

A theory of limits in artificial selection

BY A. ROBERTSON

A.R.C. Unit of Animal Genetics, Institute of Animal Genetics, Edinburgh

(Communicated by C. H. Waddington, F.R.S.—Received 14 March 1960)

(1) The paper presents a theory of selection limits in artificial selection. It is, however, developed primarily in terms of single genes.

(2) For a single gene with selective advantage s , the chance of fixation (the expected gene frequency at the limit) is a function only of Ns , where N is the effective population size. In artificial selection based on individual measurements, where the selection differential is \bar{x} standard deviations, the expected limit of individual selection in any population is a function only of $N\bar{x}$.

(3) For low values of $N\bar{x}$, the total advance by selection is, for additive genes, $2N$ times the gain in the first generation but may be much greater than this for recessives, particularly if their initial frequency is low.

(4) The half-life of any selection process will, for additive genes, not be greater than $1.4N$ generations but may for rare recessives equal $2N$.

(5) The effect of an initial period of selection or inbreeding or of both together on the limits in further selection is discussed. It appears that the effects of restrictions in population size on the selection limit may be a useful diagnostic tool in the laboratory.

(6) The treatment can be extended to deal with the limits of further selection after the crossing of replicate lines from the same population when the initial response has ceased.

(7) In a selection programme of individual selection of equal intensity in both sexes, the furthest limit should be attained when half the population is selected from each generation.

(8) The treatment can also be extended to include selection based on progeny or family records. Consideration of the optimum structure, as far as the limit is concerned, shows that the use of the information on relatives is always a sacrifice on the eventual limit for the sake of immediate gain in the early generations. The loss may, however, be small in large populations.

Current theories of artificial selection have been concerned almost entirely with the prediction of the rate of response of the population mean to selection pressures of different kinds. The selection may be expected to increase the frequency of favourable alleles until, in a large population, they eventually reach fixation. But if the population size is finite there is a possibility that one allele may be fixed by chance even though there is a more desirable one in the population. The smaller the population, the greater will this possibility be. In this paper, a theory of selection limits is developed taking into account at the same time the forces of selection and those of chance fixation. Some of the conclusions are not very precise, but, in so far as they indicate qualitatively what we might expect, they may be of value. They bear on many practical problems, for instance the attainment of the maximum possible advance from the existing variation with a population and also the retention of the potential for change in a population while keeping it in 'cold storage'.

It may of course happen that a population will reach a selection limit while still retaining genetic variation, due to continued selection for heterozygotes. It is hoped to deal with the effects of restricted population size in such a situation in a separate paper.

We have first to develop the theory with respect to individual genes whose selective advantage we know and then to modify the conclusions to deal with

the situation when we are merely selecting the animals extreme for some quantitative measurement. Two kinds of agencies affecting gene frequency are at work in any selection programme. The first is genetic sampling, or genetic drift, causing a random change from generation to generation in the frequency, q , of any gene. The mean change is zero and its variance is $[q(1-q)]/2N$, where N is the effective population size. The second is the directed change in gene frequency due to selection which we can write in the case of two alternative alleles as $sf(q)$, where s depends on the relative selective advantages of the three genotypes at that locus and $f(q)$ is a function of the gene frequency, depending on the type of gene action involved. The approach to such problems is essentially due to Wright (1931). He introduced the concept of the distribution of gene frequencies, which may be considered alternatively to refer to loci of the same kind and magnitude of effect in one population or to an individual locus in many such populations. The central concept in this present approach is that of the chance of fixation of the gene in question, to which Kimura (1957) has given the symbol $u(q)$ where q is the frequency of the gene in the initial population. u is then alternatively the proportion of equivalent loci which would be expected to be fixed in any line, perhaps the easiest model in considering artificial selection, or the proportion of replicate selected lines in which an individual gene would be expected to be fixed. u may then have a value intermediate between 0 and 1 even though homozygosis is complete. We may refer to $u(q)$ as the 'expected limit' where the extreme possible limit will be a value of 1. $u(q) - q$ is then the expected total change in gene frequency and, corresponding to it in artificial selection, we have the change in the character under selection when fixation is reached, which we shall describe as the 'total advance'. At the extreme, we have a population in which all desirable genes in the initial population have been fixed and we shall describe by the phrase 'possible advance' the difference in the selected character between this and the initial population.

Genetic sampling causes a broadening of the gene-frequency distribution leading to eventual fixation at the extreme values and selection causes a general shift of the distribution in the direction of selection. If we write $\phi(q, t)$ for the distribution of gene frequencies at time t , we can, using a continuous model, write down the process of change with time as

$$\frac{\partial \phi}{\partial t} = \frac{\partial^2}{\partial q^2} \left(\frac{q(1-q)\phi}{4N} \right) - \frac{\partial}{\partial q} (\phi sf(q)) \quad (1)$$

the first term on the right-hand side representing drift and the second selection. This may be rearranged as

$$\frac{\partial \phi}{\partial (t/N)} = \frac{\partial^2}{\partial q^2} \left(\frac{q(1-q)\phi}{4} \right) - \frac{\partial}{\partial q} (\phi N sf(q)). \quad (2)$$

Thus the change in ϕ at a particular value of q in an amount of time t/N is dependent only on Ns and on the initial function $\phi(q, 0)$. It then follows that the pattern of the change is determined by Ns and its time scale is directly proportional to N .

For a given value of q , the value of ϕ is a function of Ns and t/N , the exact form of the function being determined by the initial gene-frequency distribution. Now the pattern when $t = \infty$ is merely the proportion of loci (or replicate populations) in which the allele in question has been fixed. The chance of fixation of the allele is then determined by the initial distribution and Ns . If we start with a single locus in a single population, we can say that the chance of fixation is a function solely of Ns and the initial gene frequency.

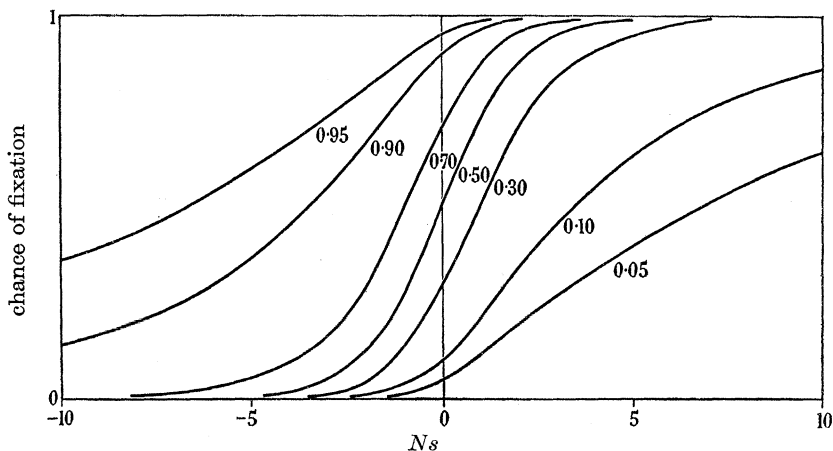


FIGURE 1. The chance of fixation of a gene acting additively. The curves are drawn for different initial gene frequencies.

The treatment that follows is an extension of some results of Kimura (1957) who developed explicit expressions for the chance of fixation. In the case of a pair of alleles with additive effects on selective advantage, i.e. the three genotypes have selective advantage $1 - \frac{1}{2}s$, 1 and $1 + \frac{1}{2}s$, respectively, the chance of fixation of a gene whose initial frequency is q is given by

$$u(q) = \frac{1 - e^{-2Nsq}}{1 - e^{-2Ns}}, \quad (3)$$

which is shown in figure 1. When $Ns = 0$, $u(q)$ equals q . This merely says that if there is no selection, the mean gene frequency is not changed. An expansion in terms of Ns gives.

$$u(q) = q + q(1-q)Ns + \dots \quad (4)$$

Thus the slope of the curves, when $Ns = 0$, is $q(1-q)$. As the total possible advance is $1-q$, it follows that the greater part of this will be achieved if Ns is greater than $1/q$. In fact, if $Nsq > 1$, it can be shown that more than 70% of the possible advance will be achieved and if $Nsq > 2$, more than 93%.

The value of $u(q)$ when Ns is small can be derived by a completely different approach. We then assume that the mean frequency will change little during the fixation process and that the mean heterozygosity declines by the usual proportion $1/2N$ each generation. In the first generation of selection, it can be shown that the

change in gene frequency is $(\frac{1}{2}s)q(1-q)$. But we know that the average value of $q(1-q)$ will decline by a fraction $1/2N$ each generation. Thus

$$\begin{aligned} u(q) - q &= \sum_{t=0}^{\infty} \frac{s}{2} q(1-q) \left(1 - \frac{1}{2N}\right)^t \\ &= Nsq(1-q). \end{aligned} \quad (5)$$

The total advance is thus $2N$ times the change in the first generation. For larger values of Ns , the total advance may be larger than this because at low initial gene frequencies the mean value of $q(1-q)$ may well increase during selection, in spite of the inbreeding, as the mean gene frequency increases. For low values of q and large values of Ns we have, by expansion of (3), $u(q) = 2Nsq$ so that the total advance approaches $4N$ times the change in the first generation.

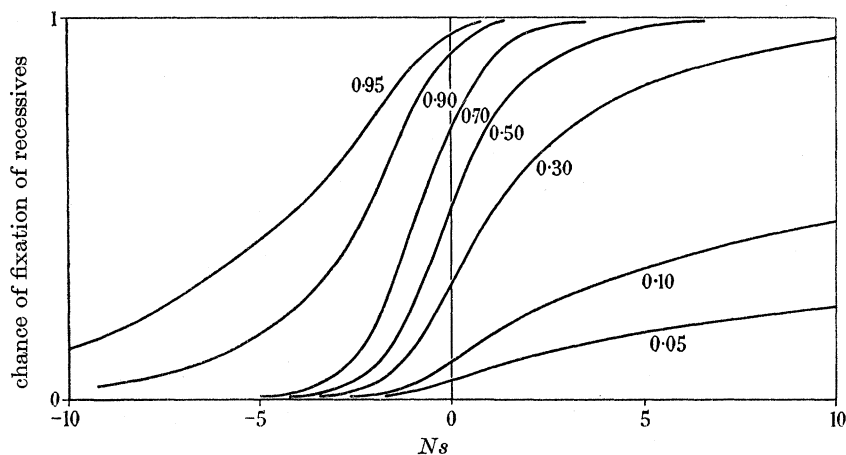


FIGURE 2. The chance of fixation of a recessive gene. The curves are drawn for different initial recessive frequencies.

If there is no additive gene action, the algebra is more complex. With selection for a recessive gene with frequency q where the homozygous recessive has a selective advantage s , we have $f(q) = q^2(1-q)$ and Kimura (1957) has shown that the chance of fixation is given by

$$u(q) = \frac{\int_0^q e^{-2Ns q^2} dq}{\int_0^1 e^{-2Ns q^2} dq}. \quad (6)$$

This function is shown in figure 2. An expansion is possible, giving

$$u(q) = q + \frac{2}{3} Nsq(1-q^2) + \dots \quad (7)$$

The alternative approach, assuming that $u(q) - q$ is small, necessitates the use of the moment generating matrix (Robertson 1952). This gives

$$\begin{aligned} u(q) - q &= s \sum_{t=0}^{\infty} \left[\frac{1}{2} q(1-q) \left(1 - \frac{1}{2N}\right)^t - \frac{1}{2} q(1-q)(1-2q) \left(1 - \frac{3}{2N}\right)^t \right] \\ &= \frac{2}{3} Nsq(1-q^2). \end{aligned} \quad (8)$$

Formulae (7) and (8) have obvious relevance to the chance that an inbred line will rid itself of deleterious recessives during inbreeding as $u(q)$ is the proportion of such recessives which will be fixed. We are then dealing with negative values of s . Thus if $\frac{2}{3}Ns$ is more negative than -1 , $u(q)$ will be close to zero, the gene will have been selected out and will not contribute to the depression of fitness on inbreeding. As would be expected, the lower the rate of inbreeding and the greater the gene effect, the greater the chance that a harmful recessive will be selected out. Then the lower the rate of inbreeding the smaller the effects of the individual harmful genes fixed by chance and the less the inbreeding depression when complete homozygosity is reached.

The ratio of expected total response to initial change in the first generation when Ns is small is $\frac{2}{3}Nsq(1-q^2)/sq^2(1-q) = 2N(1+q)/3q$, rather than $2N$ as in the additive case. This ratio can then be very much greater for recessives at low frequencies than for additive genes. This is a consequence of the increase of the genetic variance within lines due to low-frequency recessives up to inbreeding coefficients of 0.5 noted by Robertson (1952).

The case of equilibrium at an intermediate gene frequency resulting from selection for heterozygotes, ($f(q) = q(1-q)(q-q_0)$, where q_0 is the equilibrium frequency), is more complex as the selection may effectively prevent fixation. It is hoped to deal in detail with this and some related problems in another paper.

THE TIME SCALE OF THE SELECTION PROCESS

We have been discussing so far the limits of selection. How long will it take to get there? As the approach to the limit will be asymptotic, this question has not much meaning but it is useful to ask how long it will take the mean gene frequency to get half-way to the limit. In the language of physics, what is the 'half-life' of the selection process?

It was shown earlier (2) that the time scale is proportional to N , the effective population size, and the pattern of change is dependent only on Ns . The expected half-life in generations will then be a multiple of N , determined by Ns and the initial gene frequency. In the case of low values of Ns , we can obtain a very simple expression. Writing q_t as the mean gene frequency after t generations, we have from (5)

$$\begin{aligned} q_t - q &= \sum_0^t \frac{1}{2}sq(1-q) \left(1 - \frac{1}{2N}\right)^t \\ &= Nsq(1-q) (1 - e^{-t/2N}) \text{ approx.} \end{aligned} \quad (9)$$

The half-life is then given by $e^{-t/2N} = \frac{1}{2}$, or $t = 1.4N$ generations. In the case of a recessive gene, we have

$$\begin{aligned} q_t - q &= s \sum_0^t \left[\frac{1}{2}q(1-q) \left(1 - \frac{1}{2N}\right)^t - \frac{1}{2}q(1-q)(1-2q) \left(1 - \frac{3}{2N}\right)^t \right] \\ &= Nsq(1-q) \left[(1 - e^{-t/2N}) - \frac{1}{3}(1-2q)(1 - e^{-3t/2N}) \right] \text{ approx.} \end{aligned} \quad (10)$$

When $q = \frac{1}{2}$, the second term vanishes and the half-life is the same as in the additive case. As q approaches zero (i.e. the recessive is at low frequency) it can be shown

that the half-life is $2.12N$ generations and as q approaches unity, the factor approaches $1.03N$. We may then expect that, if Ns is low, the half-life will probably be between N and $2N$ generations.

If Ns is not small, the problem becomes extremely difficult to solve explicitly. It can, however, be explored by evaluating empirically the change in the gene-frequency distribution as the generations proceed. We then make use of the fact that the pattern of change is determined only by Ns . For a given value of N , there are $2N + 1$ possible gene frequencies. If a population has a certain frequency at a particular generation, the probability distribution of its frequency in the next generation can be calculated as a binomial distribution with index $2N$ and mean correspondingly modified by selection. We can then write down the transformation matrix to convert the gene-frequency distribution in one generation into that in the next. It only proved possible to deal with values of N up to 5 on a desk calculator because the total increases as N^3 .

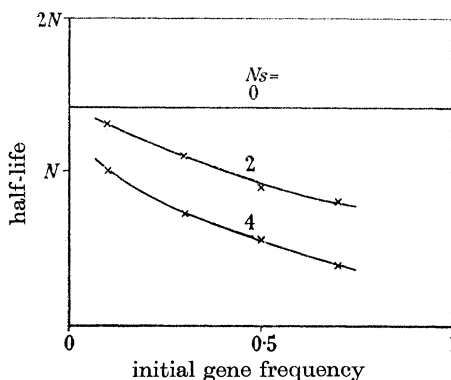


FIGURE 3. The average 'half-life' in generations of the selection process for a gene acting additively.

Some of the half-lives so calculated with $N = 5$ are given in figure 3. It appears that the half-life decreases continuously as Ns increases. In a sense this might have been expected. The higher Ns , the greater is the chance that the favourable allele becomes fixed by selection before it is lost from the line by sampling. It seems then that the values calculated theoretically for low Ns are probably upper limits. If the half-life of a selection programme is reached well before the range of N to $2N$, expected when the chance of fixation is not high, we may perhaps conclude that we have fixed all the desirable alleles.

THE EFFECT OF SELECTION AND INBREEDING ON SUBSEQUENT SELECTION LIMITS

We have been discussing so far the chance of fixation of a gene with a known initial frequency and have shown that it is dependent only on Ns . We now proceed to ask how this dependence on Ns is affected by an initial period of inbreeding or selection or of both together. It may be of advantage here to think of the gene-frequency distribution as referring to genes at different loci, all with the same selective advantage. In the initial population we shall assume that they all have

the same gene frequency. This frequency will be altered by the initial period of selection or inbreeding and we wish to know how the chance of fixation (or the proportion of the genes likely to be fixed) considered as a function of Ns , has been altered in this initial period. The detailed discussion will be devoted to the case of additive action.

There are three alternative treatments in the initial period that we shall consider.

(a) *Selection in a very large population*

The frequency of all genes will have been altered by the same amount and the curve of chance of fixation against Ns is that corresponding to the new frequency.

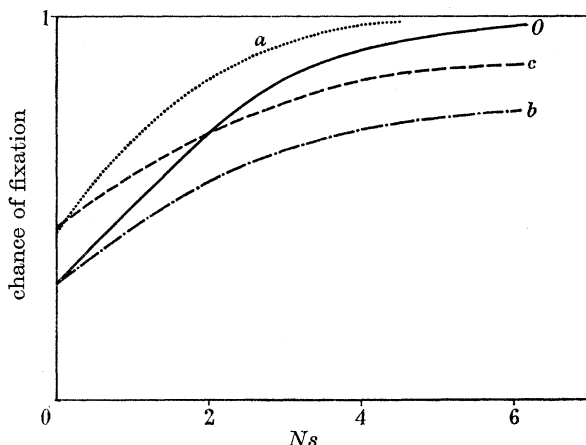


FIGURE 4. The effect of various treatments on the curve of chance of fixation against Ns for a gene with initial frequency of 0.3. The treatments are three generations of (a) selection with $s = 0.4$ in a large population, (b) restriction of effective population size to 5, (c) selection with $s = 0.4$ and effective population size 5. *O* = original.

(b) *Restriction of population size without selection*

The mean gene frequency will stay the same but different genes will have different initial frequencies. Some may be lost altogether from the population so that the ultimate limits of selection at high values of Ns will be reduced. This effect will be most marked at low initial gene frequencies. For additive genes, the slope of $u(q)$ plotted against Ns at low values of Ns will be reduced due to the decline in heterozygosity. The slope will also be less for recessives.

(c) *Restricted population size with selection*

We have now a mixture of the two effects. The mean gene frequency will be higher when $Ns = 0$ because of the selection but the ultimate limit will be reduced because of chance fixation in the initial period. But the new curve of $u(q)$ against Ns must intersect the old one at the point corresponding to the value of Ns used in the initial period. The further selection process is then merely a continuation of the old and will have the same expected limit.

Figure 4 shows the curves of chance of fixation against Ns for an initial gene frequency of 0.3 for the initial population and after three generations of (a)

selection with $s = 0.4$ in a large population (*b*) restriction of population size to $N = 5$ and (*c*) selection and restriction together. The curves for the last two were obtained by calculating the gene-frequency distribution after the three generations using the transformation matrix as in the calculation of the half-life. The curves illustrate well the points made earlier.

The effect of restriction or 'bottleneck' in populations for some generations may be enlarged upon in relation to initial gene frequency. The lower the initial frequency the greater the chance a gene will be lost from the population by a sudden reduction in population size. Figure 5 shows the curve of limit against Ns for the 'bottleneck' of restriction of parents to a single pair, and of such a restriction for three consecutive generations, in each case followed by expansion. The effect of

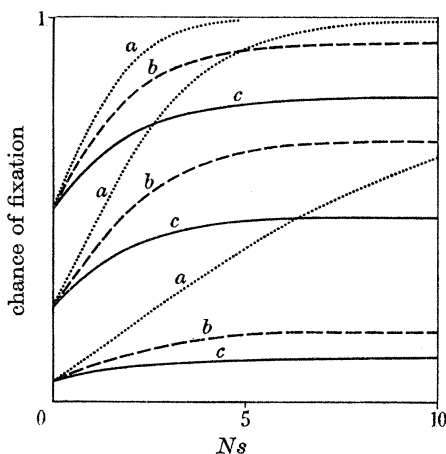


FIGURE 5. The effects of 'bottlenecks' in population size on the curve of chance of fixation against calculated for initial gene frequencies of 0.1, 0.3 and 0.5. (*a*) initial population, (*b*) restriction to a single mating for one generation only, (*c*) restriction to a single mating for three consecutive generations.

the bottleneck on the maximum advance possible in further selection is very marked for the genes with low initial frequency. If the initial frequency is 0.5, a single restriction of parents to one pair reduces the possible further advance to 0.80, so that the possible advance is only 60% of that in the initial population. If the initial gene frequency is 0.05, such a restriction reduces the expected ultimate advance to 0.185 so that the possible advance is only 14% of that in the initial population. But a further two generations of restriction only reduce the possible advance by a further factor of two. In the first generation, many such genes would be lost altogether, but those that were retained would have their mean frequency increased to at least 0.25 so that further restriction would not have such a large effect. In selecting from a previously unselected large population, it is the first generation that is critical in losing low-frequency genes. Thereafter the population has its segregating genes at higher frequencies.

As a consequence, for genes with an initial low frequency in populations that have been through a bottleneck, the values of Ns necessary to attain a major part

of the possible advance will be less than before the restriction. In this they are similar to selected lines. In both cases, the genes that are still segregating are at a higher frequency than they were before. The figure shows that to get 70 % of the distance to the possible limit, the Ns value should be 12.2 if the initial gene frequencies are 0.05 but declines to 3.0 after one restriction to a single pair mating.

Highly selected populations or those which have passed through a severe 'bottleneck' in population size will be tolerant of any further size restrictions in the sense that the desirable alleles will be harder to lose because, if they are present at all, they will have a reasonable frequency. This has a paradoxical practical consequence to the storage of populations for possible use in a selection programme, a problem now facing many poultry breeders. The more highly selected a strain the smaller the numbers needed for keeping it. Any desirable genes that are still segregating are probably at a high frequency. Much more care should be given to a completely unselected strain because there the desirable alleles are more likely to be at low frequency and therefore to be lost by accident. The extreme type of population of this kind is one in which the desirable alleles are all at very low frequency, as would happen in attempting to produce new useful variation by irradiation. As Dempster (1958) has pointed out, it is then very important to keep the population size high in storage and in the early generations of selection.

THE CROSSING OF SELECTED LINES

If we cross together two selected lines from the same population after they have reached fixation, we may expect to make further progress in the cross if either line contained a desirable gene which the other did not. Let us assume that we take a series of lines which reached fixation with a given value of Ns and cross them together in pairs. What can we predict about the limits to further selection as a function of Ns ? If the chance of fixation in the initial selection is u , then this will be the expected gene frequency when Ns is zero. We may expect further that a proportion u^2 of the pairs will both have the gene in question, in $2u(1-u)$ one will have it and one will not and in $(1-u)^2$ neither will have it. The possible limit for high values of Ns is $1 - (1-u)^2$. Any further gain will come from the $2u(1-u)$ pairs in which the gene frequency is one-half.

We therefore know the values for $Ns = 0$ and $Ns = \infty$. A surprising result holds for additive genes if the further selection from the cross has twice the value of Ns as had the original selection. The expected limit is then exactly the same as the limit for selection from the original population with twice the original value of Ns . The result in fact holds more generally. If we make populations by crossing any number of lines together and then use a proportionally higher value of Ns , we have the same expected limit as if we had used the higher value all the time.

These results only hold approximately for recessive genes. The discrepancies are not large but, in single crosses, the expected limit on selecting the cross with double the original value is for high-frequency recessives rather higher than if all the selection had been at the higher value and rather lower for low-frequency recessives.

SELECTION FOR A QUANTITATIVE CHARACTER

We have been dealing so far with genes whose selective advantage we know. But in many of our laboratory selection experiments, we select animals on the basis of their measurement for some 'metric character'. We presume that we increase the frequency of the desirable genes, but we know neither their frequency nor the selective advantage that we confer on them by our artificial selection. We merely observe that the mean of the population changes.

We can take the first step in applying the earlier results by using a formula originally derived by Haldane (1931). He showed that under artificial selection the selective advantage associated with any small difference on the metric scale, on which selection is based, is equal to that difference multiplied by I/σ^2 , where I is the superiority of chosen parents above the mean of the population and σ^2 is the phenotypic variance in the population. If we express the intensity of selection as a dimensionless character, by putting $\bar{i} = I/\sigma$, the factor becomes \bar{i}/σ . For a gene acting additively with a difference of 'a' units on the metric scale between the mean of the two homozygotes, we then know that it will act additively as far as concerns selective advantage under artificial selection and that

$$s = \frac{\bar{i}a}{\sigma}.$$

We saw that, for all individual genes, the chance of fixation is a function of Ns and the initial frequency q . In artificial selection, we may write

$$u(q) = f(N\bar{i}a, q)$$

and since the mean phenotype at fixation can, in the absence of interaction between genes at different loci, be written by summing over loci $\Sigma au(q) = \Sigma af(N\bar{i}a, q)$ it follows that in any population the expected limit of selection is a function only of $N\bar{i}$. The exact form of this function will depend on the distribution of gene frequencies and effects and on the type of gene action involved.

We saw earlier that, for an additive gene, at least 70% of the possible gain in gene frequency at the limit would be obtained provided $Nsq > 1$. In artificial selection, the condition may be written $N\bar{i}aq > \sigma$. At low values of $N\bar{i}$ we can only be sure to fix genes either with large effect or with high frequency. As $N\bar{i}$ increases we begin to fix the rarer genes with smaller effects.

But in an artificial selection programme we do not observe changes in gene frequency, we only see a change in the mean of the measurement in the population. The total change in the mean, $X - \bar{X}$, will for additive genes be given by

$$\begin{aligned} X - \bar{X} &= \Sigma a(u(q) - q) \\ &= \Sigma a(q(1 - q)Ns + \dots) \\ &= 2N\bar{i} \Sigma \frac{a^2 q(1 - q)}{2\sigma} \quad \text{if } N\bar{i} \text{ is low} \\ &= N\bar{i} \frac{2\sigma^2}{\sigma}, \end{aligned}$$

where σ_g^2 is the additive genetic variance. We can therefore predict the slope of the curve of $X - \bar{X}$ plotted against $N\bar{i}$ for low values. As in the case of single genes, the total advance is $2N$ times the response in the first generation ($\bar{i}\sigma_g^2/\sigma$) as Dempster (1955) has pointed out.

Again, as with single genes, this does not hold for recessives. Then

$$\begin{aligned} X - \bar{X} &= \Sigma a(u(q) - q^2) \\ &= \Sigma \left(aq(1-q) + \frac{2N\bar{i}}{3} \frac{a^2q(1-q^2)}{\sigma} + \dots \right) \end{aligned}$$

of which the first term is the inbreeding depression. The coefficient of $N\bar{i}$, $[2\Sigma a^2q(1-q^2)]/3\sigma$, may be much greater than $2\sigma_g^2/\sigma$, which equals $[4\Sigma a^2q^3(1-q)]/\sigma$, especially for recessives at low frequency, so that the total advance due to recessives may be much greater than $2N$ times the response in the first generation.

The possible advance in the case of additive genes is given by the expression $\Sigma a(1-q)$ and this we cannot predict merely by measuring the genetic variance in the initial population. The smaller the number of genes contributing to any given additive genetic variance (and in consequence the greater their individual average effect) the lower will be the possible advance by selection and the quicker will it be reached. Though we can in some measure predict the initial slope of the curve of limit against $N\bar{i}$ there is no way in which we can predict the possible limit at high values of $N\bar{i}$.

Some of these points are illustrated in figures 6 and 7 which show the curves of advance plotted against $N\bar{i}$ for two imaginary populations. In figure 6, all genes have the same initial frequency of 0.5, but it is assumed that the advance at high values of $N\bar{i}$ is contributed equally by three classes of genes with a/σ values of 0.5, 0.3 and 0.1, respectively. In figure 7, it is assumed that $a/\sigma = 0.3$ for all genes but that the ultimate advance is contributed equally by genes with frequencies of 0.5, 0.3 and 0.1, respectively. It will be seen that the further increase in limit when $N\bar{i}$ is high is due almost entirely to the genes with small effects or low initial frequencies.

We are now in a position to take over into the context of artificial selection the results that were obtained for single genes. We cannot now observe the chance of fixation, as we cannot observe gene frequency. If $u(q)$ is very different from unity for many genes, we will notice that replicate lines from the same initial population will be very different in the limit they reach. The variation will be greatest when the average value of u for the different genes controlling the character is in the neighbourhood of 0.5. The variation between replicates will again be a function of $N\bar{i}$, but the type of function will depend on the gene frequencies in the initial population. If they are low, the variation in limit between replicates may pass through a maximum as $N\bar{i}$ increases but will decline to zero at high values. This is important in any experimental work on this problem as many replicates may be needed at low values of $N\bar{i}$. The possibility of utilizing the variation between replicate programmes may too have important consequences in practical animal breeding.

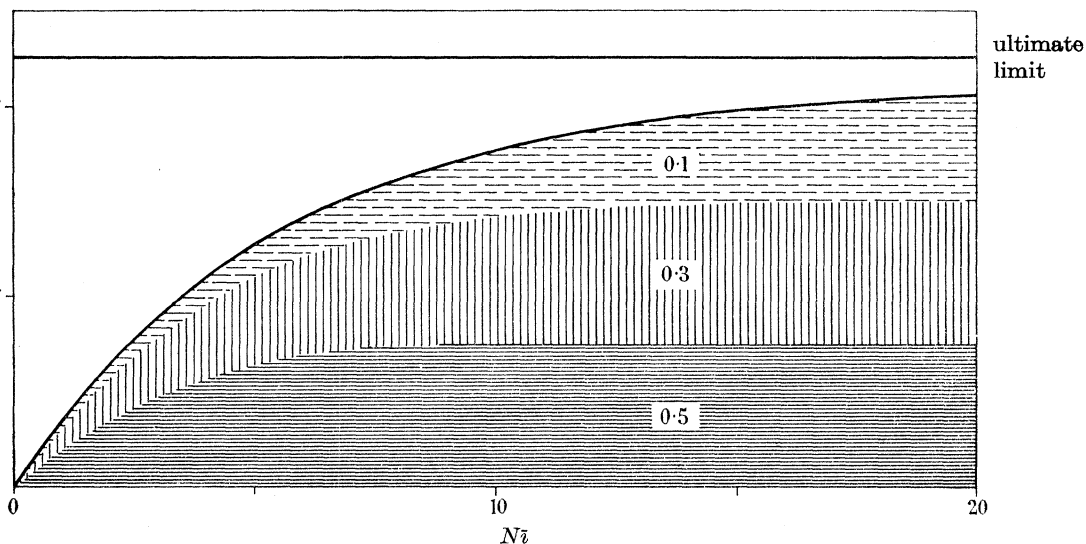


FIGURE 6. The expected limits to artificial selection in a population in which all genes have initial frequency 0.5 and in which the possible advance is contributed equally by genes with $a/\sigma = 0.1, 0.3$ and 0.5 , respectively.

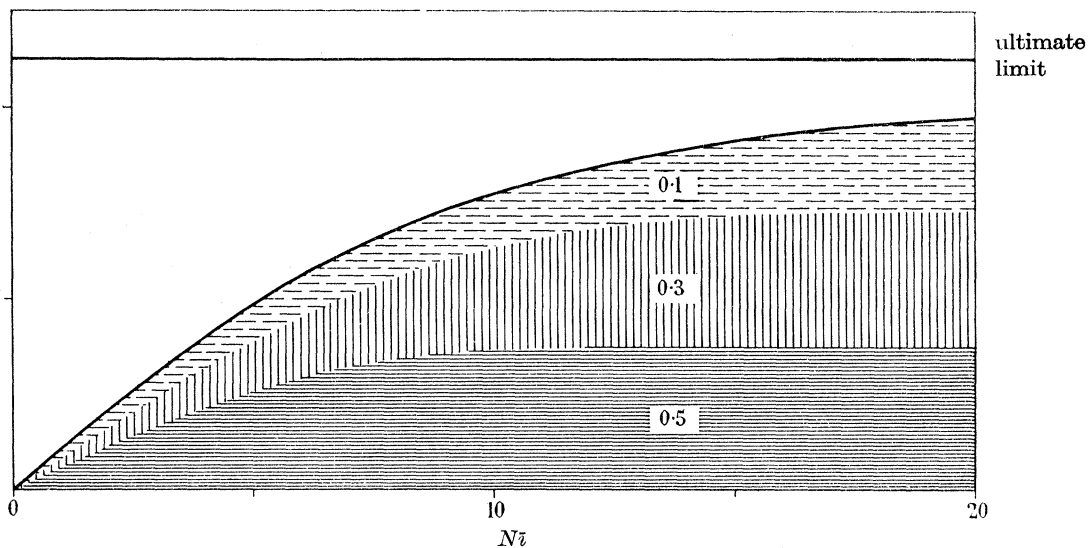


FIGURE 7. Similar to figure 6 except all genes now have $a/\sigma = 0.3$ and the possible advance has equal contributions from genes with frequency $0.1, 0.3$ and 0.5 .

The discussion of the time-scale of the selection process is immediately relevant to the quantitative case. We may then expect that selection programmes should have an upper limit to their half-life in the region of $1.4N$ generations if the genetic variation is mainly additive. If the half-life is reached earlier than this, we may presume that the majority of desirable alleles have been fixed and that there is little point in crossing two such lines derived from the same base population.

The effect of the initial treatment of the population on the curves of advance as a function of $N\bar{i}$ is directly referable from the single gene to the quantitative situation. The section on the effects of a restriction in population size on the possible limits at high $N\bar{i}$ values may be important here. In dealing with continuous variation, it is usual to describe the situation in terms of variance components. But it is very difficult to penetrate these to discover anything about the effects and frequencies of the genes which give rise to them. In the effect of a 'bottleneck' in population size on the ultimate limit of selection, we have a phenomenon which depends almost entirely on the frequencies of the desirable genes in the initial population and hardly at all on the magnitude of their individual effects. The investigation of the effects of bottlenecks on selection limits may thus be valuable in the analysis of laboratory populations.

These results are also relevant to the problem of how to get the maximum possible advance from a given initial population. To stand a high chance of fixing all the rare but desirable genes we shall start obviously with as high a value of $N\bar{i}$ as we could. How large this should be will depend on the previous history of the population. If it has been previously selected or restricted in population size, the need would not be so great. Then as selection proceeded, the size of the programme could be reduced because the frequency of the desirable genes would continually be increasing. Unfortunately we cannot give a recipe of how rapid this reduction could be.

The control we have over $N\bar{i}$ lies mostly in N , the effective population size, rather than in \bar{i} , the selection intensity. In most selection programmes that are at all efficient, \bar{i} lies between 1 and 2, these figures corresponding to a proportion selected, p , of 40% and 6%, respectively.

Occasionally, a selection programme is arranged so that the genetic contribution of all parents to the next generation is equal in order to minimize the increase in inbreeding each generation. It is rather doubtful whether this would affect the total advance by selection. It prevents the loss of variation by chance fixation but at the price of exposing only a part of the genetic variation to selection. In the simplest case of each mating contributing one male and one female as parents of the next generation, the effective breeding size of the population is twice the actual size. But the selection acts only on the variation within full-sib groups, i.e. half the total additive variance. The limit would not then be changed as the increase in N is exactly balanced by the decrease in the effective value of σ_g^2 . This argument would not hold if there were considerable non-genetic differences between full-sib groups when the accuracy of the selection might thereby be increased.

THE OPTIMUM INTENSITY OF INDIVIDUAL SELECTION

Suppose we have a selection programme in which we measure T animals and select the proportion p that are highest for some metric character. What is the optimum value of p ?

If the character is normally distributed, we may put $\bar{i} = z/p$, where z is the ordinate of the unit normal curve at the point where the area cut off is p . But N , the number of parents selected, is equal to Tp so that $N\bar{i} = Tz$. z has a maximum when

$p = \frac{1}{2}$ so that the greatest advance will be attained when on half the population is selected each generation, as Dempster (1955) has shown. But the maximum of the curve may be very flat, especially when T is large, because of the asymptotic nature of the curve of limit against $N\bar{v}$. Figure 8, which is based on the results for the synthetic population given in figure 7, shows the expected limit plotted against p for two different values of T . When $T = 50$, the curve has become extremely flat-topped. In practice, the problem is to find the value which will combine a high rate of initial response with a reasonable approach to the ultimate possible limit.

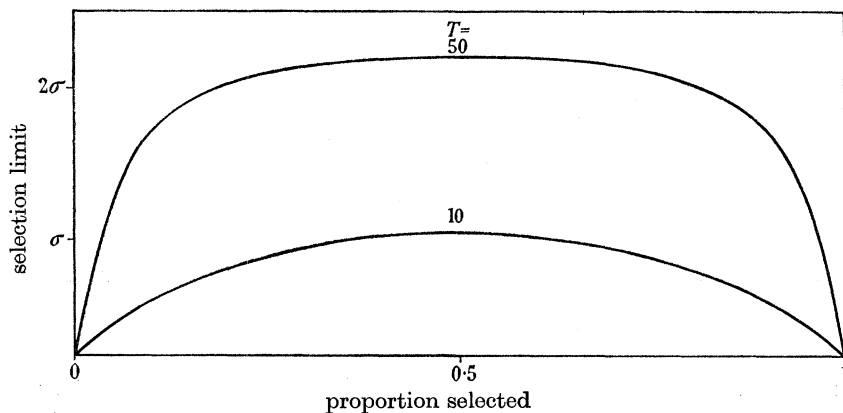


FIGURE 8. The expected limits to individual selection in the population in figure 6 when the number of animals measured is T .

THE EXTENSION TO FAMILY SELECTION

It is a frequent practice in a selection programme to use the records of the relatives of the animals under selection. It can be shown that, if the information on which selection is based has a correlation of r_{IG} with the animal's breeding value, the selective advantage of an additive gene is $\bar{v}ar_{IG}/h$, where h^2 is the heritability of the measurement under selection. Thus the use of measurements on relatives in a selective programme changes by the same factor the selection pressure on all the genes affecting the character. If we have the same response in two large selection programmes of different kinds from the same initial population, we may expect the gene frequencies in the two lines to be very similar. We can now generalize our earlier results to state that the selection limit is a function of $N\bar{v}r_{IG}/h$. This will not apply if other characters are taken into account in selection. The relative selection pressures on the different genes may then be altered, and the effect on selection limits cannot be predicted. Finally, if the value of \bar{v} is different in the two sexes the value to be inserted in the formula is the arithmetic mean of the two.

The effect of different kinds of selection programme may be illustrated by the following table which gives the value of $N\bar{v}r_{IG}/h$ for some of the different kinds of selection programmes for bristles in *Drosophila* which have been carried out in this laboratory. Following Crow (1954), we shall assume that in mass-mating populations of *Drosophila* the effective population size is 0.55 of the actual size.

The character selected for will be assumed to have a heritability of 0.50. The symbol 20/100 means that the extreme 20 of the 100 flies measured were chosen as parents.

selection method	intensity	$N\bar{r}_{IG}/h$
individual	20/100	$\left. \begin{array}{l} 31 \\ 21 \\ 8 \\ 10 \\ 2.4 \end{array} \right\}$
	20/50	
	20/25	
	10/25	
	1/5	
half-sib family (size 20)	2/10	9
full-sib family (size 10)	4/20	10

This leads naturally to the question of the optimum values of selection intensity and family size, if selection is based on family average, in order to achieve the greatest advance. The not very useful answer turns out to be that, for a given number of animals measured, the highest limit (determined by the maximum value of $N\bar{r}_{IG}$) is reached when one-half of the families are selected and the family size is one, which would mean that selection was based only on the individual's own measurement. In other words, a family selection or progeny testing programme always involves some sacrifice of ultimate response for the sake of greater immediate gain.

DEFECTS OF THE MODEL

From the point of view of the frequencies of single genes, the model is clearly of direct applicability. But, in transferring the conclusions to selection for quantitative characters, several hidden assumptions have been made which should now be mentioned.

The first one is that a/σ will remain constant throughout the selection. On simple theory, we would expect the genetic part of σ to decline gradually as the genes at other loci become fixed. On the other hand, as the level of homozygosity rises, the environmental part of σ might be expected to increase in some characters. And in many selection experiments there is evidence that even the genetic part does not decline greatly as the limit is reached because of selection for heterozygotes. Rather of the same kind are problems of scale. These are almost impossible to cope with in a completely logical and satisfactory manner and each case has to be treated on its merits. But scaling difficulties should be more of a nuisance in discussing the extent of the selection limit rather than how long it takes to get there.

Linkage presents the second problem. One might expect *a priori* that the greater the number of generations over which a given selection differential was spread, the greater the response because of the high probability of recombination within chromosomes. If this is true, one might expect that the optimum proportion selected in individual selection would be rather less than the value of one-half which holds for the independent segregation of genes. But the theoretical treatment of linkage and selection in a population of limited size is a complex one. The use of Monte Carlo methods with digital computers, now being tried by Fraser (1957) and by Cockerham & Martin (1960), is probably the most satisfactory way of handling the problem.

We must also consider natural selection, probably for heterozygotes, in the

sense that any population may have many gene frequencies held in equilibrium. The effect of this tendency to return to the original set of gene frequencies will be to reduce the effective value of \bar{i} by a constant proportion (see Robertson 1956, p. 246). The optimum proportion selected in individual selection would then remain in the neighbourhood of 0.5.

The problem has been discussed entirely in terms of two alternative alleles. The existence of many alleles at each locus tremendously complicates the detailed treatment but some useful statements can still be made. It will still be possible to recast the basic differential equations, similar to (1), into a form similar to (2) so that again the pattern of change will be determined in artificial selection by $N\bar{i}$ and the time scale will be proportional to N . The limits of selection in any population will then still be a function of $N\bar{i}$. When $N\bar{i}$ is small, we can still say that the slope of the curve of advance against $N\bar{i}$ will have slope $2\sigma_g^2/\sigma$.

Finally, the effective population size in any artificial selection programme may be dependent on the intensity of selection. All parents chosen in one generation will not have an equal chance of contributing progeny to the group chosen in the next. The parents with a higher breeding value will have a higher chance of contributing than those with a low breeding value. A theoretical examination has been made of this effect in individual selection which will be published elsewhere in detail. It has been found that the ratio of the actual number of parents to the effective number is very roughly equal to $1 + K\bar{i}^2h^2$ where K depends on the relationship within families. This factor would have to be borne in mind in a more detailed discussion of the problem.

It must be emphasized that this theoretical investigation has probably its real value not in predicting exactly what is going to happen in reality but in enabling one to design experiments on selection limits and to interpret them when we have done this.

I am much indebted to my colleagues for their critical reading of the manuscript, in particular to Dr D. S. Falconer.

REFERENCES

- Cockerham, C. C. & Martin, F. G. 1960 *The IBM 650 and its uses in Selective Studies: in Biometrical Genetics*. Pergamon Press (In the Press).
- Crow, J. F. 1954 Breeding structure of populations; in *Statistics and Mathematics in Biology*, ed. Kempthorne, Bancroft, Gower and Lush, Iowa State College Press, Ames, Ia.
- Dempster, E. R. 1955 Genetic models in relation to animal breeding problems. *Biometrics*, **11**, 535-6.
- Dempster, E. R. 1958 The fate of mutations in closed populations: in *Proc. 11th Pacific Poultry Breeders Roundtable*. University of California, California: Davis.
- Fraser, A. S. 1957 The simulation of genetic systems by automatic digital computers. II. The effects of linkage on the rates of advance under selection. *Aust. J. Biol. Sci.* **10**, 492-9.
- Haldane, J. B. S. 1931 A mathematical theory of natural and artificial selection. VII. Selection intensity as a function of mortality rate. *Proc. Camb. Phil. Soc.* **27**, 131-6.
- Kimura, M. 1957 Some problems of stochastic processes in genetics. *Ann. Math. Stat.* **28**, 882-901.
- Robertson, A. 1952 The effect of inbreeding on the variance due to recessive genes. *Genetics*, **37**, 189-207.
- Robertson, A. 1956 The effect of selection against extreme deviants based on deviation or on homozygosis. *J. Genet.* **54**, 236-48.
- Wright, S. 1931 Evolution in mendelian populations. *Genetics*, **16**, 97-159.