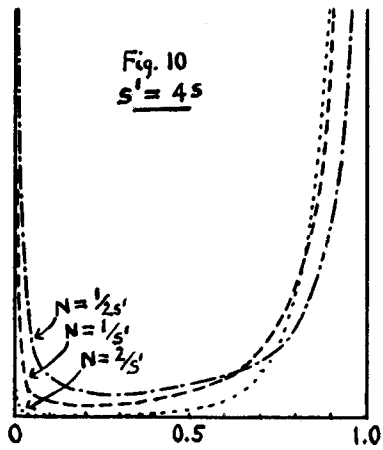
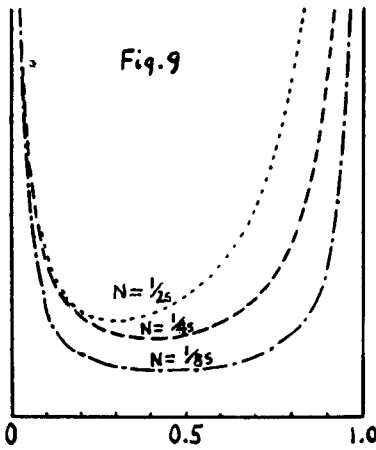
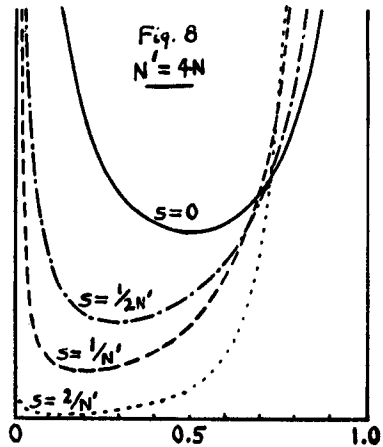
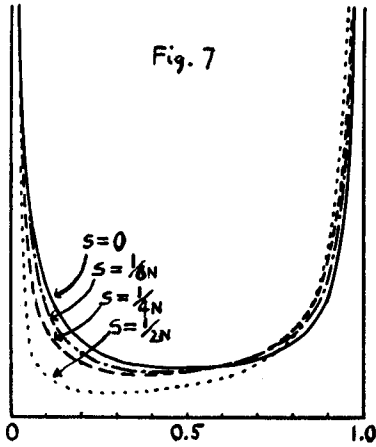
The selection coefficient refers here to the difference between the selection in the group under consideration and that in the species as a whole, the effect of the latter being taken account of in the mean gene frequency of the species  $q_m$ .

#### *The distribution curves*

Some of the forms taken by the probability array of gene frequencies, in cases involving selection, are illustrated in figures 7 to 14. Figures 7 to 10 deal with the case in which the rates of mutation are negligibly small compared with  $1/4N$ . The curves are thus all variants of the form  $y = Ce^{4Ns}q^{-1}(1-q)^{-1}$ . Figure 7 illustrates the relatively slight effect of selection below a certain relation to size of population. All conditions are the same in figure 8 except that the populations are four times as large as in figure 7. Thus while the absolute intensities of the selection coefficients are the same, the relations to size of population are altered.<sup>11</sup> The curves bring out the great effect of selection beyond the critical point,  $s = \frac{1}{2N}$  (where mutation rates are low). Figures 9 and 10 are intended to show the effects of change in size of population where the intensity of selection remains constant (low in figure 9, four times as severe in figure 10). Up to a certain point ( $N = \frac{1}{2s}$ , figure 9) increase in population raises the middle portion of the curve. Above this point (figure 10) increase in population depresses the middle portion. In the former case, the increase in unfixed factors brings increased variability of cumulative characters, in the latter there is little change of variability in relation to population size, the depression among middle frequencies being balanced by the accumulation of nearly but not quite fixed factors. All of these fig-

<sup>11</sup> The probability that increase in the number of unfixed genes would react on the individual gene selection coefficients, reducing them, is here ignored.

ures (7 to 10) may be taken as representative of conditions in small inbred populations which have been isolated sufficiently long to reach equilibrium in relation to mutation. It will be recalled that figures 3 and 4 represent successive stages preceding the attainment of such equilibrium.



FIGURES 7, 8, 9, and 10.—Distribution of gene frequencies in relation to size of population and intensity of selection where rates of mutation and migration are small compared with  $1/4N$ . Formulae all of type  $y = Cy^{4Nsq}q^{-1}(1-q)^{-1}$ .

Figure 7. Small population, four degrees of selection. Figure 8. Population four times as large as in figure 7 under the same four (absolute) degrees of selection. Figure 9. Three sizes of population under given weak selection. Figure 10. Same three sizes of population as in figure 9, under selection four times as severe.

Figures 11 to 14 present exactly the same series of comparisons as figures 7 to 10, for small populations that are not completely isolated from the main body of a species.<sup>12</sup> In all cases the gene frequency ( $q_m$ ) of the

Fig. 11

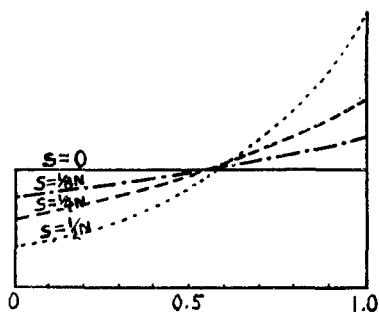


Fig. 12

$N' = 4N$

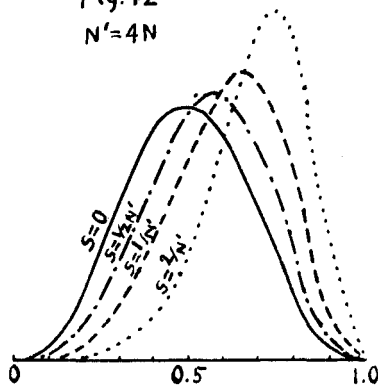


Fig. 13

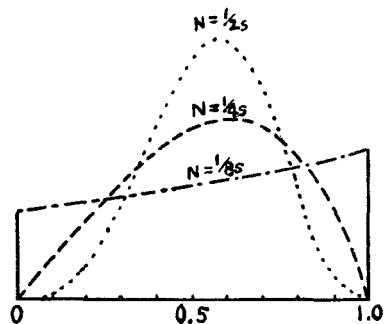
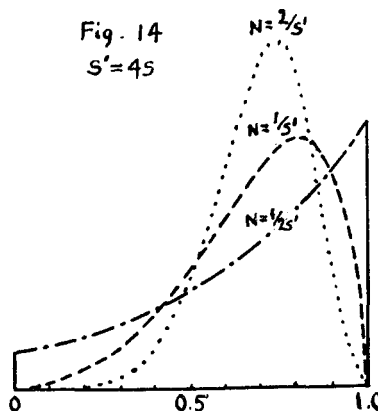


Fig. 14

$S' = 4S$



FIGURES 11 to 14.—Distribution of gene frequencies in subgroups of large population (mean frequency  $q_m = 1/2$ ) in relation to size of population and intensity of selection. Formulae all of the type  $y = Ce^{4Ns}q^{4Nm}q_m^{-1}(1-q)^{4Nm(1-q_m)^{-1}}$ . Same comparisons as in figures 7 to 10.

Figure 11. Small subgroups ( $2Nm = 1$ ), four degrees of selection. Figure 12. Subgroups four times as large as in figure 11, under same four (absolute) degrees of selection. Figure 13. Subgroups of three sizes under given weak selection. Figure 14. Same three sizes as in figure 13 under selection four times as severe.

The figures may also be used to illustrate cases of equal mutation to and from a gene ( $u = v$ ).  $y = Ce^{4Ns}q^{4Nv-1}(1-q)^{4Nu-1}$

<sup>12</sup> These figures also represent the distribution of gene frequencies in population in which mutation and reverse mutation are equally frequent, but this seems to be so exceptional a case especially under multiple allelomorphism, as to be of little importance.



whole species is assumed to be  $1/2$ . The relations of migration to size of population are such that there is very little complete fixation of genes. In figure 11,  $m = 1/2N$  and the purely exponential curves show how increasing intensity of genic selection shifts a uniform distribution in the direction favored by the selection. The fourfold greater population of figure 12 brings about concentration, in curves approaching the normal in form. Figure 13 brings out the concentrating effect of increase in population in the case of weak selection while figure 14 does the same for the case of selection four times as severe.

The important case in which mutation is balanced by selection in a moderately large population (both  $s$  and  $u$  large compared with  $1/4N$ ) is illustrated in figure 19. The four curves represent four degrees of selection, rising by doubling of severity at each step from a case in which mutation pressure practically overwhelms the effect of selection to the reverse situation. The limiting condition in populations so large that  $1/4N$  is very small compared with both  $s$  and  $u$  is that of concentration of factor frequency almost at a single value (figure 20, page 148).

#### *Dominance ratio*

The form of the distribution of the frequencies of the dominant genes affecting a character is of interest in connection with the dominance ratio. Since different genes have different mutation rates and selection coefficients, this distribution is a composite of curves of the types discussed. In small populations which have reached equilibrium, all of these arrays and hence their composite are of the type  $Cq^{-1}(1-q)^{-1}$ . The dominance ratio is  $1/5$  in so far as dominance is complete. FISHER gives the value as  $3/13 = 0.23$  for the case "when in the absence of selection, sufficient mutation takes place to counteract the effect of random survival." The difference from the value 0.20 given above is due solely to the difference in the formula for the curve, discussed earlier.

In the case of the isolation of a small part of a large population, the dominance ratio takes the value  $1/3$  in so far as dependent on dominant type genes in equilibrium with recessive mutation but not with reverse mutation ( $y = C/(1-q)$ ). Where following isolation both fixation and loss are substantially irreversible ( $y = 1$ ) the dominance ratio is  $1/4$  in agreement with FISHER's result. In both of these cases, of course, the dominance ratio falls to  $1/5$  when equilibrium is finally attained.

The foregoing discussion applies practically only to very small completely isolated populations. In large populations where the distribution of gene frequencies, even in partially isolated subgroups, tends to approach

the normal type, the dominance ratio comes to depend mainly on the mean gene frequency which depends on the relation of selection to mutation, or on selection against both homozygotes in favor of heterozygotes. In the extreme case in which the gene frequency is reduced to a single value, the dominance ratio is  $q/(2-q)$ . Values close to unity should not be uncommon, especially where gene frequency is controlled by the balance of selection and mutation. Such a dominance ratio has rather surprising effects on the correlation between relatives. The correlation between parent and offspring approaches 0 although that between brothers may remain as high as 0.25. However, the occurrence of an appreciable number of genes at lower frequencies, for example, held in equilibrium by selection favoring the heterozygote against both homozygotes would greatly lower the dominance ratio.

All of these figures are on the assumption that dominance is complete. Dominance, however, is frequently not complete. Among 22 heterozygotes in the guinea pig which have been studied with some care, at least 9 or about 40 percent are to some extent intermediate. Most of these have to do with color characters. It is not unlikely that incomplete dominance will be found to be even more frequent on careful study of size characters.

FISHER (1922) comes rather definitely to the conclusion that the dominance ratio is typically in the neighborhood of 1/3. This was based primarily on a distribution of factor frequencies which he reached for the case of selection against a recessive,<sup>13</sup> with which the results of the present study are not at all in agreement. He also finds, however, that the differences between fraternal and parent-offspring correlations in data which he analyzes indicate the same figure. The analysis of a large number of correlations of these sorts would undoubtedly furnish valuable information with regard to the statistical situation in populations. It is to be noted, however, that similarity in the environment of brothers as compared with parent and offspring may also contribute to a higher fraternal correlation and that in any case one cannot reason from the dominance ratio deduced from correlations to the distribution of factor frequencies without making some assumption as to the prevalence of dominance.

About all that seems justified by the present analysis, is the statement that for permanently small populations under low selection the value should

<sup>13</sup> This was given as  $df \propto \frac{d\theta}{\sin \frac{1}{2}\theta \cos^3 \frac{1}{2}\theta}$  or  $\phi(q) = Cq^{-1}(1-q)^{-2}$  on transformation of scale.

FISHER does not discuss dominance ratio in connection with his recent revision of his results in *The genetical theory of natural selection*, 1930.

be less and probably considerably less than 0.20 but that this figure may be raised by severe selection (favoring dominants) and especially by increase in size of population. It may even approach unity in very large populations under severe selection, if complete dominance is the rule.

*Mean and variability of characters*

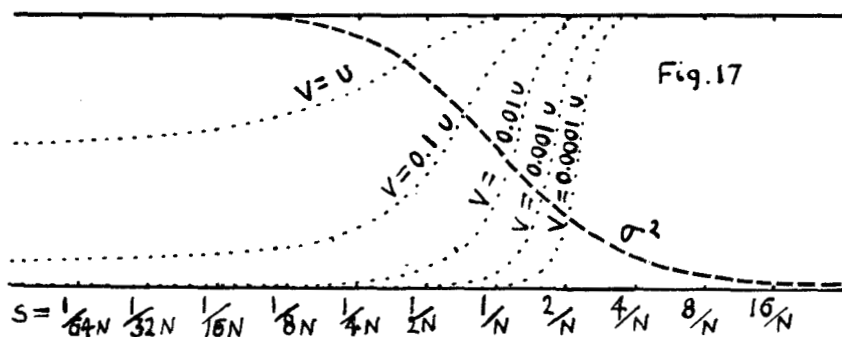
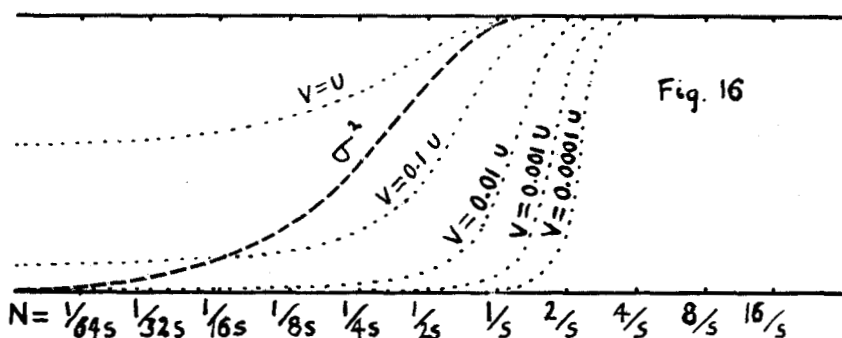
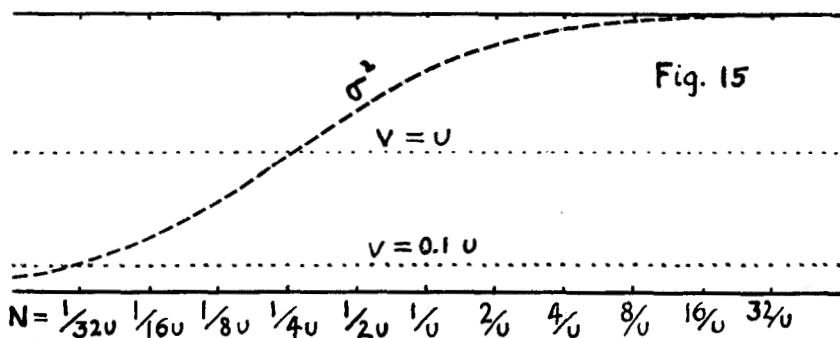
In the case of genes which are indifferent to selection ( $s$  less than  $u$ ), the mean frequency  $\bar{q} = v/(u+v)$  remains unchanged through all transformations from a U-shaped distribution in small populations to an I-shaped one in large populations. The variance, due to such genes, is small in small populations, rises in nearly direct proportion to size of population up to a certain critical point (about  $N = 1/4u$ ) and then approaches a limiting value. For the case in which mutations in one direction ( $u$ ) occur at a much greater rate than in the other ( $v$ ), the general formula reduces to  $\sigma^2 = \sigma_\infty^2 \left( \frac{4Nu}{1+4Nu} \right)$ , in which  $\sigma_\infty^2 (= 2La^2v/u)$  is the limiting value.

This case is illustrated in figure 15. The dotted lines represent mean gene frequencies and the line of dashes the variance.

Actual changes in the size of a given population are not of course accompanied by instant adjustment of the distribution of gene frequencies. A decrease in size to a point well below the critical value is followed by decrease in heterozygosis and variance at a rate between  $1/4N$  and  $1/2N$  per generation depending on the number of allelomorphs. This may be a fairly rapid process in terms of geologic time but the recovery of heterozygosis through growth of the population to its original size occurs more slowly, since this depends on mutation pressure. On the other hand, the intercrossing of a number of isolated strains, in each of which the reduction of variance has occurred, is followed by immediate recovery of the original statistical situation (except with respect to factor combinations in which there is some delay).

In the opposite case of genes under vigorous selection ( $s$  much greater than  $u$ ) mean frequency as well as variance is affected by size of population and by severity of selection. As in the preceding case, variance is small in small populations, rises in nearly direct proportion to growth of population<sup>14</sup> until a critical point is approached (here about  $N = 1/4s$ ) and then rapidly approaches a limiting value  $\sigma^2 = \sigma_\infty^2 \left( \frac{e^{4Ns} - 1}{e^{4Ns} + 1} \right)$  where

<sup>14</sup> As before, the probability that the increase in variance, due to growth of population, would react on the selection coefficients of the individual genes, reducing them, requires some qualification of this statement in application to actual populations.



FIGURES 15 to 17.—The variance ( $\sigma^2$ ) and mean gene frequencies (dotted lines) at equilibrium under various conditions of mutation, selection and size of population. Figure 15. Effects of increasing population where selection is negligible relative to mutation. Figure 16. Effects of increasing population where mutation rates are small compared with  $1/4N$ . Figure 17. Effects of increasing selection where mutation rates are small compared with  $1/4N$ .

$\sigma_{\infty}^2 = \mu a^2/s$ . The mean factor frequency in large populations ( $\bar{q} = 1 - u/s$ ) is close to 1, nearly complete fixation of the favorable gene. In small populations, on the other hand, the equilibrium point approaches that of the opposing mutation pressures ( $\bar{q} = v/(u+v)$ ) and hence practically 0, with complete loss ( $v=0$ ) as the inevitable ultimate fate in an extended multiple allelomorph series. Up to the point at which mutation pressure seriously disturbs the form of the distribution curve, the mean gene frequencies are simply the ratios of the chances of fixation at each extreme, namely,  $ve^{4Ns}/(u + ve^{4Ns})$ .

The relations of mean frequency and variance to size of population in this case are both shown in figure 16, the former for various relative values of  $u$  and  $v$ . Inspection of figures 9 and 10 may also be of assistance in understanding this situation.

As in the other case, actual change in size of population is not accompanied by immediate attainment of the new equilibrium. Decrease in population to a number well below the critical point is followed by decrease in heterozygosis at the rate described, bringing with it at the same rate the well known inbreeding effects, loss of variance and, in general, decline in vigor toward a new level. This immediate decline in vigor is not due to change in mean gene frequency, but merely to the greater proportion of recessive phenotypes as homozygosis increases, and thus comes to an end when the degree of homozygosis has reached equilibrium. The change in mean gene frequency proceeds more slowly since it depends on mutation pressure. Long continued isolation should thus involve two distinct degeneration processes, a rapid but soon completed process of fixation and a very slow process of accumulation of injurious genes. The recovery on increase in size of population is slow in both cases, depending on mutation pressure. The intercrossing of isolated lines, on the other hand, is followed by immediate return to the original status of the population if only the immediate inbreeding effect has occurred, but must wait on favorable mutations if there has been time for the slower process.

The effects of different intensities of selection on mean gene frequency and variance (population size constant) are illustrated in figure 17, still assuming that the selection coefficient is of higher order than mutation rate. Figures 7 and 8 showing the distribution of gene frequencies in this case may also be of assistance here. Selection has little effect on variability until it reaches about the value  $1/8N$ , about half the variance is eliminated when selection reaches  $1/N$  and most of it at  $s = 4/N$ . The formula is  $\sigma^2 = \sigma_0^2 \left( \frac{e^{4Ns} - 1}{2Ns(e^{4Ns} + 1)} \right)$ . Selection, of course, affects the mean gene

frequency, the formula being the same as that given above under the effect of size of population. On actual increase in the intensity of selection, the rate of change toward the new equilibrium both in mean and variance is controlled by selection pressure and may thus be fairly rapid in terms of geologic time in a large population. This is the case in which HALDANE's formulae for progress under selection are most applicable. The increase in variance and in the proportion of unfavorable genes following relaxation of selection, on the other hand, are controlled by mutation pressure and thus approach equilibrium relatively slowly. A shift in gene frequency at rate  $uq$  may well mean no more than 0.000,001 per generation.

The type of result where the selection coefficient is of the same order of magnitude as mutation rate can be inferred, qualitatively, at least from the preceding extreme cases. Inspection of figure 19 may also be of assistance here.

#### THE EVOLUTION OF MENDELIAN SYSTEMS

##### *Classification of the factors of evolution*

In attempting to draw conclusions with respect to evolution one is apt, perhaps, to assume that factors which make for great variation are necessarily favorable while those which reduce variation are unfavorable. Evolution, however, is not merely change, it is a process of *cumulative* change: fixation in some respects is as important as variation in others. Live stock breeders like to compare their work to that of a modeller in clay. They speak of moulding the type toward the ideal which they have in mind. The analogy is a good one in suggesting that in both cases it is a certain intermediate degree of plasticity that is required.

The basic cumulative factor in evolution is the extraordinary persistence of gene specificity. This doubtless rests on a tendency to precise duplication of gene structure in the proper environment. The basic change factor is gene mutation, the occasional failure of precise duplication. Since the time of LAMARCK, a school of biologists have held that the primary changes in hereditary constitution must be adaptive in direction in order to account for evolutionary advance. Unfortunately, the results of experimental study have given no support to this view. Instead, the characteristics of actually observed gene mutations seem about as unfavorable as could be imagined for adaptive evolution. In the first place, is their fortuitous occurrence. No correlation has been found between external conditions and direction of mutation, and those few agents which have been found to affect the rate (X-ray, radium, and to a relatively unimportant extent, temperature) merely speed up the rate of random mutation. The great

majority of mutations are either definitely injurious to the organism or produce such small effects as to be seemingly negligible. MULLER has graphically compared the range of mutations to a spectrum in which the nonlethal conspicuous mutations form a narrow field between broad regions of individually inconspicuous mutations on the one hand and of sublethal and lethal mutations on the other. In addition, the great majority of mutations are more or less completely recessive to the type genes from which they arise. These effects are easily understood if mutation is an accidental process. Random changes in a complex organization are more likely to injure than to improve it, and with respect to the immediate products of the gene, random changes are more likely to be of the nature inactivation (and hence probably recessive) than of increased activation. Finally is to be mentioned the extreme rarity of gene mutation. Even in *Drosophila*, mutation rates as high as  $u = 10^{-5}$  per locus seem to be exceptional and  $10^{-6}$  or less more characteristic. This infrequency seems unfavorable to rapid evolution, yet it is a necessary corollary of the usually injurious effect, if life is to persist at all. Moreover, the more advanced the evolution, the slower must be the time rate of mutation. In one-celled organisms, dividing several times a day, a rapid time rate of mutation will not prevent the production of sufficient normal offspring to maintain the species. The same time rate in *Drosophila* with an interval of some two weeks between generations would mean such an accumulation of lethals in every gamete that the species would come to an abrupt end. The time rate of lethal mutation in *Drosophila* (7 per 1000 chromosomes per month under ordinary conditions according to MULLER (1928)) would be quite impossible in the human species. The problem is to determine how an adaptive evolutionary process may be derived from such unfavorable raw material as the infrequent, fortuitous and usually injurious gene mutations.

It will be convenient here to classify factors of evolution according as they tend toward genetic homogeneity or heterogeneity of the species. They are grouped below in more or less definitely opposing pairs.

<i>Factors of Genetic Homogeneity</i>	<i>Factors of Genetic Heterogeneity</i>
Gene duplication	Gene mutation ( $u, v$ )
Gene aggregation	Random division of aggregate
Mitosis	Chromosome aberration
Conjugation	Reduction (meiosis)
Linkage	Crossing over
Restriction of population size ( $1/2N$ )	Hybridization ( $m$ )
Environmental pressure ( $s$ )	Individual adaptability
Crossbreeding among subgroups ( $m_1$ )	Subdivision of group ( $1/2N_1$ )
Individual adaptability	Local environments of subgroups ( $s_1$ )

The first pair have been discussed above. They enter into the formulae through the mutation rates  $u$  and  $v$ . MULLER has pointed out the necessary similarity, in order of size, of genes and of filterable viruses and has suggested the possibility that the latter may consist of single genes. If so, their evolution rests wholly on a not too high rate of mutation, and selection, which seems possible enough in organisms as simple as these presumably are, especially as in this case where the gene is the organism, the mutation of the gene need not be expected to be as fortuitously related to the activities of the organism as in more complex cases.

Presumably the first step toward higher organisms is the aggregation of such genes with multiplication of the aggregate by random division. Given occasional gene mutation, this leads to a new kind of variation, that in proportional abundance of the different kinds of genic material. The larger the aggregate, the less violent the variation. Large aggregates present a labile system capable of quantitative variation in response (perhaps physiologically as well as through selection) to changing conditions. As far as observation goes, the bacteria, and blue green algae have no mechanism of division beyond a random division of the protoplasmic constituents. Such apportionment of more or less autonomous materials may also be important in the differentiating cell lineages of multicellular organisms, but, except for a few plastid characters, seems to play no important role in heredity from generation to generation, as far as has been determined by experiment. There seems here an adequate basis for an evolutionary process in organisms so simple that the handing on of a few different protoplasmic constituents can determine all of the characteristics of the species but the conditions are not favorable for an extensive cumulative process.

Mitosis provides the mechanism by which an indefinitely large number of qualitatively different elements may be maintained in the same proportions. But it provides so perfectly for the persistence of complex organization that further change is difficult. Irregularities in mitosis provide a source of variation but of so violent a nature for the most part as to be of infrequent evolutionary importance, although the differences in chromosome numbers of related species demonstrate that they play a genuine rôle. Complete duplication (tetraploidy) is important in doubling the possible number of different kinds of genes. Other aberrations, especially translocations, are probably more important in isolating types, than for the character changes which they bring. Gene mutation remains the principal factor of variation, but seems inadequate as the basis of an evolutionary process under exclusively mitotic (asexual) reproduction.

The most important factor in transcending the evolutionary difficulties



inherent in the characteristics of gene mutation is undoubtedly the attainment of biparental reproduction (EAST 1918). This involves two phases, conjugation, a factor which makes the entire interbreeding group a physiological unit in evolution, and meiosis, with its consequence, Mendelian recombination which enormously increases the amount of variability within the limits of the species. Each additional viable mutation in an asexually reproducing form merely adds one to the number of types subject to natural selection. The chance that two or more indifferent or injurious mutations may combine in one line to produce a possibly favorable change is of the second or higher order. Under biparental reproduction, each new mutation doubles the number of potential variations which may be tried out. The contrast is between  $n+1$  and  $2^n$  types from  $n$  viable mutations.

Biparental reproduction solves the evolutionary requirement of a rich field of variation. But by itself it provides rather too much plasticity. It makes a highly adaptable species, capable of producing types fitted to each of a variety of conditions, but a successful combination of characteristics is attained in individuals only to be broken up in the next generation by the mechanism of meiosis itself.

An excellent illustration of the principle that a balance between factors of homogeneity and of heterogeneity may provide a more favorable condition for evolution than either factor by itself may be found in the effects of an alternation of a series of asexual generations with an occasional sexual generation. Evolution is restrained under exclusive asexual reproduction by the absence of sufficient variation, and under exclusive sexual reproduction by the noncumulative character of the variation, but, on alternating with each other, any variety in the wide range of combinations provided by a cross may be multiplied indefinitely by asexual reproduction. The selection of individuals is replaced by the much more effective selection of clones and leads to rapid statistical advance which, however, comes to an end with reduction to a single successful clone. On the other hand a new cross (before reduction to a single clone) may provide a new field of variation making possible a repetition of the process at a higher level. This method has been a favorite of the plant breeder and is perhaps the most successful yet devised for human control of evolution in those cases to which it can be applied at all. Under natural conditions, alternation of asexual and sexual reproduction is characteristic of many organisms and doubtless has played an important rôle in their evolution.

The demonstration of the evolutionary advantages of an alternation of the two modes of reproduction seems to prove too much. Asexual re-

production is practically absent in the most complex group of animals, the vertebrates, and is rather sporadic in its occurrence elsewhere. The purpose of the present paper has been to investigate the statistical situation in a population under exclusive sexual reproduction in order to obtain a clear idea of the conditions for a degree of plasticity in a species which may make the evolutionary process an intelligible one.

First may be mentioned briefly a modification of the meiotic mechanism which has been introduced only qualitatively into the investigation where at all. This is the aggregation of genes into more or less persistent systems, the chromosomes. Complete linkage cuts down variability by preventing recombination. Wholly random assortment gives maximum recombination but does not allow any important degree of persistence of combinations once reached. An intermediate condition permits every combination to be formed sooner or later and gives sufficient persistence of such combinations to give a little more scope to selection than in the case of random assortment. Close linkage, moreover, brings about a condition in which selection tends to favor the heterozygote against both homozygotes and so helps in maintaining a store of unfixed factors in the population.

Restriction of size of population, measured by  $1/2N$ , is a factor of homogeneity and conversely with increase of size. The effects of restricted size may also be balanced by those of occasional external hybridization, measured by  $m$ .

Environmental pressure on the species as a whole is a factor of homogeneity. It has been urged by some that because natural selection is a factor which reduces variability, and most conspicuously by eliminating extreme types, it cannot be the guiding principle in adaptive evolution. From the viewpoint of evolution as a moving equilibrium, however, the guiding principle may be found on the conservative as well as on the radical side. The selection coefficient,  $s$ , depends on the balance between environmental pressure and individual adaptability. High development of the latter permits the survival of genetically diverse types in the face of severe pressure.

Subdivision of a population into almost completely isolated groups, whether by prevailing self fertilization, close inbreeding, assortative mating, by habitat or by geographic barriers is a factor of heterogeneity with effects measured by  $1/2N_1$ ,  $N_1$  being here the size of the subgroup. This factor may be balanced by crossbreeding between such groups, measured by  $m_1$ .

It is interesting to note that restriction of population size is a factor of homogeneity or of heterogeneity for the species, depending on whether it relates to the species as a whole or to subgroups and conversely with the crossbreeding coefficient. Similarly the selection pressures of varied environments within the range of the species ( $s_1$ ) constitute factors of heterogeneity, restrained from excessive genetic effect by the same individual adaptability which appears in the opposite column in relation to the general environment of the species. Individual adaptability is, in fact, distinctly a factor of evolutionary poise. It is not only of the greatest significance as a factor of evolution in damping the effects of selection and keeping these down to an order not too great in comparison with  $1/4N$  and  $u$ , but is itself perhaps the chief object of selection. The evolution of complex organisms rests on the attainment of gene combinations which determine a varied repertoire of adaptive cell responses in relation to external conditions. The older writers on evolution were often staggered by the seeming necessity of accounting for the evolution of fine details of an adaptive nature, for example, the fine structure of all of the bones. From the view that structure is never inherited as such, but merely types of adaptive cell behavior which lead to particular structures under particular conditions, the difficulty to a considerable extent disappears. The present difficulty is rather in tracing the inheritance of highly localized structural details to a more immediate inheritance of certain types of cell behavior.

*Labiality as the condition for evolution*

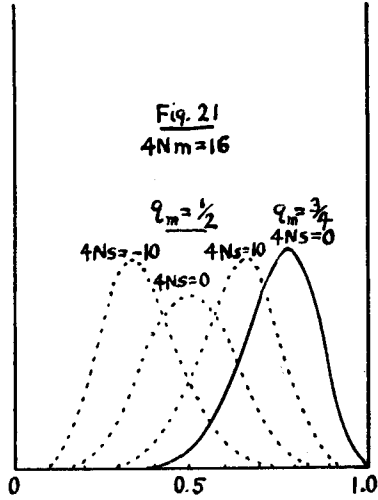
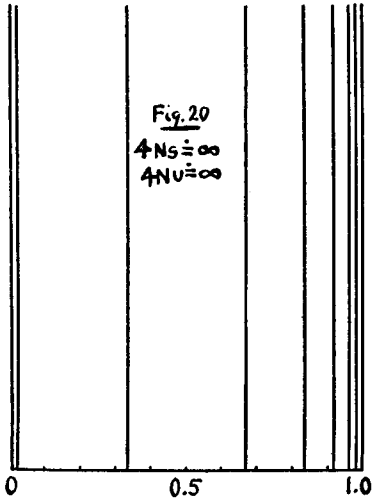
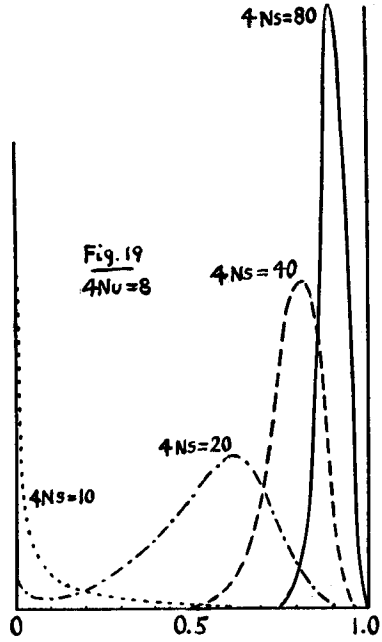
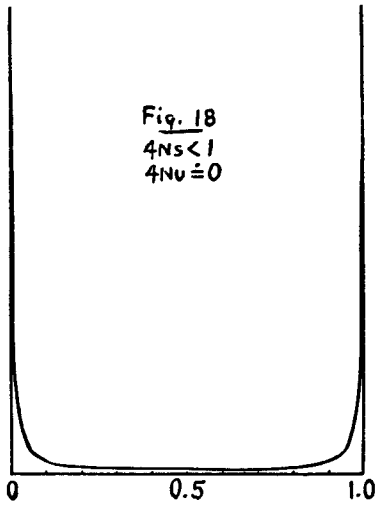
The statistical effects of the more important of these factors in a freely interbreeding population are brought together in the formula

$$y = Ce^{4Nsq}q^{-1}(1 - q)^{4Nu-1}.$$

The term  $4Nv$  in the exponent of  $q$  is here assumed to be negligible and the terms applicable in case of external hybridization are also omitted.

Consider first the situation in a small population in which  $1/4N$  is much greater than  $u$  and than  $s$  (figure 18). Nearly all genes are fixed in one phase or another. Even rather severe selection is without effect. There is no equilibrium for individual genes. They drift from one state of fixation to another in time regardless of selection, but the rate of transfer is extremely slow. Such evolution as there is, is random in direction and tends toward extinction of the group.

Consider next the opposite extreme, a very large undivided population under severe selection. Assume that  $s$  is in general much greater than  $u$  and that the latter is much greater than  $1/4N$ . There is almost complete



FIGURES 18 to 21.—Distributions of gene frequencies in relation to size of population, selection, mutation and state of subdivision. Figure 18. Small population, random fixation or loss of genes ( $y = Cq^{-1}(1-q)^{-1}$ ). Figure 19. Intermediate size of population, random variation of gene frequencies about modal values due to opposing mutation and selection ( $y = Ce^{4Nsqq^{-1}}(1-q)^{4Nu-1}$ ). Figure 20. Large population, gene frequencies in equilibrium between mutation and selection ( $q = 1-u/s$ , etc.). Figure 21. Subdivisions of large population, random variation of gene frequencies about modal values due to immigration and selection. ( $y = Ce^{4Nsqq^{Nm}q_m^{-1}}(1-q)^{4Nm(1-q_m)-1}$ ).

fixation of the favored gene for each locus. Here also there is little possibility of evolution. There would be complete equilibrium under uniform conditions if the number of allelomorphs at each locus were limited. With an unlimited chain of possible gene transformations, new favorable mutations should arise from time to time and gradually displace the hitherto more favored genes but with the most extreme slowness even in terms of geologic time.<sup>15</sup>

Even if selection is relaxed to such a point that the selection coefficients of many of the genes are not much greater than mutation rates, the conditions are not favorable for a rapid evolution (figure 20). The amount of variability in the population may be great, maximum in fact, but if the distributions of gene frequencies are closely concentrated about single values, the situation approaches one of complete equilibrium and hence of complete cessation of evolution. At best an extremely slow, adaptive, and hence probably orthogenetic advance is to be expected from new mutations and from the effects of shifting conditions.

It should be added that a relatively rapid shift of gene frequencies can be brought about in this case by vigorous increase in the intensity of selection. The effects of unopposed selection of various sorts and in various relations of the genes has been studied exhaustively by HALDANE, with regard to the time required to bring about a shift of gene frequency of any required amount. The end result, however, is the situation previously discussed. The rapid advance has been at the expense of the store of variability of the species and ultimately puts the latter in a condition in which any further change must be exceedingly slow. Moreover, the advance is of an essentially reversible type. There has been a parallel movement of all of the equilibria affected and on cessation of the drastic selection, mutation pressure should (with extreme slowness) carry all equilibria back to their original positions. Practically, complete reversibility is not to be expected, and especially under changes in selection which are more complicated than can be described as alternately severe and relaxed. Nevertheless, the situation is distinctly unfavorable for a continuing evolutionary process.

Thus conditions are unfavorable for evolution both in very small and in very large, freely interbreeding, populations, and largely irrespective of severity of selection. We have next to consider the intermediate situa-

<sup>15</sup> This, nevertheless, seems to be the case which FISHER (1930) considers most favorable to evolution. The greatest difference between our conclusions seems to lie here. His theory is one of complete and direct control by natural selection while I attribute greatest immediate importance to the effects of incomplete isolation.

tion in which  $s$  is not much greater than  $u$  for many genes and the latter is not much greater than  $1/4N$ . Such a case is illustrated in figure 19. The size of population is sufficient to prevent random fixation of genes, but insufficient to prevent random drifting of gene frequencies about their mean values, as determined by selection and mutation. It is to be supposed that the relations of the selection and mutation coefficients vary from factor to factor. The more indifferent ones drift about through a wide range of frequencies in the course of geologic time while those under more severe selection oscillate about positions close to complete fixation. In any case, all gene frequencies are continually changing even under uniform environmental conditions. But the selection coefficients themselves are in general to be considered functions of the entire array of gene frequencies and will therefore also be continually changing. The probability arrays of some genes will travel to the right and close up as their selection coefficients stiffen, while some of the genes which have been nearly fixed will come to be less severely selected and their probability arrays will shift to the left and open out or even move to the extreme left under displacement by another allelomorph. A continuous and essentially irreversible evolutionary process thus seems inevitable even under completely uniform conditions. The direction is largely random over short periods but adaptive in the long run. The less the variation of gene frequency about its mean value, the closer the approach to an adaptive orthogenesis. Complete separation of the species into large subspecies should be followed by rather slow more or less closely parallel evolutions, if the conditions are similar, or by adaptive radiation, under diverse conditions, while isolation of smaller groups would be followed by a relatively rapid but more largely nonadaptive radiation.

As to rate, since the process depends mainly on the value of  $1/4N$ , assumed to be somewhat less than  $u$  (and  $s$ ) the process cannot be as rapid as one due temporarily to either unopposed selection or unopposed mutation pressure. Hundreds of thousands of generations seem to be required at best for important nonadaptive evolutionary changes of the species as a whole; while adaptive advance, depending on the chance attainment of favorable combinations would be much slower. Even so the process is much the most rapid non-self-terminating one yet considered.

In reaching the tentative conclusion that the situation is most favorable for evolution in a population of a certain intermediate size, one important consideration has been omitted. This is the tendency toward subdivision into more or less completely isolated subgroups in widely distributed populations. Within each subgroup there is a distribution of gene fre-

quencies dependent largely on its own size ( $N_1$ ), amount of crossing with the rest of the species ( $m_1$ ), selection due to local conditions ( $s_1$ ) and the mean gene frequency of migrants from the rest of the species ( $q_m$ ). It may be assumed that  $1/4N_1$  is so much larger than  $u$  that mutation pressure (and also the average selection coefficient,  $s$ , for the whole species) can be ignored.

$$y = Ce^{4N_1s_1q} q^{4N_1m_1q_m-1} (1 - q)^{4N_1m_1(1-q_m)-1}.$$

Gene frequency in each subgroup oscillates about a mean value, which is that of the whole species only if conditions of selection are uniform. Figure 21 represents various cases. The random variations of gene frequency have effects similar to those described above within each group. The result is a partly nonadaptive, partly adaptive radiation among the subgroups. Those in which the most successful types are reached presumably flourish and tend to overflow their boundaries while others decline, leading to changes in the mean gene frequency of the species as a whole. In this case, the rate of evolution should be much greater than in the previous cases. The coefficients  $1/4N_1$  and  $s_1$  may be relatively large and bring about rapid differentiation of subgroups, while the competition between subgroups will bring about rapid changes in the gene frequencies of the species as a whole. The direction of evolution of the species as a whole will be closely responsive to the prevailing conditions, orthogenetic as long as these are constant, but changing with sufficiently long continued environmental change.

A question which requires consideration is the effect of alternation of conditions, large and small size of population, severe and low selection. The effects of changes in the conditions of selection have already been touched upon. Persistence of small numbers or of severe selection for such periods of time as to bring about extensive fixation of factors compromises evolution for a long time following, there being no escape from fixation except by mutation pressure. Many thousands of generations may be required after restoration to large size and not too severe selection, before evolutionary plasticity is restored. Short time oscillations in population number or severity of selection, on the other hand, probably tend to speed up evolutionary change by causing minor changes in gene frequency.

#### *Control of evolution*

With regard to control of the process, it is evident that little is possible either within a small stock or a freely interbreeding large one. Even drastic

selection is of little effect in the former, and in the latter, while it may bring about a rapid immediate change in the particular respect selected, this must be at the expense of other characters, and in any case, soon leads to a condition in which further advance must wait on the occurrence of mutations more favorable than those fixed by the selection. The limitations in this case have been well brought out in a recent discussion by KEMP (1929). Maximum continuous progress in a homogeneous population requires an intensity of selection for each of the more indifferent genes not much greater than its mutation rate and also a certain size of population. Even so, the direction of advance is somewhat uncertain and the rate to be measured in geologic time.

If infrequency of mutation is the limiting factor here, it would seem that a considerable increase in the rate of evolution should be made possible by a speeding up of mutation, as by X-rays. There is a limit, however, imposed by the prevailing injurious character of mutations. Even the most rigorous culling of individuals means in general, only a low selection coefficient (in absolute terms) for each of the presumably numerous unfixed genes, which are not in themselves lethal or sublethal in effect. Such culling would become insufficient to hold mutation pressure in check when the latter had increased beyond a certain point ( $u > s$ ). Moreover, as the number of unfixed genes becomes greater under an increased mutation rate, the smaller becomes the separate gene selection coefficients, making it certain that mutation rate could not increase very much before the possibility of effective selection (in all respects at once) rather than infrequency of mutation would become the limiting factor. With respect to lethal mutations, it has already been noted that the observed natural time rate in *Drosophila* is such as would mean immediate extinction, if transferred to the human species. It is clear that an evolution in the direction of increased gene stability, rather than mutability, has been a necessary phase, in the evolution of the longer lived higher animals. This makes it unlikely that a general increase in mutation rate would increase the rate of evolutionary advance along adaptive lines.

The only practicable method of bringing about a rapid and non-self-terminating advance seems to be through subdivision of the population into isolated and hence differentiating small groups, among which selection may be practiced, but not to the extent of reduction to only one or two types (WRIGHT 1922a). The crossing of the superior types followed by another period of isolation, then by further crossing and so on *ad infinitum* presents a system by means of which an evolutionary advance through the field of possible combinations of the genes present in the original stock, and



arising by occasional mutation, should be relatively rapid and practically unlimited. The occasional use of means for increasing mutation rate within limited portions of the population should add further to the possibilities of this system.

*Agreement with data of evolution*

We come finally to the question as to how far the characteristics of evolution in nature can be accounted for on a Mendelian basis. A review of the data of evolution would go far beyond the scope of the present paper. It may be suggested, however, that the type of moving equilibrium to be expected, according to the present analysis, in a population comparable to natural species in numbers, state of subdivision, conditions of selection, individual adaptability, etc. agrees well with the apparent course of evolution in the majority of cases, even though heredity depend wholly on genes with properties like those observed in the laboratory. Adaptive orthogenetic advances for moderate periods of geologic time, a winding course in the long run, nonadaptive branching following isolation as the usual mode of origin of subspecies, species and perhaps even genera, adaptive branching giving rise occasionally to species which may originate new families, orders, etc.; apparent continuity as the rule, discontinuity the rare exception, are all in harmony with this interpretation.

The most serious difficulties are perhaps in apparent cases of nonadaptive orthogenesis on the one hand and extreme perfection of complicated adaptations on the other. In so far as extreme degeneration of organs is concerned, there is little difficulty—this is to be expected as a by-product of other evolutionary changes. Because of their multiple effects, there can be no really indifferent genes, whatever may be true of organs which have been reduced beyond a certain size. Zero as the value of a selection coefficient is merely a mathematical point between positive and negative values. It is common observation that mutations are more likely to reduce the development of an organ than to stimulate it. It follows that evolutionary change in general will have as a by product the gradual elimination of indifferent organs. Nonadaptive orthogenesis of a positive sort, increase of size of organs to a point which threatens the species, constitutes a more difficult problem, if a real phenomenon. Probably many of the cases cited are cases in which the line of evolution represents the most favorable immediately open to a species doomed by competition with a form of radically different type or else cases in which selection based on individual advantage leads the species into a cul-de-sac. The nonadaptive differentiation of small subgroups and the great effectiveness of subsequent

selection between such groups as compared with that between individuals seem important factors in the origin of peculiar adaptations and the attainment of extreme perfection. It is recognized that there are specific cases which seem to offer great difficulty. This should not obscure the fact that the bulk of the data indicate a process of just the sort which must be occurring in any case to some extent as a statistical consequence of the known mechanism of heredity. The conclusion seems warranted that the enormous recent additions to knowledge of heredity have merely strengthened the general conception of the evolutionary process reached by DARWIN in his exhaustive analysis of the data available 70 years ago.

*“Creative” and “emergent” evolution*

The present discussion has dealt with the problem of evolution as one depending wholly on mechanism and chance. In recent years, there has been some tendency to revert to more or less mystical conceptions revolving about such phrases as “emergent evolution” and “creative evolution.” The writer must confess to a certain sympathy with such viewpoints philosophically but feels that they can have no place in an attempt at scientific analysis of the problem. One may recognize that the only reality directly experienced is that of mind, including choice, that mechanism is merely a term for regular behavior, and that there can be no ultimate explanation in terms of mechanism—merely an analytic description. Such a description, however, is the essential task of science and because of these very considerations, objective and subjective terms cannot be used in the same description without danger of something like 100 percent duplication. Whatever incompleteness is involved in scientific analysis applies to the simplest problems of mechanics as well as to evolution. It is present in most aggravated form, perhaps, in the development and behavior of individual organisms, but even here there seems to be no necessary limit (short of quantum phenomena) to the extent to which mechanistic analysis may be carried. An organism appears to be a system, linked up in such a way, through chains of trigger mechanisms, that a high degree of freedom of behavior as a whole merely requires departures from regularity of behavior among the ultimate parts, of the order of infinitesimals raised to powers as high as the lengths of the above chains. This view implies considerable limitations in the synthetic phases of science, but in any case it seems to have reached the point of demonstration in the field of quantum physics that prediction can be expressed only in terms of probabilities, decreasing with the period of time. As to evolution, its entities, species and ecologic systems, are much less closely knit than individual organisms.

One may conceive of the process as involving freedom, most readily traceable in the factor called here individual adaptability. This, however, is a subjective interpretation and can have no place in the objective scientific analysis of the problem.

## SUMMARY

The frequency of a given gene in a population may be modified by a number of conditions including recurrent mutation to and from it, migration, selection of various sorts and, far from least in importance, mere chance variation. Using  $q$  for gene frequency,  $v$  and  $u$  for mutation rates to and from the gene respectively,  $m$  for the exchange of population with neighboring groups with gene frequency  $q_m$ ,  $s$  for the selective advantage of the gene over its combined allelomorphs and  $N$  for the effective number in the breeding stock (much smaller as a rule than the actual number of adult individuals) the most probable change in gene frequency per generation may be written:

$$\Delta q = v(1 - q) - uq - m(q - q_m) + sq(1 - q)$$

and the array of probabilities for the next generation as  $[(1 - q - \Delta q)a + (q + \Delta q)A]^{2N}$ . The contribution of zygotic selection (reproductive rates of  $aa$ ,  $Aa$  and  $AA$  as  $1 - s^1 : 1 - hs^1 : 1$ ) is  $\Delta q = s^1 q(1 - q)[1 - q + h(2q - 1)]$ . In interpreting results it is necessary to recognize that the above coefficients are continually changing in value and especially that the selection coefficient of a particular gene is really a function not only of the relative frequencies and momentary selection coefficients of its different allelomorphs but also of the entire system of frequencies and selection coefficients of non-allelomorphs. Selection relates to the organism as a whole and its environment and not to genes as such. The mutation rate to a gene ( $v$ ) can usually be treated as of negligible magnitude assuming the prevalence of multiple allelomorphs.

In a population so large that chance variation is negligible, gene frequency reaches equilibrium when  $\Delta q = 0$ . Among special cases is that of opposing mutation rates ( $q = \frac{v}{u+v}$ ), of selection against both homozygotes ( $q = \frac{1-h}{1-2h}$ ), of mutation against genic selection ( $q = 1 - \frac{u}{s}$ ), of mutation against zygotic selection ( $q = 1 - \frac{u}{hs^1}$  unless  $h$  approaches 0, when  $q = 1 - \sqrt{\frac{u}{s}}$ ), of selection and migration ( $q = 1 - \frac{m}{s}(1 - q_m)$ ) or

$-\frac{mq_m}{s}$  if  $s$  is much greater than  $m$ ,  $q = q_m \left( 1 + \frac{s}{m}(1 - q_m) \right)$  if  $s$  is much smaller than  $m$ , while the values  $q = \sqrt{q_m}$  or  $1 - \sqrt{1 - q_m}$  when  $s = \pm m$  illustrate the intermediate case).

Gene frequency fluctuates about the equilibrium point in a distribution curve, the form of which depends on the relations between population number and the various pressures. The general formula in the case of a freely interbreeding group, assuming genic selection, is

$$y = Ce^{4Nsq}q^{4N[mq_m+v]-1}(1 - q)^{4N[m(1-q_m)+u]-1}.$$

The correlation between relatives is affected by the form of the distribution of gene frequencies through FISHER'S "dominance ratio." It appears that this is less than 0.20 in small populations under low selection but may even approach 1 in large populations under severe selection against recessives.

In a large population in which gene frequencies are always close to their equilibrium points, any change in conditions other than population number is followed by an approach toward the new equilibria at rates given by the  $\Delta q$ 's. Great reduction in population number is followed by fixation and loss of genes, each at the rate  $1/4N$  per generation, where  $N$  refers to the new population number. This applies either in a group of monoecious individuals with random fertilization or, approximately, in one equally divided between males and females (9.6 percent instead of 12.5 percent, however, under brother-sister mating,  $N=2$ ). More generally with an effective breeding stock of  $N_m$  males and  $N_f$  females, the rates of fixation and of loss are each approximately  $(1/16N_m + 1/16N_f)$  until mutation pressure at length brings equilibrium in a distribution approaching first the form  $y = C(1 - q)^{-1}$  with decay at rate  $u$  and ultimately  $Cq^{-1}(1 - q)^{-1}$ . The converse process, great increase in the size of a long inbred population, is followed by a slow approach toward the new equilibrium at a rate dependent in the early stages on mutation pressure.

With respect to genes which are indifferent to selection, the mean frequency is always  $q = v/(u + v)$ . The variance of characters, dependent on such genes, is proportional (at equilibrium) to population number up to about  $N = 1/4u$ . Beyond this, there is approach of variance to a limiting value.

In the presence of selection ( $s$  considerably greater than  $2u$ ) the mean frequency at equilibrium varies between approximate fixation of the favored genes ( $q = 1 - u/s$ ) in large populations and approximate, if not complete, fixation of mutant allelomorphs ( $q = v/(u + v)$ ) in small popula-

tions, the rate of change from one state to the other being the mutation rate ( $u$ ). A consequence is a slow but increasing tendency to decline in vigor in inbred stocks, to be distinguished from the relatively rapid but soon completed fixation process, described above as occurring at rate  $1/2N$ . The variance of characters in this as in the preceding case, is approximately proportional to population number up to a certain point ( $N$  less than  $1/4s$ ) and above this rapidly approaches a limiting value. Variance is inversely proportional to the severity of selection in large populations unless the selection is very slight but in small populations is little affected by selection unless the latter is very severe ( $s$  greater than  $1/4N$ ).

Evolution as a process of cumulative change depends on a proper balance of the conditions, which, at each level of organization—gene, chromosome, cell, individual, local race—make for genetic homogeneity or genetic heterogeneity of the species. While the basic factor of change—the infrequent, fortuitous, usually more or less injurious gene mutations, in themselves, appear to furnish an inadequate basis for evolution, the mechanism of cell division, with its occasional aberrations, and of nuclear fusion (at fertilization) followed at some time by reduction make it possible for a relatively small number of not too injurious mutations to provide an extensive field of actual variations. The type and rate of evolution in such a system depend on the balance among the evolutionary pressures considered here. In too small a population ( $1/4N$  much greater than  $u$  and  $s$ ) there is nearly complete fixation, little variation, little effect of selection and thus a static condition modified occasionally by chance fixation of new mutations leading inevitably to degeneration and extinction. In too large a freely interbreeding population ( $1/4N$  much less than  $u$  and  $s$ ) there is great variability but such a close approach to complete equilibrium of all gene frequencies that there is no evolution under static conditions. Change in conditions such as more severe selection, merely shifts all gene frequencies and for the most part reversibly, to new equilibrium points in which the population remains static as long as the new conditions persist. Such evolutionary change as occurs is an extremely slow adaptive process. In a population of intermediate size ( $1/4N$  of the order of  $u$ ) there is continual random shifting of gene frequencies and a consequent shifting of selection coefficients which leads to a relatively rapid, continuing, irreversible, and largely fortuitous, but not degenerative series of changes, even under static conditions. The rate is rapid only in comparison with the preceding cases, however, being limited by mutation pressure and thus requiring periods of the order of 100,000 generations for important changes.

Finally in a large population, divided and subdivided into partially isolated local races of small size, there is a continually shifting differentiation among the latter (*intensified by local differences in selection but occurring under uniform and static conditions*) which inevitably brings about an indefinitely continuing, irreversible, adaptive, and much more rapid evolution of the species. Complete isolation in this case, and more slowly in the preceding, originates new species differing for the most part in nonadaptive respects but is capable of initiating an adaptive radiation as well as of parallel orthogenetic lines, in accordance with the conditions. It is suggested, in conclusion, that the differing statistical situations to be expected among natural species are adequate to account for the different sorts of evolutionary processes which have been described, and that, in particular, conditions in nature are often such as to bring about the state of poise among *opposing tendencies on which an indefinitely continuing evolutionary process depends.*

## LITERATURE CITED

- BERNSTEIN, F., 1925 Zusammenfassende Betrachtungen über die erblichen Blutstrukturen des Menschen. *Z. induct. Abstamm.-u. VererbLehre* **37**: 237-269.
- CALDER, A., 1927 The role of inbreeding in the development of the Clydesdale breed of horse. *Proc. Roy. Soc. Edinb.* **47**: 118-140.
- EAST, E. M., 1918 The role of reproduction in evolution. *Amer. Nat.* **52**: 273-289.
- FISHER, R. A., 1918 The correlation between relatives on the supposition of Mendelian inheritance. *Trans. Roy. Soc., Edinb.* 52 part **2**: 399-433.
- 1922 On the dominance ratio. *Proc. Roy. Soc., Edinb.* **42**: 321-341.
- 1928 The possible modification of the response of the wild type to recurrent mutations. *Amer. Nat.*, **62**: 115-126.
- 1929 The evolution of dominance; reply to Professor Sewall Wright: *Amer. Nat.* **63**: 553-556.
- 1930 The genetical theory of natural selection. 272 pp. Oxford: Clarendon Press.
- HAGEDOORN, A. L., and HAGEDOORN A. C., 1921 The relative value of the processes causing evolution. 294 pp. The Hague: Martinus Nijhoff.
- HALDANE, J. B. S., 1924-1927 A mathematical theory of natural and artificial selection. Part I. *Trans. Camb. Phil. Soc.* **23**: 19-41. Part II. *Proc. Camb. Phil. Soc. (Biol. Sci)* **1**: 158-163. Part III. *Proc. Camb. Phil. Soc.* **23**: 363-372. Part IV. *Proc. Camb. Phil. Soc.* **23**: 607-615. Part V. *Proc. Camb. Phil. Soc.* **23**: 838-844.
- HARDY, G. H., 1908 Mendelian proportions in a mixed population. *Science* **28**: 49-50.
- JENNINGS, H. S., 1916 The numerical results of diverse systems of breeding. *Genetics* **1**: 53-89.
- JONES, D. F., 1917 Dominance of linked factors as a means of accounting for heterosis. *Genetics* **2**: 466-479.
- KEMP, W. B., 1929 Genetic equilibrium and selection. *Genetics* **14**: 85-127.
- LOTKA, A. J., 1925 *Elements of Physical Biology*. Baltimore: Williams and Wilkins.
- MCPHEE, H. C., and WRIGHT S., 1925, 1926 Mendelian analysis of the pure breeds of livestock. III. The Shorthorns. *J. Hered.*, **16**: 205-215. IV. The British dairy Shorthorns. *J. Hered.* **17**: 397-401.

- MORGAN, T. H., BRIDGES, C. B., STURTEVANT, A. H., 1925 The genetics of *Drosophila*. 262 pp. The Hague: Martinus Nijhoff.
- MULLER, H. J., 1922 Variation due to change in the individual gene. *Amer. Nat.* **56**: 32-50.  
1928 The measurement of gene mutation rate in *Drosophila*, its high variability, and its dependence upon temperature. *Genetics* **13**: 279-357.  
1929 The gene as the basis of life. *Proc. Int. Cong. Plant Sciences* **1**: 897-921.
- ROBBINS, R. B., 1918 Some applications of mathematics to breeding problems. III. *Genetics* **3**: 375-389.
- SMITH, A. D. B., 1926 Inbreeding in cattle and horses. *Eugen. Rev.* **14**: 189-204.
- WAHLUND, STEN., 1928 Zusammensetzung von Populationen und Korrelationserscheinungen vom Standpunkt der Vererbungslehre aus betrachtet. *Hereditas* **11**: 65-106.
- WEINBERG, W., 1909 Über Vererbungsgesetze beim Menschen. *Z. induct. Abstamm.-u. Vererb.-Lehre* **1**: 277-330.  
1910 Weiteres Beiträge zur Theorie der Vererbung. *Arch. Rass.-u. Ges. Biol.* **7**: 35-49, 169-173.
- WRIGHT, S., 1921 Systems of mating. *Genetics* **6**: 111-178.  
1922 Coefficients of inbreeding and relationship. *Amer. Nat.* **61**: 330-338.  
1922a The effects of inbreeding and crossbreeding on guinea-pigs. III. Crosses between highly inbred families. U. S. Dept. Agr. Bull. No. 1121.  
1923 Mendelian analysis of the pure breeds of live stock. Part I. *J. Hered.* **14**: 339-348. Part II, *J. Hered.* **14**: 405-422.  
1929 FISHER'S theory of dominance. *Amer. Nat.* **63**: 274-279.  
1929a The evolution of dominance. Comment on Doctor FISHER'S reply. *Amer. Nat.* **63**: 556-561.  
1929b Evolution in a Mendelian population. *Anat. Rec.* **44**: 287.
- WRIGHT, S., and MCPHEE, H. C., 1925 An approximate method of calculating coefficients inbreeding and relationship from livestock pedigrees. *J. Agric. Res.* **31**: 377-383.