

## The genetic variability of polygenic characters under optimizing selection, mutation and drift

By M. G. BULMER

*Department of Biomathematics, University of Oxford*

(Received 10 May 1971)

### SUMMARY

The effect of optimizing selection, mutation and drift on a metric character determined by a large number of loci with equal effects without dominance was investigated theoretically. Conditions for a stable equilibrium under selection and mutation, in the absence of drift, have been obtained, and hence the amount of genetic variability which can be maintained by mutation has been determined. An approximate expression for the average amount of genetic variability to be expected in the presence of drift in a population of finite size has also been obtained and evaluated.

### 1. INTRODUCTION

The effect of selection for an optimal value on a polygenic, metric character has been investigated by Fisher (1930), Haldane (1932, 1953), Wright (1935), Robertson (1956) and Bulmer (1971*a*). The conclusion of these authors is that optimizing selection tends to eliminate genetic variability, but the tendency to fixation due to this cause and to genetic drift will at some point be balanced by the introduction of fresh variability by mutation. The purpose of this paper is to determine the amount of genetic variability which can be maintained by mutation under optimizing selection both in an infinite and in a finite population.

Because of the complexity of the problem we shall only discuss the simplest case of a metric character determined by a large number of loci with equal effects and without dominance; the effect of linkage disequilibrium will also be ignored. Consider, then, a metric character,  $y$ , whose genetic component is determined by  $n$  loci. We shall suppose that each locus has two alleles,  $C_1$  and  $C_2$ , and that the effects of the three possible genotypes,  $C_1C_1$ ,  $C_1C_2$  and  $C_2C_2$ , are  $-a$ ,  $0$  and  $a$  respectively. If the frequency of the  $C_1$  allele at the  $i$ th locus is  $p_i$ , then the genetic variance (before selection) is

$$V_G = 2a^2 \sum_{i=1}^n p_i q_i. \quad (1)$$

We shall also suppose that there is an independent, normally distributed environmental component of  $y$  with variance  $V_E$ , so that the total phenotypic variance (before selection) is  $V = V_G + V_E$ . The purpose of this paper is to determine the expected value of the genetic variance under the joint action of selection, mutation and drift.

We shall also suppose that the population is subject to selection for an optimal

value,  $\theta$ . As a mathematical model of optimizing selection we shall suppose that the fitness of an individual with phenotypic value  $y$  is

$$w(y) = \exp[-c(y - \theta)^2], \quad (2)$$

where  $c$  is a measure of the intensity of selection. If  $y$  is normally distributed in the population before selection with mean  $M$  and variance  $V$ , then it is quite easy to show (see Latter, 1970, and Bulmer, 1971*b*) that it will also be normally distributed after selection with mean  $M + DM$  and variance  $V + DV$ , where

$$\left. \begin{aligned} DM &= 2cV(\theta - M)/(1 + 2cV), \\ DV &= -2cV^2/(1 + 2cV). \end{aligned} \right\} \quad (3)$$

It is often more convenient to express the intensity of selection in terms of the dimensionless quantity

$$k = 2cV/(1 + 2cV), \quad (4)$$

so that

$$DV = -kV. \quad (5)$$

The quantity  $k$  therefore represents the proportionate reduction in the variance as a result of selection; it has been called by Latter (1970) 'the coefficient of centripetal selection'.

The above model of optimizing selection is a form of stabilizing selection since  $DV$  is negative, that is to say the effect of selection is to reduce the phenotypic variance. Under this form of selection, and in the absence of overdominance, it has already been shown (Bulmer, 1971*a*) that there is no non-trivial stable equilibrium under selection alone, in the absence of mutation, so that all genetic variability must eventually be eliminated. In the next section we shall consider how much genetic variability can be maintained in this situation by mutation; for the time being we shall suppose that the population size is effectively infinite so that there is no genetic drift. For the sake of simplicity we shall only consider the symmetrical case in which  $M = \theta$  when all the gene frequencies are equal to one half, or more generally when the average of the gene frequencies is equal to one half - that is to say when  $\Sigma p_i/n = \frac{1}{2}$ .

## 2. SELECTION AND MUTATION

It has been shown previously (Bulmer, 1971*a*) that the change in the gene frequency at the  $i$ th locus as a result of selection is given approximately by

$$\Delta p_i(\text{selection}) = p_i q_i [Aa - \frac{1}{2}Ba^2(p_i - q_i)]. \quad (6)$$

If all loci are subject to equal forward and backward mutation rates,  $u$ , then the total change in the gene frequency is approximately

$$\Delta p_i(\text{total}) = p_i q_i [Aa - \frac{1}{2}Ba^2(p_i - q_i)] - u(p_i - q_i). \quad (7)$$

If  $y$  is normally distributed the coefficients  $A$  and  $B$  are given by

$$\left. \begin{aligned} A &= -DM/V, \\ B &= (DV + DM^2)/V^2, \end{aligned} \right\} \quad (8)$$

where  $DM$  and  $DV$ , the change in the phenotypic mean and variance as a result of selection, are given by equation (3). The above expression for  $A$  is only sufficiently

accurate if the skewness of the genetic contribution to  $y$  is of order  $1/n$ . However, it is clear from the symmetry of the situation that for any equilibrium position the gene frequencies must be symmetrically distributed about  $\frac{1}{2}$ , so that the skewness of the distribution is zero; furthermore,  $DM$  must be zero at equilibrium since the regression of offspring on parent is linear in the absence of dominance. At equilibrium we may therefore write

$$\left. \begin{aligned} A &= 0, \\ B &= DV/V^2 = -2c/(1+2cV). \end{aligned} \right\} \tag{9}$$

Since the total change in the gene frequency must be zero at equilibrium we may therefore write

$$P_i Q_i \frac{ca^2}{(1+2cV)} (P_i - Q_i) - u(P_i - Q_i) = 0, \tag{10}$$

if  $P_i$  denotes the gene frequency at the  $i$ th locus at an equilibrium position. It follows that either

$$\left. \begin{aligned} P_i &= Q_i = \frac{1}{2} \\ P_i Q_i &= u(1+2cV)/ca^2 = \delta \end{aligned} \right\} \tag{11}$$

or

at an equilibrium position. The second alternative is only possible if  $\delta < \frac{1}{4}$ ; if this is the case, the equation  $P_i Q_i = \delta$  will have two solutions between 0 and 1, which will be denoted by  $\pi_1$  and  $\pi_2$ , located at equal distances on either side of  $\frac{1}{2}$ . Since the gene frequencies must be symmetrically distributed about  $\frac{1}{2}$ , there must be equal numbers of loci at  $\pi_1$  and  $\pi_2$ .

To test the stability of an equilibrium position, let us suppose that the gene frequencies are perturbed from their positions by small perturbations,  $e_i$ , so that  $p_i = P_i + e_i$ . The coefficient  $A$  will be changed by the perturbation from zero to  $2aB \sum e_i$  (see Bulmer, 1971 a); the change in the coefficient  $B$  can be ignored. Hence, ignoring terms like  $e_i^2$ , we may write

$$\left. \begin{aligned} \Delta p_i &= \frac{1}{2} a^2 B \sum_{j=1}^n e_j - \frac{1}{4} a^2 B e_i - 2u e_i \\ &= -\frac{u}{\delta} \sum_{j=1}^n e_j + \frac{1}{2} \frac{u}{\delta} e_i - 2u e_i \quad \text{if } P_i = Q_i = \frac{1}{2}, \\ \Delta p_i &= 2a^2 B P_i Q_i \sum_{j=1}^n e_j + \frac{1}{2} a^2 B (P_i - Q_i)^2 e_i \\ &= -4u \sum_{j=1}^n e_j + 4u e_i - (u/\delta) e_i \quad \text{if } P_i Q_i = \delta. \end{aligned} \right\} \tag{12}$$

If we write  $e_i^* = e_i + \Delta p_i$  for the perturbation in the next generation, then

$$\left. \begin{aligned} e_i^* &= \left( 1 + \frac{1}{2} \frac{u}{\delta} - 2u \right) e_i - \frac{u}{\delta} \sum_{j=1}^n e_j \quad \text{if } P_i = Q_i = \frac{1}{2}, \\ e_i^* &= \left( 1 - \frac{u}{\delta} + 4u \right) e_i - 4u \sum_{j=1}^n e_j \quad \text{if } P_i Q_i = \delta. \end{aligned} \right\} \tag{13}$$

Let us first suppose that  $\delta < \frac{1}{4}$ . In this case the equilibrium will be unstable if there is more than one locus at a gene frequency of  $\frac{1}{2}$ . For if the  $i$ th and  $j$ th loci have gene frequencies of  $\frac{1}{2}$ , then

$$(e_i^* - e_j^*) = (1 + \frac{1}{2}(u/\delta) - 2u) (e_i - e_j), \tag{14}$$

which is unstable since  $\frac{1}{2}u/\delta > 2u$  when  $\delta < \frac{1}{4}$ . But the number of loci at  $\frac{1}{2}$  must be even or odd according as the total number of loci is even or odd, since there are equal numbers of loci at  $\pi_1$  and  $\pi_2$ . Thus a necessary condition for a stable equilibrium is that there be no loci or one locus at  $\frac{1}{2}$ , depending on whether the total number of loci is even or odd. To show that this is also a sufficient condition for stability when the total number of loci is even, suppose that there are equal numbers of loci at  $\pi_1$  and  $\pi_2$ . If we consider the quantities  $\bar{e} = \Sigma e_i/n$  and  $d_i = e_i - \bar{e}$ , then

$$\left. \begin{aligned} \bar{e}^* &= [1 - (u/\delta) - 4u(n-1)]\bar{e}, \\ d_i^* &= (1 - (u/\delta) + 4u)d_i. \end{aligned} \right\} \tag{14}$$

The quantity  $d_i$  is stable since  $u/\delta > 4u$  when  $\delta < \frac{1}{4}$ ; the quantity  $\bar{e}$  is clearly also stable. It follows that the  $e_i$ 's are also stable. On the other hand, if  $\delta > \frac{1}{4}$ , then the only equilibrium position is with all the loci at  $\frac{1}{2}$ ; it can easily be verified by the above argument that this is a stable equilibrium.

A slight complication arises from the fact that the critical quantity

$$\delta = u(1 + 2cV)/ca^2$$

itself depends on the gene frequencies which determine the genetic variance and hence affect the total variance,  $V$ . If we assume that  $P_iQ_i$  has the same value  $x$  at all loci (which must be the case for a stable equilibrium if the number of loci is even), then

$$\delta = \frac{1 + 2cV_E}{ca^2} + 4nux. \tag{15}$$

This straight line will intersect the straight line  $\delta = x$  at a single point, when  $x = \xi$ , say. If  $\xi < \frac{1}{4}$ , then  $\delta$  must clearly be less than  $\frac{1}{4}$  when  $x = \frac{1}{4}$ , so that there is a unique stable equilibrium at  $x = \xi$ . If  $\xi > \frac{1}{4}$ , then  $\delta$  must be greater than  $\frac{1}{4}$  when  $x = \frac{1}{4}$ , so that there is a unique stable equilibrium at  $x = \frac{1}{4}$ .

To interpret the above results in terms of the genetic variability maintained by mutation, we observe that if  $\delta < \frac{1}{4}$  at the stable equilibrium point, then the genetic variance is

$$V_G = 2a^2 \Sigma P_iQ_i = 2nu(1 + 2cV)/c, \tag{16}$$

so that the heritability is

$$h^2 = V_G/V = 2nu(1 + 2cV)/cV = 4nu/k, \tag{17}$$

where  $k$  is the measure of the intensity of selection defined in equation (4). (This agrees with the result obtained by Latter (1960) by a less rigorous argument, with the proviso that the measure of the intensity of selection,  $I$ , used by Latter is equal to  $\frac{1}{2}k$ .) On the other hand if  $\delta > \frac{1}{4}$  at the stable equilibrium point, then  $P_iQ_i = \frac{1}{4}$  at all loci, so that

$$V_G = \frac{1}{2}na^2 \tag{18}$$

and

$$h^2 = \frac{1}{2}na^2/V = \frac{1}{2}na^2/(\frac{1}{2}na^2 + V_E). \tag{19}$$

Thus the heritability maintained by mutation is the smaller of the two quantities (17) and (19); it would of course be impossible to maintain a heritability higher

than (19) since the genetic variance has a maximum value given by (18). Thus if the observed heritability of a metric character is  $h^2$ , and if it is postulated that this heritability is maintained by mutation, then it must be inferred that  $4nu/k \geq h^2$ . For example, if  $h^2 = \frac{1}{2}$  and  $k = 0.01$ , and if it can be assumed that the mutation rate,  $u$ , is about  $10^{-5}$ , then it is necessary to postulate that the number of loci is at least 125 unless some other means of maintaining genetic variability is operating.

It has been assumed in the above account that there are only two alleles at each locus. This assumption may be rather unrealistic in the light of modern understanding of the nature of the gene, and it is therefore of interest to compare the above result with the model investigated by Kimura (1965), in which there are an effectively infinite number of alleles at each locus. Under this model it is assumed that there is a probability  $u$  that a mutation will occur at a particular locus, and that the change in the quantitative character produced by any mutation can be regarded as a random variable,  $\zeta$ , such that  $E(\zeta^2) = \sigma^2$ . Kimura also supposes that the character is subject to selection depending on the squared deviation from an optimal value, so that the fitness of an individual with phenotypic value  $y$  is  $w(y) = 1 - c(y - \theta)^2$ ; for weak selective pressures this is almost equivalent to the fitness function defined in equation (2). The genetic variance maintained by mutation, on the assumptions that there is no dominance and that the mutation rates,  $u$ , and genetic effects,  $\sigma^2$ , are the same at all loci, is then found to be

$$V_G = n\sigma\sqrt{(2u/c)}. \quad (20)$$

It also has been assumed that all loci have equal effects. If this assumption is relaxed under a model with two alleles at each locus, it can be shown that at equilibrium the loci with large effects will have gene frequencies near 0 or 1, while loci with small effects will have gene frequencies near or equal to  $\frac{1}{2}$ . The reader is referred to the papers of Latter (1960, 1969) for further discussion.

### 3. SELECTION, MUTATION AND DRIFT

If the population is finite, of effective size  $N$ , the gene frequencies,  $p_i$ , at the various loci will be subject to drift and will be random variables with a joint probability distribution. It has been shown by Wright (1937) and more rigorously by Kimura (1964) that the stationary distribution of gene frequencies is given by

$$\phi(p_1, p_2, \dots, p_n) \propto \bar{w}^{2N} \prod_{i=1}^n (p_i q_i)^{4Nu-1}, \quad (21)$$

where  $\bar{w}$  is the average fitness in a population with the given gene frequencies. If  $y$  is normally distributed with mean  $M$  and variance  $V$ , given the gene frequencies  $p_1, p_2, \dots, p_n$ , then, under the model of selection considered previously,

$$\begin{aligned} \bar{w} &= E\{\exp[-c(y - \theta)^2]\} = \frac{1}{\sqrt{(2\pi V)}} \int_{-\infty}^{\infty} \exp[-c(y - \theta)^2 - \frac{1}{2}(y - M)^2/V] dy \\ &= \exp[-c(M - \theta)^2/(1 + 2cV)]/\sqrt{(1 + 2cV)} \quad (\text{by completing the square}) \\ &= \exp\left[-\frac{c(M - \theta)^2}{1 + 2cV} - \frac{1}{2} \ln(1 + 2cV)\right]. \end{aligned} \quad (22)$$

If  $cV$  is small, this can be written to sufficient accuracy as

$$\bar{w} \simeq \exp [-c(M - \theta)^2 - cV] = \exp -c[(M - \theta)^2 + V]. \tag{23}$$

In the symmetrical case when  $M = \theta$  when  $\Sigma p_i/n = \frac{1}{2}$ , we may write

$$\left. \begin{aligned} M - \theta &= -2a\Sigma(p_i - \frac{1}{2}), \\ V &= V_E + 2a^2\Sigma p_i q_i. \end{aligned} \right\} \tag{24}$$

We may thus write approximately

$$\phi(p_1, p_2, \dots, p_n) \propto \exp -2Nca^2\{4[\Sigma(p_i - \frac{1}{2})]^2 + 2\Sigma p_i q_i\} \prod_{i=1}^n (p_i q_i)^{4Nu-1}. \tag{25}$$

The gene frequencies are not independently distributed in equation (25) because of the factor  $[\Sigma(p_i - \frac{1}{2})]^2$ . It is, however, the distribution which would be obtained if we start with independently distributed gene frequencies each having the distribution

$$f(p_i) \propto \exp (-4Nca^2 p_i q_i) (p_i q_i)^{4Nu-1} \tag{26}$$

and then weight this distribution with a probability depending on the value of  $\Sigma(p_i - \frac{1}{2})$ . (The distribution (26) is that of a locus with the homozygotes having the same fitness, and the heterozygotes having a selective disadvantage of  $ca^2$ .) Let us now consider the joint distribution of the random variables  $X = \Sigma(p_i - \frac{1}{2})$  and  $Y = \Sigma p_i q_i$ , on the assumption that the  $p_i$ 's are all independently distributed according to the distribution (26). Since this distribution is symmetrical about  $\frac{1}{2}$ , the covariance of  $p_i$  and  $p_i q_i$  must be zero, since knowledge that  $p_i q_i$  takes the particular value,  $b$ , say, only tells us that  $p_i$  is equally likely to take the values  $\frac{1}{2} + \sqrt{(\frac{1}{4} - b)}$  and  $\frac{1}{2} - \sqrt{(\frac{1}{4} - b)}$ . Hence  $X$  and  $Y$  must also be uncorrelated. Furthermore, it follows from the bivariate form of the central limit theorem that  $X$  and  $Y$  must asymptotically follow a bivariate normal distribution when the number of loci is large; since  $X$  and  $Y$  are uncorrelated they must thus be asymptotically independent. If we now weight the joint distribution of the  $p_i$ 's by a factor depending only on  $X$  in order to produce the distribution (25), the distribution of  $Y$  must remain unaltered. Hence, in order to obtain information about the distribution of  $Y$ , which determines the genetic variance, we can assume that the gene frequencies are independently distributed with the distribution (26).

In order to find the expected value of  $p_i q_i$ , and hence of  $\Sigma p_i q_i$  which determines the genetic variance, from the distribution (26) we must evaluate an integral of the type

$$\begin{aligned} I(\alpha, \beta) &= \int_0^1 e^{-\alpha x(1-x)} [x(1-x)]^{\beta-1} dx \\ &= \int_0^{\frac{1}{2}} e^{-\alpha z} z^{\beta-1} (\frac{1}{4} - z)^{-\frac{1}{2}} dz, \end{aligned} \tag{27}$$

where  $z = x(1-x)$ . This integral is fortunately a standard Laplace transform given in Erdélyi (1954, vol. 1, p. 139, no. 23) and is expressible as

$$I(\alpha, \beta) = B(\beta, \frac{1}{2}) (\frac{1}{4})^{\beta-\frac{1}{2}} e^{-\frac{1}{4}\alpha} M(\frac{1}{2}, \beta + \frac{1}{2}, \frac{1}{4}\alpha) \tag{28}$$

where  $B(\dots, \dots)$  is the complete Beta function and  $M(\dots, \dots, \dots)$  is the confluent hypergeometric function. (I am grateful to Dr F. H. C. Marriott for drawing my attention to this solution of equation (27).) Hence

$$E(p_i q_i) = I(\alpha, \beta + 1) / I(\alpha, \beta) = \frac{1}{4} \frac{\beta}{(\beta + \frac{1}{2})} \frac{M(\frac{1}{2}, \beta + \frac{3}{2}, \frac{1}{4}\alpha)}{M(\frac{1}{2}, \beta + \frac{1}{2}, \frac{1}{4}\alpha)} \tag{29}$$

The parameter  $\beta = 4Nu$  can be taken as a measure of the size of the population, while the ratio  $\beta/\alpha = u/ca^2$  can be equated to the critical quantity  $\delta$  of the previous section since it is assumed that  $cV$  is small. Values of  $E(p_i q_i)$  calculated from equation (29) for different values of  $\beta = 4Nu$  and of  $\beta/\alpha = u/ca^2 = \delta$  are given in Table 1; the irregular intervals of tabulation were dictated by available tables of the confluent hypergeometric function (Airey, 1926, 1927; Rushton & Lang, 1954; Slater, 1964). The average probability of heterozygosity is  $2E(p_i q_i)$ , and the expected value of the genetic variance is  $2na^2 E(p_i q_i)$ .

When the population size is infinite, it will be seen from the last column of Table 1 that  $E(p_i q_i)$  is the smaller of  $\delta$  and  $\frac{1}{4}$ ; this follows from the results of the preceding section. When the selection intensity is zero, so that  $c$  is zero and  $\delta$  infinite, it will be seen from the last row of Table 1 that  $E(p_i q_i) = \frac{1}{4}\beta/(\beta + \frac{1}{2})$ ; this follows from putting  $c = 0$  in equation (26), so that the distribution becomes a straightforward Beta distribution. As a reasonable approximation in the body of the table it can be seen that  $E(p_i q_i)$  is roughly equal to whichever is the smaller of  $\delta$  and  $\frac{1}{4}\beta/(\beta + \frac{1}{2})$ . This implies that the genetic variability is almost entirely determined by mutation and drift, and is affected little by selection, when  $\delta$  is greater than  $\frac{1}{4}\beta/(\beta + \frac{1}{2})$ ; on the other hand, when the converse is true, the genetic variability is affected little by drift and is determined almost entirely by mutation and selection.

Table 1. *The expected value of  $p_i q_i$  as a function of  $4Nu$  and  $\delta$*

$\frac{\beta}{\alpha} = \frac{u}{ca^2} = \delta$	$\beta = 4Nu$				
	0.1	0.5	1	3	$\infty$
0.025	0.0249	0.0294	0.0268	0.0255	0.0250
0.083	0.0356	0.0811	0.0935	0.0934	0.0833
0.250	0.0396	0.1095	0.1427	0.1843	0.2500
0.833	0.0410	0.1203	0.1598	0.2074	0.2500
2.50	0.0414	0.1234	0.1644	0.2122	0.2500
$\infty$	0.0417	0.1250	0.1667	0.2143	0.2500

Since the level of heterozygosity for randomly chosen loci in natural populations has been found to have a value of about 0.1, it is of interest to consider the values of  $\beta$  and  $\delta$  necessary to maintain such a level of heterozygosity under the present model. From the previous paragraph it follows that an approximate condition for  $E(p_i q_i)$  to be equal to 0.05 is that the smaller of  $\delta$  and  $\frac{1}{4}\beta/(\beta + \frac{1}{2})$  should be equal to 0.05; hence

$$\left. \begin{aligned} \text{either } \delta &= 0.05 \text{ and } \beta > 0.125 \\ \text{or } \delta &> 0.05 \text{ and } \beta = 0.125. \end{aligned} \right\} \tag{30}$$



If we assume that  $u = 10^{-5}$ , these conditions can be written in the form

$$\left. \begin{array}{l} \text{either } ca^2 = 2 \times 10^{-4} \text{ and } N > 3000 \\ \text{or } ca^2 < 2 \times 10^{-4} \text{ and } N = 3000. \end{array} \right\} \quad (31)$$

For example, let us suppose that the genetic variance of some quantitative character is 50 and the total variance 100, so that the heritability is  $\frac{1}{2}$ . It will be assumed that the coefficient of selection,  $k$ , is equal to 0.01, that the mutation rate,  $u$ , is  $10^{-5}$  and that the heterozygosity is 0.1. If the effective population size is considerably larger than 3000, then the genetic variability must be determined almost entirely by the joint effects of mutation and selection as discussed in the previous section. Hence the number of loci must be about 125 from equation (17) (unless some other means of maintaining genetic variability is operating); furthermore we can deduce that  $c = 0.5 \times 10^{-4}$  and that  $a^2 = 4$  from the facts that  $ca^2 = 2 \times 10^{-4}$  and that  $k = 2cV = 200c = 0.01$ . On the other hand, if the effective population size is about 3000, so that the genetic variability is determined by the joint effects of mutation and drift, then it can only be inferred that  $ca^2$  is less than  $2 \times 10^{-4}$ ; it can be shown as before that  $c = 0.5 \times 10^{-4}$ , but it only follows that  $a^2$  must be less than 4 and that the number of loci must be larger than 125.

#### REFERENCES

- AIREY, J. R. (1926). The confluent hypergeometric function. *British Association Reports*, pp. 276–294.
- AIREY, J. R. (1927). The confluent hypergeometric function. *British Association Reports*, pp. 220–244.
- BULMER, M. G. (1971*a*). The stability of equilibria under selection. *Heredity* **27**, 157–162.
- BULMER, M. G. (1971*b*). Stable equilibria under the two island model. *Heredity* **27**, 321–330.
- ERDÉLYI, A. (1954). *Tables of Integral Transforms*. (Bateman Manuscript Project.) New York: McGraw-Hill.
- FISHER, R. A. (1930). *The Genetical Theory of Natural Selection*. Oxford: Clarendon Press.
- HALDANE, J. B. S. (1932). *The Causes of Evolution*. London: Longmans, Green.
- HALDANE, J. B. S. (1953). The measurement of natural selection. *Proceedings of the IXth International Congress of Genetics*, pp. 480–483.
- KIMURA, M. (1964). Diffusion models in population genetics. *Journal of Applied Probability* **1**, 177–232.
- KIMURA, M. (1965). A stochastic model concerning the maintenance of genetic variability in quantitative characters. *Proceedings of the National Academy of Sciences, U.S.A.* **54**, 731–736.
- LATTER, B. D. H. (1960). Natural selection for an intermediate optimum. *Australian Journal of Biological Sciences* **13**, 30–45.
- LATTER, B. D. H. (1969). Models of quantitative variation and computer simulation of selection response. In *Computer Applications in Genetics* (ed. N. E. Morton), pp. 49–60. University of Hawaii Press.
- LATTER, B. D. H. (1970). Selection in finite populations with multiple alleles. II. Centripetal selection, mutation, and isoallelic variation. *Genetics* **66**, 165–186.
- ROBERTSON, A. (1956). The effect of selection against extreme deviants based on deviation or on homozygosis. *Journal of Genetics* **54**, 236–248.
- RUSHTON, S. & LANG, E. D. (1954). Tables of the confluent hypergeometric function. *Sankhyā: The Indian Journal of Statistics* **13**, 377–411.



- SLATER, L. J. (1964). Confluent hypergeometric functions. In *Handbook of Mathematical Functions* (ed. M. Abramowitz & I. A. Stegun), pp. 503–535. National Bureau of Standards, Applied Mathematics Series 55.
- WRIGHT, S. (1935). Evolution in populations in approximate equilibrium. *Journal of Genetics* **30**, 257–266.
- WRIGHT, S. (1937). The distribution of gene frequencies in populations. *Proceedings of the National Academy of Science* **23**, 307–320.