

THE EFFECTS OF INBREEDING AT LOCI WITH HETEROZYGOTE ADVANTAGE

W. G. HILL AND ALAN ROBERTSON*

Institute of Animal Genetics, Edinburgh 9

Received May 20, 1968

SINCE the early studies of FISHER and WRIGHT the theory of selection within populations of finite size has received much attention. KIMURA (1964) has reviewed the part of the theory that is based on continuous models in which it is usually assumed that individuals mate at random within small closed sub-populations or lines. In such a situation we are concerned with the distribution of the frequency of individual genes over many replicate lines, or, equivalently, the distribution of the frequency of identical genes within the same line. In this report we study selection favouring heterozygous individuals with random mating within lines and no selection or crossing occurring between lines. The model for inbreeding which we discuss must be distinguished from an alternative situation, perhaps more common in plants, in which inbreeding occurs within an infinitely large population as a result of non-random mating, for example by selfing or mixed selfing and outcrossing. In the latter type of model, selection also may occur between sublimes and recurrent mutation is not required for equilibria of gene frequency to occur without fixation, whereas it is in our model. These equilibrium situations have been analysed recently in some detail by ALLARD and co-workers. Many of their results for single loci are reviewed by JAIN and WORKMAN (1967) and analysis of a two locus model is given by JAIN and ALLARD (1966).

The effect of selection for heterozygous individuals in small lines when there is no between-line selection has been studied by REEVE (1955) using transition probability matrices for mating types in lines of only a few individuals, and by ROBERTSON (1962). The latter considered two situations—firstly when there is a balance between mutation and fixation and secondly when, in the absence of mutation, the amount of heterozygosity is declining at a steady rate. In both, the critical factor proved to be the equilibrium gene frequency, which depends on the relative fitness of the two homozygotes. If the equilibrium frequency lies outside the range 0.2 to 0.8 then selection may have an effect opposite to that usually expected and increase the rate of fixation.

In the present paper we shall be concerned with the intermediate stages of selection for the heterozygote in small lines with a known initial gene frequency. Selection may alter the mean gene frequency and the proportion of heterozygotes

* Member of the Agricultural Research Council Unit of Animal Genetics.

as well as the mean of quantitative traits. Two particular situations are of interest. In the context of natural selection, we may be concerned with the effect on fitness due to such loci when a large population is suddenly reduced in size. The initial gene frequency can then be assumed to be that at equilibrium in large populations. We might then speak of the effect of inbreeding on fitness though it is in fact the joint effect of natural selection and inbreeding.

The second situation is one involving artificial selection in which the heterozygote has an advantage because it is the best of the three genotypes for the character under selection. We now have no reason to assume any particular initial gene frequency.

MODEL

We shall consider only the case of two alleles at a single locus at which there is no mutation and which is not linked to other loci under selection. Let us assume that the relative selective advantages of the genotypes A_1A_1 , A_1A_2 and A_2A_2 are $1-s_1$, 1 and $1-s_2$, respectively. Let q denote the frequency of the A_1 allele and let \bar{q} , given by $\bar{q} = s_2/(s_1+s_2)$, be the equilibrium frequency in large populations. We shall only consider values of $\bar{q} \leq 0.5$ since there is symmetry about this point.

Transition matrix for monecious individuals: In the model which we shall investigate in most detail we assume that there are non-overlapping generations and that the parents comprise N monecious individuals which undergo random mating including random selfing. At some generation t let there be i A_1 alleles among the $2N$ alleles at the A locus in the adults, where $0 \leq i \leq 2N$. For brevity let $q = i/2N$. The genotypic frequencies among the zygotes at generation $t+1$ will be q^2 , $2q(1-q)$ and $(1-q)^2$ for A_1A_1 , A_1A_2 and A_2A_2 individuals, respectively. These genotypic frequencies depend only on the gene frequencies in their parents since there is random mating. Assuming that selection acts through differences in viability, the N individuals which become parents of the next generation will have a multinomial distribution of genotypic frequencies.

$$f_i(x, \gamma, z) = \binom{N}{x \ \gamma \ z} \left\{ \frac{q^2(1-s_1)}{\bar{w}} \right\}^x \left\{ \frac{2q(1-q)}{\bar{w}} \right\}^\gamma \left\{ \frac{(1-q)^2(1-s_2)}{\bar{w}} \right\}^z$$

where $f_i(x, \gamma, z)$ is the probability that there are x A_1A_1 , γ A_1A_2 and z A_2A_2 individuals surviving, with $q = i/2N$. The average fitness, \bar{w} , is

$$\begin{aligned} \bar{w} &= 1 - s_1q^2 - s_2(1-q)^2 \\ &= 1 - (s_1 + s_2) [\bar{q}(1-\bar{q}) + (q-\bar{q})^2] \end{aligned} \quad (1)$$

The probability that the N survivors have exactly j A_1 alleles is given by summation of the probabilities of all combinations of genotypic frequencies for which $2x + \gamma = j$. Thus we obtain p_{ij} , the probability that there are j A_1 alleles at generation $t+1$ given that there were i at generation t as

$$p_{ij} = \sum_{\substack{2x+\gamma \\ =j}} f_i(x, \gamma, z), \quad i, j = 0, \dots, 2N$$

We let \mathbf{P} be the transition probability matrix with elements p_{ij} . Since the selective

values are assumed to be independent of generation, \mathbf{P} is independent of generation number, t , also.

The expected change in gene frequency, $\delta q = E(j/2N - q|q = i/2N)$, is

$$\delta q = -(s_1 + s_2)q(1 - q)(q - \bar{q})/\bar{w}$$

which is the usual formula for response with a model of heterozygote advantage.

The expected fitness, gene frequency and heterozygosity were computed in successive generations by repeated multiplication using the transition matrix. For example, let the expected gene frequency at generation t , conditional on the initial frequency being q_0 , be $E(q_t|q_0)$ or simply $E(q)$. Also let $\mathbf{v}_{(t)}$ be a vector with elements $v_{i(t)} = E(q_t|q_0 = i/2N)$.

Thus

$$v_{i(0)} = i/2N, i = 0, \dots, 2N,$$

and we obtain

$$\mathbf{v}_{(1)} = \mathbf{P}\mathbf{v}_{(0)},$$

$$\mathbf{v}_{(2)} = \mathbf{P}\mathbf{v}_{(1)}$$

and, in general,

$$\mathbf{v}_{(t)} = \mathbf{P}\mathbf{v}_{(t-1)} \tag{2}$$

Iteration of (2) was repeated on a computer for up to $t = 8N$ generations, but was terminated earlier if there was almost complete fixation or the distribution of gene frequency among lines still segregating appeared to reach a state of steady decline. Then the value of λ , given by

$$\lambda = [v_{i(t)} - v_{i(t-1)}]/[v_{i(t-1)} - v_{i(t-2)}]$$

is constant for sufficiently large t and all i , $0 \leq i \leq 2N$. Results for later generations were obtained by assuming that the steady state had been reached, computing the dominant non-unit latent root, λ , and using this to predict subsequent changes. The expected heterozygosity, $E[2q(1 - q)]$ within lines was computed in a similar manner, with $v_{(i)0}$ becoming $2(i/2N)(1 - i/2N)$. The mean fitness (equation 1) is a linear function of $(q - \bar{q})^2$, with high values denoting low fitness, and, for simplicity, fitness has been expressed in this form. Since

$$E[(q - \bar{q})^2] = \bar{q}^2 + (1 - 2\bar{q})E(q) - E[q(1 - q)] \tag{3}$$

the expected fitness could be evaluated from the expected gene frequency and heterozygosity.

Transition matrix for dieocious individuals: Although the model with monocious individuals lends itself to simple numerical evaluation on a computer, it does not represent the real situation in most species. A model with two distinct sexes and random mating between the two sexes was therefore investigated for small values of population size, with other model assumptions as in the monocious case. In general there are $(2N_m + 1)(2N_f + 1)$ possible states of gene frequency in the two sexes if there are N_m males and N_f females in each replicate line every generation. We shall only consider the case where the population sizes and selective values are the same for the two sexes, and we let $N_m = N_f = L$ and a state i specify that there are i_m A_1 alleles in male parents and i_f A_1 alleles in female parents, ($0 \leq i_m, i_f \leq 2L$). With random mating, the zygotic frequencies in the progeny from parents in state i are

$q_m q_f A_1 A_1$, $q_m(1 - q_f) + (1 - q_m)q_f A_1 A_2$, $(1 - q_m)(1 - q_f) A_2 A_2$ where $q_m = i_m/2L$, $q_f = i_f/2L$. The mean gene frequency among the progeny is $q = (q_m + q_f)/2$, and letting $r = (q_m - q_f)/2$, the proportion of heterozygotes becomes $2q(1 - q) + 2r^2$. Of course, the states could be defined in terms of q and r among the zygotes, instead of q_m and q_f in their parents. The mean fitness in terms of q and r is

$$\bar{W} = 1 - (s_1 + s_2) [\bar{q}(1 - \bar{q}) + (q - \bar{q})^2 - r^2] \quad (4)$$

With selection operating independently in males and females, the probability b_{ij} that a line is in state j at generation $t + 1$, given that it was in state i at generation t is

$$b_{ij} = P(j_m | i). P(j_f | i)$$

where $P(j_m | i)$, $P(j_f | i)$ are the marginal probabilities of obtaining $j_m A_1$ alleles in males and $j_f A_1$ alleles in females, respectively. For example,

$$P(j_m | i) = \sum_{\substack{2x+y \\ =j_m}} \binom{L}{xyz} [q_m q_f (1 - s_1)]^x [q_m(1 - q_f) + (1 - q_m)q_f]^y \cdot [(1 - q_m)(1 - q_f)(1 - s_2)]^z / \bar{W}^L \quad (5)$$

and, since the selection coefficients are assumed to be the same in males and females, $P(j_f | i)$ is obtained by substituting j_f for j_m in (5). The expected change in gene frequency, δq , is in one generation

$$\delta q = - \frac{(s_1 + s_2)q(1 - q)(q - \bar{q}) + r^2(s_1 + s_2)(q + \bar{q} - 1)}{\bar{W}} \quad (6)$$

If terms of order $(r^2)^a (s_1 + s_2)^b$ are ignored if $a + b > 2$, \bar{W} in equation (4) may be replaced by the \bar{w} of equation (1) relating to the monocious model.

Some simplification of the matrix \mathbf{B} with elements b_{ij} is possible because of symmetry. It is shown in the APPENDIX that iteration can be performed with a vector of dimension $(2L + 1)(L + 1)$ and a square matrix of the same dimension, rather than with a vector and matrix of dimension $(2L + 1)^2$. Even so, it was necessary to restrict computation to matrices with $L = 5$ giving an effective population size of 10, whereas with the model with only one sex it was possible to work with N as large as 40.

Continuous model approximation: As N becomes infinitely large, but with $N(s_1 + s_2)$ remaining finite, the selection process can be approximated by a continuous model using a diffusion equation (WATERSON 1962; KIMURA 1964). The KOLMOGOROV forward equation has not been solved explicitly, although the dominant latent root has been evaluated (MILLER 1962). However, we can use the form of the equation to make generalisations about our results, since the inbreeding and selection process becomes only a function of $N(s_1 + s_2)$, \bar{q} and the initial frequency, q_0 , so long as time is measured on a scale proportional to N . Tests were made to find the adequacy of this generalisation for small values of N , and results are shown in Table 1. The linear function of fitness, $E[(q - \bar{q})^2]$, is tabulated for various values of $N(s_1 + s_2)$, \bar{q} and t/N generations, for $N = 10$, 20 and 40 in the monocious model. The results obtained with different values of N are seen to be very similar, except with the largest value of $s_1 + s_2$ (0.8). Also,

TABLE 1

$E[(q - \bar{q})^2] \times 10^5$ computed for several values of population size (N) with a monecious model (M), and for 5 males and 5 females ($N = 10$) with a diecious model (D).

$q_0 = \bar{q}$, except for $\bar{q} = 0.0$ when $q_0 = 0.1$.

\bar{q}	Generations $\times 1/N$	$N(s_1 + s_2)$ Method (N)	1				4				16	
			D(10)	M(10)	M(20)	M(40)	D(10)	M(10)	M(20)	M(40)	M(20)	M(40)
0.0	0.5		2654	2665	2638	2625	1924	1809	1803	1800	558	581
	1.0		3488	3472	3450	3440	1687	1495	1541	1563	205	220
	2.0		4068	4001	4019	4025	892	716	789	828	27	30
	4.0		4195	4115	4172	4199	208	141	179	203	0	1
	$\rightarrow \infty$		4166	4095	4178	4218	9	9	25	37	0	0
0.1	0.5		1841	1883	1857	1845	1528	1498	1487	1481	841	865
	1.0		2969	3009	2982	2969	1914	1923	1856	1871	929	950
	2.0		4083	4093	4090	4088	1733	1615	1697	1739	977	984
	4.0		4844	4820	4852	4866	1288	1219	1292	1335	998	999
	$\rightarrow \infty$		5312	5235	5293	5320	1046	1046	1095	1127	1000	1000
0.3	0.5		4213	4372	4336	4319	3203	3326	3399	3429	1348	1560
	1.0		7148	7418	7378	7357	4711	4919	5105	5188	1708	2006
	2.0		11071	11424	11398	11384	6355	6659	6997	7155	2344	2781
	4.0		15163	15465	15470	15471	8083	8434	8903	9131	3451	4083
	$\rightarrow \infty$		18530	18470	18518	18542	10545	10571	11127	11403	9000	9000
0.5	0.5		4980	5193	5160	5143	3626	3831	3964	4020	1162	1461
	1.0		8588	8982	8943	8922	5471	5875	6182	6316	1179	1522
	2.0		13828	14401	14356	14333	8171	8895	9455	9701	1184	1552
	4.0		19796	20339	20301	20282	12366	13463	14272	14622	1195	1602
	$\rightarrow \infty$		25000	25000	25000	25000	25000	25000	25000	25000	25000	25000

a doubling of N from 20 to 40 has less effect than a doubling from 10 to 20, and, of course, the results must converge to the diffusion equation result as N becomes infinite. Thus, in order to describe the situation, it seems satisfactory to use results from just one value of N . Since we are in practice likely to be interested in values of N much larger than we can handle on the computer, we have only analyzed results obtained with $N = 40$, our largest value. Satisfactory agreement between diffusion equation and exact methods has been found in earlier studies by EWENS (1963), who also derived correction terms for approximations (EWENS 1964). However, these results were for a haploid model with additive selective advantages.

Also included in Table 1 are some results obtained with the diecious model, using the same parameters but with $L = 5$, equivalent to $N = 10$. The function plotted is $E[(q_m + q_f)/2 - \bar{q}]^2$ for comparison with the monecious model, but the mean fitness in the diecious model is also affected by departures from Hardy-Weinberg equilibrium due to gene frequency differences between the sexes of the parents. Initially it is assumed that there is the same gene frequency in each sex, so that, for example, with $q_0 = 0.1$, $i_m = i_f = 1$. There is generally adequate

agreement between the one and two-sex models with $N = 10$. At larger values of N better correspondence can be anticipated, since the main effect of having different sexes would appear to be that this causes departure from Hardy-Weinberg equilibrium among the progeny when the parents have different frequencies. However, $E[(q_m - q_f)^2] = 4E(r^2)$ is inversely proportional to N , and so, for large N will become relatively unimportant in the prediction of change of gene frequency (equation 5). FELDMAN (1966) uses the theory of WATTERSON (1962) to show that the same diffusion equation approximates both models. We therefore seem justified in drawing conclusions about populations with random mating between two sexes from our results with populations in which there is only one sex with random mating, including random selfing.

SELECTION FROM INITIAL GENE FREQUENCY EQUILIBRIUM

When a large population is suddenly reduced in size, the initial frequency at those loci in which segregation had been maintained in the population by superior fitness of the heterozygote may be assumed to be close to the equilibrium frequency, \bar{q} . This situation is clearly of importance and we shall consider it first and in some detail.

The effect on mean fitness, which is the character under selection, can be calculated from the average value of $(q - \bar{q})^2$. It proves useful to use as a modified time scale, $1 - e^{-t/2N} = F^*$, which is approximately equal to the inbreeding coefficient measured from pedigrees. In the absence of selection, the heterozygosity declines as $1 - F^*$. The mean of the character under selection is plotted in Figure 1 for a range of $N(s_1 + s_2)$ values for $\bar{q} = 0.1, 0.3$ and 0.5 . For comparison the curve for the expected value of a character controlled by recessives ($\bar{q} = 0$) is also included, with initial recessive frequency 0.05 . Figures 2 and 3 show the average heterozygosity and the average gene frequency, respectively, during the inbreeding and selection process. Results for $\bar{q} = 0.5$ are not included in Figures 2 and 3 since for $\bar{q} = q_0 = 0.5$, there can be no change in the mean gene frequency when starting from equilibrium, and thus $E(q) = 0.5$ and $E[2q(1 - q)] = 0.5 - 2E[(q - \bar{q})^2]$ for all t . Some loci may show heterozygote superiority for fitness and yet have additive effects on some observed metric trait. Changes in the mean of this character will therefore be a linear function of the mean gene frequency, $E(q)$, shown in Figure 3.

When the inbreeding is so rapid that selection has very little effect (as might happen for instance by using special crossing programmes in *Drosophila*) the mean fitness declines linearly with F^* in all cases ($N(s_1 + s_2) = 0$). Some aspects of the results are rather surprising, when it is borne in mind that heterozygote superiority has its greatest effect in maintaining segregation in small populations when the equilibrium frequency is 0.5 . However it can be seen that at low values of F^* and $N(s_1 + s_2)$, selection has a greater effect when the equilibrium frequency is 0.1 . Figures 2 and 3 show that we are here dealing with two quite separate phenomena which may act in opposite directions in particular cases.

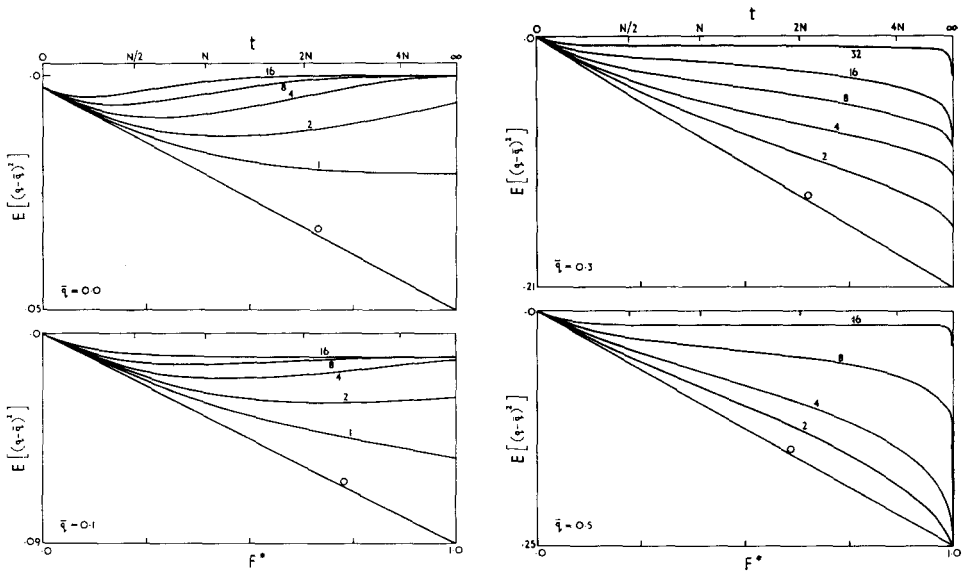


FIGURE 1.—The effect on the mean value of the selected trait $E[(q - \bar{q})^2]$ is plotted against $F^* = 1 - e^{-t/2N}$ for values of \bar{q} of 0.0, 0.1, 0.3 and 0.5. The initial frequency equals \bar{q} , except for $\bar{q} = 0.0$ (recessive) when the initial frequency is 0.05. Curves are plotted for several values of $N(s_1 + s_2)$.

These correspond to the last two terms in equation (3) which represent respectively the change in gene frequency and in heterozygosity. When the equilibrium gene frequency equals 0.5, the mean gene frequency does not change, so that changes in the mean of the selected character and in the level of heterozygosity are proportional to each other. Inbreeding decline is reduced by the maintenance of a high level of heterozygosity since the second term in (1) vanishes. On the other hand, when $\bar{q} = 0.1$, we find that selection now *reduces* the heterozygosity but at the same time reduces the effect on the mean of the character by preventing fixation of the poorer homozygote. Thus the effect on the mean is due to completely different phenomena in the two cases.

Considering separately the curves for the different equilibrium gene frequencies as \bar{q} approaches zero (including the recessive case as the most extreme value), the inbreeding decline may be halted after a certain time and the selected character then rises again. When there is heterozygote superiority and the initial gene frequency equals \bar{q} , the final mean can never be as high as that at the outset since at complete fixation ($F^* = 1$) we cannot do better than fix all populations for the better homozygote, which is inferior to the mean in the initial population at equilibrium. On the other hand, with a single recessive gene in which segregation is maintained by mutation, the final mean may be above the initial value, due to complete exclusion of the recessive at larger values of Ns_1 . It is known that

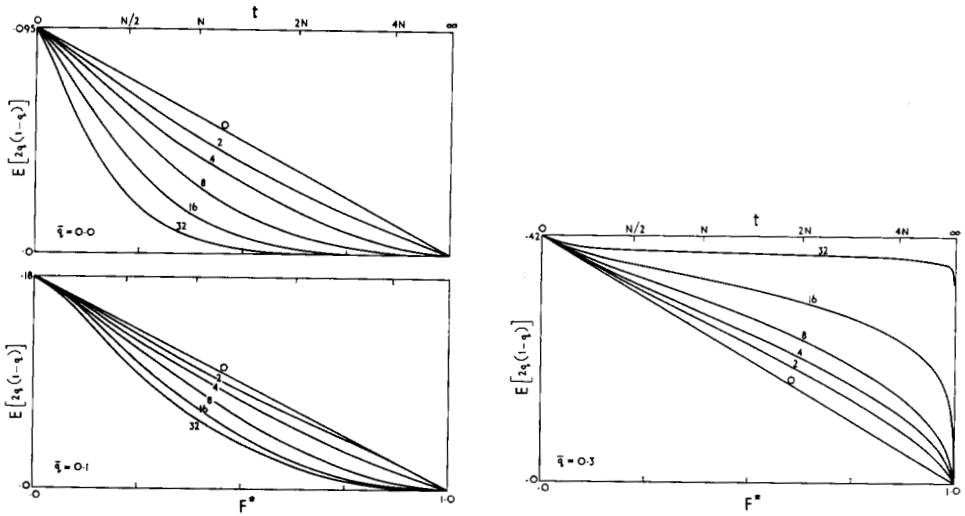


FIGURE 2.—The mean heterozygosity, $E[2q(1-q)]$, plotted as for Figure 1, with $\bar{q} = 0.5$ excluded.

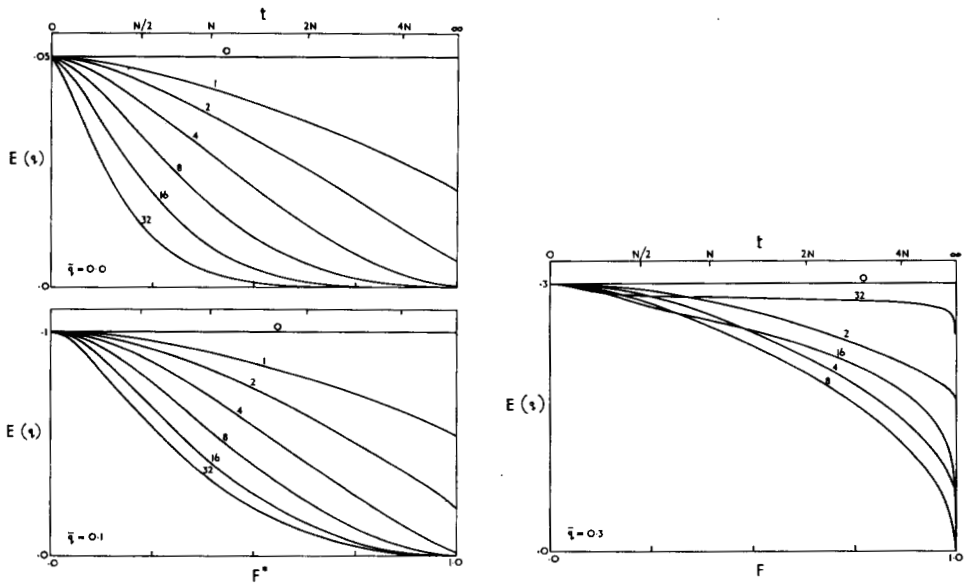


FIGURE 3.—The mean gene frequency, $E(q)$, plotted as for Figure 1, with $\bar{q} = 0.5$ excluded.

in the absence of selection the additive genetic variance within populations due to an initially rare recessive gene will increase up to inbreeding coefficients of 0.5 (ROBERTSON 1952). This provides an explanation of the curves for the extreme values of \bar{q} . The first effect of the reduction in population size is a decline in the mean due to an increase in the proportion of homozygotes and selection does not become effective until the additive genetic variance has increased as a result of the spread of gene frequencies. The inbreeding decline is then halted and the mean increased.

When the equilibrium gene frequency is 0.5 and $N(s_1 + s_2)$ is high the mean of the character under selection becomes almost constant after a few generations. There is thereafter a very slow approach to fixation. Selection for the heterozygote under these conditions retards rather than prevents fixation and ultimately all replicate lines will become fixed. As F^* approaches unity the time scale in generations is very much contracted and very small changes in F^* lead to relatively large changes in $E[(q - \bar{q})^2]$.

The curves for $\bar{q} = 0.3$ have a pattern in between those for the other values. In the earlier generations they are similar to those for $\bar{q} = 0.5$, but the effect of selection can be seen from the gene frequency at fixation which is much higher when $N(s_1 + s_2)$ is large and only the better homozygote is fixed. The mean gene frequency (Figure 3) for $\bar{q} = 0.3$ declines more rapidly for $N(s_1 + s_2) = 8$ than for $N(s_1 + s_2) = 32$ since fixation occurs earlier, but the limiting value of $E(q)$ is almost the same in each case.

INITIAL GENE FREQUENCY NOT AT THE EQUILIBRIUM VALUE

There are several situations in which the initial gene frequency may not be at equilibrium. In natural populations there may be a change in environment, which alters the relative fitness of the genotypes, coinciding with a reduction in population size, or there may be departures from equilibrium resulting from random drift at previous reductions in population size. In populations of plants and domestic animals, artificial selection may be applied to a trait which had not previously been important. The selective values (s_1, s_2) are then approximately equal to linear functions of the average genotypic values for the quantitative trait. The adequacy of this approximation for study of truncation selection in infinite populations has been investigated by HILL (1969) and found suitable for most descriptive purposes. We shall illustrate the effects of departures from initial equilibrium for only one value of $N(s_1 + s_2)$, from which we can readily infer the results in other situations.

In Figures 4 and 5 the mean of the quantitative trait, expressed as $[(q - \bar{q})^2]$, and the average gene frequency, respectively, are plotted as a function of initial frequency and generations of inbreeding for $N(s_1 + s_2) = 8$ and $\bar{q} = 0.3$ and 0.5. The curves are drawn for values of t such as 0, $N/2$, N , $2N$, $8N$ and ∞ generations, corresponding to $F^* = 0, 0.22, 0.39, 0.63, 0.98$ and 1, respectively.

When $\bar{q} = 0.5$ and $q_0 < 0.5$, there is an initial period of advance in the mean

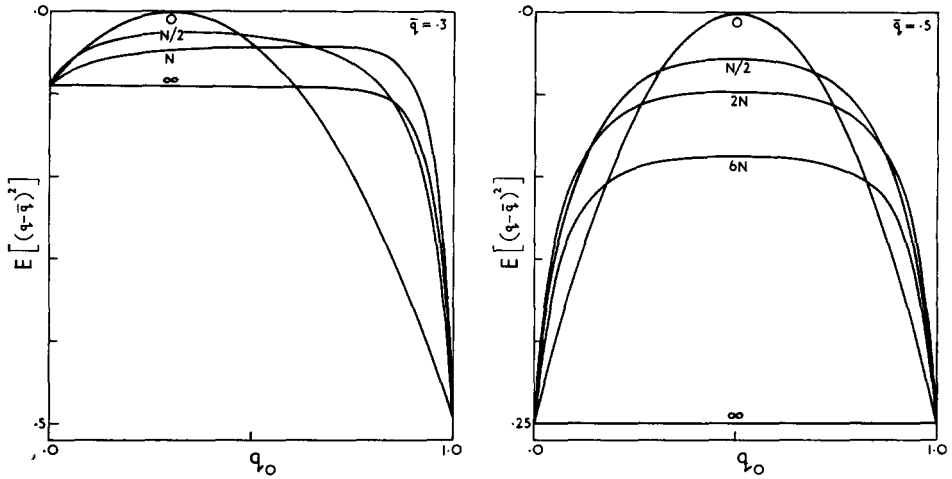


FIGURE 4.— $E[(q-\bar{q})^2]$ for the selected trait is shown as a function of initial frequency for $\bar{q} = 0.3$ and 0.5 and $N(s_1 + s_2) = 8$. Curves are plotted for different numbers of generations.

of the selected trait (Figure 4). This is later lost, and the final mean is the same for all values of q_0 since both homozygous genotypes have the same value. By contrast, when the equilibrium frequency differs from 0.5, the final mean depends on the relative proportions of the two homozygotes fixed, and therefore on q_0 . When \bar{q} is less than 0.5 and q_0 high, most of the early advance is retained and there is an overall advance in the mean if $q_0 > 2\bar{q}$, approximately.

If the equilibrium frequency is 0.5, selection always changes the average gene frequency (Figure 5) towards 0.5 until the steady state is reached. After this the average gene frequency remains constant, because the distribution of unfixed

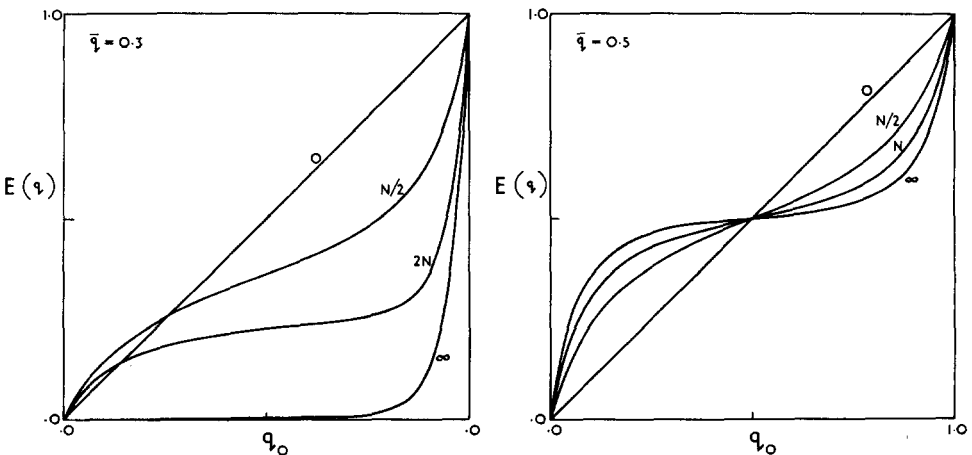


FIGURE 5.—As Figure 4, but a plot of the mean gene frequency, $E(q)$.

classes is symmetric and fixation takes place at the same rate for each homozygote. The mean of the selected trait must then slowly decline. However, if $\bar{q} < 0.5$ and $q_0 > \bar{q}$, selection reduces the gene frequency throughout the inbreeding and selection process. If $q_0 < \bar{q} < 0.5$, there is an initial increase in the average gene frequency as selection towards the equilibrium frequency occurs. The gene frequency distribution of unfixed classes is asymmetric and the poorer homozygote is rarely fixed so that the average gene frequency then declines. We have the interesting phenomenon of unidirectional selection in which there is a reversal of the direction of gene frequency change during selection.

The average gene frequency within segregating lines: Only very large populations can remain at the equilibrium frequency for long periods of time. If we wish to estimate the equilibrium frequency at a locus, as an indirect way of measuring the relative fitnesses of the homozygotes, we may have to use information from populations of finite size. The observed frequency within such populations might be thought to be a good estimator of the equilibrium frequency providing there has been no recent change in environment or immigration. However, the following discussion will show that the observed frequency in small populations is a biased estimator of the equilibrium frequency in large populations.

Consider an infinitely large population in equilibrium for a locus with heterozygote advantage, from which many identical sub-lines are drawn. The average gene frequency for all lines can be predicted from Figure 3, but this combines data from two types of populations: those which are already fixed, in which the gene frequency is 0 or 1, and those still segregating. We are concerned here with the average frequency within these *segregating* populations, which will reach a steady state value denoted by \bar{q} .

In Figure 6 the relation between \bar{q} and \bar{q} is plotted for a range of $N(s_1 + s_2)$ values, where results were obtained using the transition probability matrix method described earlier. When $N(s_1 + s_2)$ is infinitely large, the rate of fixation will be very low and \bar{q} will equal \bar{q} . On the other hand, when $N(s_1 + s_2)$ approaches zero, the unfixed classes have a uniform distribution and $\bar{q} = 0.5$. When $\bar{q} = 0.5$, the distribution of unfixed classes is symmetric about 0.5 for all values of $N(s_1 + s_2)$ so that $\bar{q} = \bar{q} = 0.5$. When $\bar{q} \neq 0.5$, we know that at final fixation the gene frequency is changed towards that of the better homozygote. If $\bar{q} = 0.3$, for instance, the first effect of the small population size is to spread the gene frequencies about this value. But only those lines with very low frequencies are likely to be fixed so that those left segregating will have a mean gene frequency greater than 0.3. We see from the figure that \bar{q} almost always lies between \bar{q} and 0.5. With intermediate $\bar{q} (\neq 0.5)$ and $N(s_1 + s_2)$ very large (≥ 16), \bar{q} may in some cases lie just outside this range. Here fixation occurs very slowly and the effect can be attributed to the asymmetry of the effect of selection. If the gene frequency drifts from $\bar{q} = 0.3$, say, it is selected more rapidly back when the drift is towards one-half than when it is towards zero because of the term $q(1 - q)$ in

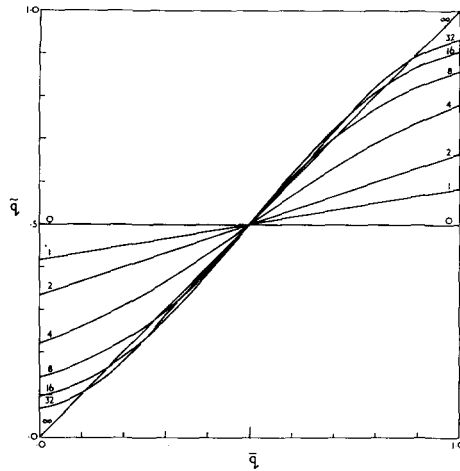


FIGURE 6.—The relations between the average frequency within segregating lines at the steady state, \bar{q} , and the equilibrium frequency for large populations, q . Curves are plotted for several values of $N(s_1 + s_2)$.

$\delta q = -(s_1 + s_2) q(1 - q)(q - \bar{q})/\bar{w}$ so that there is a relative excess of populations with extreme gene frequencies.

For most combinations of effective population size and selective values, the average gene frequency within segregating lines is seen to be biased towards 0.5. Thus if we search for polymorphism within a single small population, we are not likely to find gene frequencies at extreme values. We are then not entitled to infer the relative selective advantages at the loci we observe.

DISCUSSION

A wide variety of consequences of inbreeding is possible when there are loci with heterozygote advantage. Perhaps the most interesting result is that inbreeding depression may be delayed for quite different reasons, depending on the equilibrium frequency, when the population is initially at equilibrium. If the population is not initially at equilibrium, the mean of the selected trait may rise initially and then fall as inbreeding progresses, as well as the reverse.

Two processes were found to reduce inbreeding decline from loci with heterozygote advantage. When the equilibrium frequency was near 0.5, this was due to the maintenance of heterozygosity whereas at extreme equilibrium frequencies it was caused by preferential fixation of the better homozygote. It might be possible to differentiate between these situations in two ways. In the first, lines which had been inbred slowly up to, say, $F^* = 0.75$ (calculated from pedigrees) could then be inbred very rapidly, perhaps by full sibbing. With an equilibrium frequency of one-half, a rapid decline in fitness would be expected to accompany fixation. However, with extreme equilibrium values, most loci will already be

fixed, so little further decline in the mean would be expected. The second method of differentiation has been mentioned in a different context by ROBERTSON (1962), and would apply to very highly inbred replicate lines from the same initial population. Crosses between these lines should show heterosis for loci with intermediate equilibrium frequencies, since both types of homozygote will be fixed in different lines, but for extreme values of the equilibrium frequency most lines will be fixed for the same allele, and no heterosis will be found.

SUMMARY

A theoretical study has been made of the process of inbreeding at loci with heterozygote superiority. Results were obtained using transition probability matrices for monocious and diecious random mating sub-populations, and these alternative models were compared numerically. It was found that, by a suitable choice of parameters, general conclusions drawn from one population size with a monocious model could be applied to other values of population size and to the diecious model.—The rate of inbreeding depression at these loci can be much reduced by selection, but selection is found to act in different ways, depending on the equilibrium frequency in large populations. If this is close to one-half, the effect is due to the maintenance of heterozygosity. With extreme values of the equilibrium frequency it is due to increased fixation of the better homozygote, and this may cause an increase in the mean after a depression during the initial generations of inbreeding.—The relationship between average gene frequency within segregating populations at the steady state and the equilibrium frequency is investigated. This average frequency usually lies between the equilibrium frequency and one-half, giving the impression of more nearly equal selective values for the two alternative homozygotes than is really the case.

LITERATURE CITED

- EWENS, W. J., 1963 Numerical results and diffusion approximations in a genetic process. *Biometrika* **50**: 241–249. — 1964 The pseudo-transient distribution and its uses in genetics. *J. Appl. Prob.* **1**: 141–156.
- FELDMAN, M. W., 1966 On the offspring number distribution in a genetic population. *J. Appl. Prob.* **3**: 129–141.
- HILL, W. G., 1969 On the theory of artificial selection in finite populations. *Genet. Res.* (in press).
- JAIN, S. K., and R. W. ALLARD, 1966 The effects of linkage, epistasis, and inbreeding on population changes under selection. *Genetics* **53**: 633–659.
- JAIN, S. K., and P. L. WORKMAN, 1967 Generalised F-statistics and the theory of inbreeding and selection. *Nature* **214**: 674–678.
- KIMURA, M., 1964 Diffusion models in population genetics. *J. Appl. Prob.* **1**: 177–232.
- MILLER, G. F., 1962 The evaluation of eigenvalues of a differential equation arising in a problem in genetics. *Proc. Camb. Phil. Soc.* **58**: 588–593.
- REEVE, E. C. R., 1955 Inbreeding with homozygotes at a disadvantage. *Ann. Human Genet.* **21**: 277–288.

ROBERTSON, A., 1952 The effect of inbreeding on the variation due to recessive genes. *Genetics* **37**: 189-207. — 1962 Selection for heterozygotes in small populations. *Genetics* **47**: 1291-1300.

WATTERSON, G. A., 1962 Some theoretical aspects of diffusion theory in population genetics. *Ann. Math. Statist.* **33**: 939-957.

APPENDIX

Reduction of the transition matrix B for the dieocious model

Order the $(2L + 1)^2$ states of **B** into 3 groups:

Group (1): $2L + 1$ states with $i_m = i_f$

Group (2): $\binom{2L + 1}{2} = 2L^2 + L$ states with $i_m < i_f$ ordered, for example, as

$$(i_m, i_f) = (0,1), (0,2), \dots, (2L - 1, 2L)$$

Group (3): $2L^2 + L$ states with $i_m > i_f$ ordered similarly to group (2) as

$$(i_m, i_f) = (1,0), (2,0), \dots, (2L, 2L - 1).$$

Since q_m and q_f and thus i_m and i_f can be interchanged in equation (5) and (6), **B** may be partitioned as follows:

$$\mathbf{B} = \begin{pmatrix} \mathbf{C} & \mathbf{D} & \mathbf{D} \\ \mathbf{E} & \mathbf{G} & \mathbf{G} \\ \mathbf{E} & \mathbf{G} & \mathbf{G} \end{pmatrix}$$

where, for example, **C** specifies transitions from states in group (1) to other states in group (1) and is square of dimensions $2L + 1$. In order to compute expectations of functions such as the mean gene frequency, $v_i(t) = E[(i_m + i_f)/2L \mid \text{initial state} = i]$, which are symmetric in i_m and i_f , we partition the vector

$$\mathbf{v}'(t) = (\mathbf{x}'(t), \mathbf{y}'(t), \mathbf{y}'(t))$$

where for example $\mathbf{x}'(t)$ relates to states of group (1) and has dimension $2L + 1$. It then follows that

$$\begin{pmatrix} \mathbf{x}(t) \\ \mathbf{y}(t) \end{pmatrix} = \begin{pmatrix} \mathbf{C} & 2\mathbf{D} \\ \mathbf{E} & 2\mathbf{G} \end{pmatrix} \begin{pmatrix} \mathbf{x}(t-1) \\ \mathbf{y}(t-1) \end{pmatrix}$$

and iteration can be performed with the reduced vector and square matrix of dimension $(2L + 1)(L + 1)$.