Genet. Res., Camb. (1973), **22**, *pp.* 9–12
Printed in Great Britain

The maintenance of the genetic variability of polygenic characters by heterozygous advantage

By M. G. BULMER
Department of Biomathematics, University of Oxford

(*Received* 13 *June* 1972)

SUMMARY

The results of a previous paper on the effect of optimizing selection, mutation and drift on a metric character determined by a large number of loci have been extended to include the possibility that, in addition to selection for an optimal value, there may be independent selection in favour of heterozygotes; it is assumed for simplicity that at each locus the heterozygote has the same advantage, s, over each of the homozygotes. Under selection alone there is a stable equilibrium if $s > ca^2$, where c is a measure of the intensity of the optimizing selection and a is the effect of a gene substitution. Under the additional forces exerted by mutation and by drift due to finite population size each locus behaves independently of the other loci as if it had a heterozygous advantage equal to $(s-ca^2)$.

The effects of optimizing selection, mutation and drift on the genetic variability of a polygenic metric character have been considered in a previous paper (Bulmer, 1972). The purpose of this paper is to extend these results to include the possibility that, in addition to selection for an optimal value, there may be independent selection in favour of heterozygotes.

As before we shall consider 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 genotypes, C_1C_1, C_1C_2 and C_2C_2, are $-a$, 0 and a respectively. In addition to the genetic variability, $V_G = 2a^2\Sigma p_i q_i$, where p_i is the frequency of the C_1 allele at the ith locus, there may also be an environmental component of the variance, V_E, so that the total phenotypic variance is $V = V_G + V_E$. As a model of optimizing selection we shall suppose that the fitness of an individual with phenotypic value y is

$$w(y) = \exp[-c(y-\theta)^2], \tag{1}$$

where c is a measure of the intensity of selection and θ is the optimal value of the character. It is often more convenient to express the intensity of selection in terms of the dimensionless quantity

$$k = 2cV/(1+2cV), \tag{2}$$

which has been called by Latter (1970) 'the coefficient of centripetal selection'. If y is normally distributed in the population before selection with mean M and

variance V, then it will also be normally distributed after selection with mean $M+DM$ and variance $V+DV$, where

$$DM = k(\theta - M), \qquad DV = -kV. \tag{3}$$

In addition to this optimizing selection it will also be supposed that at each locus the heterozygote, C_1C_2, has a small selective advantage, s, over either of the homozygotes.

We shall first consider how large a heterozygote advantage is required to maintain a stable equilibrium in the absence of mutation or drift. The change in the gene frequency in one generation at the ith locus due to the selection for an optimal value is given by

$$\Delta p_i \text{ (optimizing selection)} = p_i q_i [Aa - \tfrac{1}{2}Ba^2(p_i - q_i)], \tag{4}$$

where

$$A = -DM/V, \quad B = (DV + DM^2)/V^2 \tag{5}$$

(see Bulmer, 1971, 1972). The change in the gene frequency due to heterosis is

$$\Delta p_i \text{ (heterosis)} = -sp_i q_i (p_i - q_i). \tag{6}$$

Hence the total change in the gene frequency due to selection is

$$\Delta p_i \text{ (selection)} = p_i q_i [Aa - (p_i - q_i)(s + \tfrac{1}{2}Ba^2)]. \tag{7}$$

(It is assumed that both s and ca^2 are small so that the changes in the gene frequency due to the two selective forces can be added together with negligible error.) At equilibrium Δp_i must be zero, so that, unless $p_i = 0$ or 1, the equilibrium gene frequency, P, must be the same at all loci and must satisfy the equation

$$Aa - (P - Q)(s + \tfrac{1}{2}Ba^2) = 0. \tag{8}$$

To investigate the stability of the equilibrium suppose that it is subjected to a small perturbation so that $p_i = P + e_i$; then in the next generation

$$e_i^* = e_i + \Delta p_i = e_i \left(1 - \frac{1}{2}\frac{h^2 VB}{n} - 2PQs\right) + \frac{h^2 VB}{n} \sum_{j=1}^{n} e_j. \tag{9}$$

after allowing for the change in A due to the perturbation. (See Bulmer, 1971; in this formula h^2 is the heritability, so that $h^2 V = V_G$.) If we write $\bar{e} = \Sigma e_i/n$, $\delta_i = (e_i - \bar{e})$, it follows that

$$\bar{e}^* = \left[1 - 2PQs + Bh^2 V \left(1 - \frac{1}{2n}\right)\right] \bar{e}, \quad \delta_i^* = \left[1 - \frac{1}{2}\frac{Bh^2 V}{n} - 2PQs\right] \delta_i. \tag{10}$$

Since $B = -k/V$, $\bar{e}$ is clearly stable; the criterion for the stability of δ_i, and hence for the whole process, is that

$$s > kh^2/4nPQ = ka^2/2V. \tag{11}$$

When the intensity of selection is low, k is approximately equal to $2cV$, so that this criterion can be expressed in the form

$$s > ca^2. \tag{12}$$

https://doi.org/10.1017/S0016672300012799 Published online by Cambridge University Press

We shall now suppose that all loci are subject to equal forward and backward mutation rates, u, and that the population is finite, of effective size N. The gene frequencies, p_i, at the various loci will thus be subject to drift and will be random variables with a joint probability distribution. It has been shown by Wright (1937) and Kimura (1964) that the stationary distribution of gene frequencies under the diffusion approximation is given by

$$\phi(p_1, p_2, \ldots, p_n) \propto \bar{w}^{2N} \prod_{i=1}^{n} (p_i q_i)^{4Nu-1}, \tag{13}$$

where $\bar{w}$ is the average fitness in a population with the given gene frequencies. To evaluate $\bar{w}$, we first observe that the fitness of an individual depends both on the phenotypic value of the character, y, and on the number of homozygous loci, h, and may be taken as

$$\begin{aligned} w(y, h) &= (1-s)^h \exp -[c(y-\theta)^2] \\ &\simeq \exp -[sh + c(y-\theta)^2]. \end{aligned} \tag{14}$$

Hence

$$\begin{aligned} \bar{w} &\simeq E\{\exp -[sh + c(y-\theta)^2]\} \\ &\simeq \exp -E[sh + c(y-\theta)^2], \end{aligned} \tag{15}$$

since, to a good approximation, the order of exponentiation and of taking the expected value can be interchanged. The average number of homozygous loci is $n - 2\Sigma p_i q_i$, since if the gene frequency at the ith locus is p_i, then the probability that an individual will be homozygous at that locus is $1 - 2p_i q_i$; if $E(y) = M$ and $V(y) = V$ under the given gene frequencies, then $E(y-\theta)^2 = V + (M-\theta)^2$. Hence

$$\bar{w} \simeq \exp -[s(n - 2\Sigma p_i q_i) + cV + c(M-\theta)^2]. \tag{16}$$

If we confine our attention to the symmetrical case in which $M = \theta$ when the average gene frequency is $\frac{1}{2}$, that is to say when $\Sigma(p_i - \frac{1}{2}) = 0$, then

$$(M-\theta) = -2a\Sigma(p_i - \tfrac{1}{2}),$$

so that

$$\begin{aligned} \phi(p_1, p_2, \ldots, p_n) \propto \exp\{&4Ns\Sigma p_i q_i - 8Nca^2[\Sigma(p_i - \tfrac{1}{2})]^2 \\ &- 4Nca^2 \Sigma p_i q_i\} \, \Pi (p_i q_i)^{4Nu-1}. \end{aligned} \tag{17}$$

The gene frequencies are not independently distributed in equation (17) because of the factor $[\Sigma(p_i - \frac{1}{2})]^2$. It can be shown, however (see Bulmer, 1972) that this factor can be ignored if we are only interested in the genetic variability maintained as determined by the quantity $\Sigma p_i q_i$; it follows that from this point of view the loci can be regarded as if they were n independent loci with gene frequencies determined by the joint effects of mutation (with mutation rate u), random drift (in a population of effective size N), and of selection with heterozygous advantage $s - ca^2$. The average genetic variability is thus na^2H, where H is the expected value of $2p_i q_i$, the average heterozygosity. It has been shown elsewhere (Bulmer, 1973) that H can be evaluated exactly in terms of the confluent hypergeometric function,

and that when $\beta = 4Nu$ is small and $\alpha = N(s - ca^2)$ is positive a good approximation is given by

$$H \simeq \left(\frac{2\sqrt{\alpha}}{\beta\sqrt{\pi}} \frac{\exp[-\alpha]}{\operatorname{erf}\sqrt{\alpha}} + 2\right)^{-1}. \tag{18}$$

(This approximation differs from the formula given by Robertson (1962, equation 2) in containing the additional quantity 2, which makes it considerably more accurate.)

In interpreting the above result it seems reasonable to equate the mutation rate, u, with the mutation rate per nucleotide site rather than per cistron since it is assumed that there is only one heterotic pair of alleles at each locus. A considerable amount of heterozygosity can nevertheless be maintained by quite a small heterozygous advantage even in moderate-sized populations (see Bulmer, 1973).

REFERENCES

BULMER, M. G. (1971). The stability of equilibria under selection. *Heredity* **27**, 157–162.

BULMER, M. G. (1972). The genetic variability of polygenic characters under optimizing selection, mutation and drift. *Genetical Research* **19**, 17–25.

BULMER, M. G. (1973). The maintenance of heterozygosity under the diffusion model. (In preparation.)

KIMURA, M. (1964). Diffusion models in population genetics. *Journal of Applied Probability* **1**, 177–232.

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. (1962). Selection for heterozygotes in small populations. *Genetics* **47**, 1291–1300.

WRIGHT, S. (1937). The distribution of gene frequencies in populations. *Proceedings of the National Academy of Sciences, U.S.A.* **23**, 307–320.