Genet. Res., Camb. (1976), **28**, *pp.* 101–117
With 3 text-figures
Printed in Great Britain

The effect of selection on genetic variability: a simulation study

By M. G. BULMER
Department of Biomathematics, Pusey Street, Oxford

(*Received* 26 *January* 1976)

SUMMARY

Computer simulations have been done to study the effects of stabilizing and disruptive selection on a polygenic character. The results are reported in terms of three components of genetic variability which represent changes in gene frequencies, departures from Hardy–Weinberg equilibrium and linkage disequilibrium respectively. Under random mating the first and third components are the most important. The observed changes in gene frequencies are interpreted in the light of previous theoretical work on the stability of equilibria under selection. In addition, large and rapid changes in the genotypic variance result from the generation of linkage disequilibrium under selection; the observed changes are in good agreement with those predicted on theoretical grounds.

1. INTRODUCTION

Selection can change both the mean and the variance of a metric character. Its theoretical effect on the mean is well known, but its effects on the variance is more difficult to study in theory and is less well understood. Theoretical understanding of this problem is nevertheless important both in interpreting the results of experiments on disruptive and stabilizing selection and in investigating problems such as the evolution of niche width in natural populations. The purpose of this paper is to report the results of computer simulations designed to illustrate the effect of selection on genetic variability and to interpret these results in the light of some theoretical results which have been obtained previously.

It is important to distinguish changes in the variance due to changes in gene frequencies from those due to departures from Hardy–Weinberg or linkage equilibrium. It is therefore convenient to partition the genetic variance of a metric character into three components representing these three factors:

$$V_G = V_g + C_{HW} + C_L. \tag{1}$$

In this formula V_g is the genotypic variance calculated from the distribution of genotypic values; V_g, which will be called the genic variance, is the variance calculated from the observed gene frequencies on the assumption that there is perfect Hardy–Weinberg and linkage equilibrium; $(V_g + C_{HW})$ is defined as the variance calculated from the observed genotype frequencies at each locus on the assumption that there is perfect linkage equilibrium between loci; and C_L is defined as the difference between V_G and $(V_g + C_{HW})$. Thus C_{HW} measures the

effect on the variance resulting from covariances between alleles at the same locus due to Hardy–Weinberg disequilibrium, and C_L measures the effect resulting from covariances between loci due to linkage disequilibrium. (It should be noted that 'linkage disequilibrium' is understood to include correlations between loci on homologous chromosomes as well as 'gametic phase disequilibrium' between loci on the same chromosome; it is not advantageous to represent these covariances separately since the distinction ceases to be meaningful for loci on different chromosomes.)

The genotypic variance can always be partitioned in this way. In this paper we shall consider a simple model of a metric character determined by n loci without dominance or epistasis; it will be assumed that each locus has two alleles labelled + and − which contribute 1 and 0 respectively to the character. (In addition there will be an environmental component with variance V_e.) The formulae for the three components of the genotypic variance under this model are:

$$\left.\begin{aligned} V_g &= 2\sum_i p_i(1-p_i), \\ C_{HW} &= \sum_i (2P_i(0)P_i(2) - \tfrac{1}{2}P_i^2(1)), \\ C_L &= 2\sum_{i<j} \mathrm{Cov}\,(i,j). \end{aligned}\right\} \qquad (2)$$

In these formulae p_i is the frequency of the +allele at the ith locus, $P_i(.)$ is the frequency with which 0, 1 or 2 +alleles are found at the ith locus, and Cov (i,j) is the covariance between the numbers of +alleles at loci i and j.

These components of variance may be affected both by sampling fluctuations due to finite population size and by systematic pressures such as selection. We shall first consider the effect of population size in the absence of selection; it will be shown that the genic variance declines fairly steadily (in the absence of mutation) due to genetic drift but that the genotypic variance may fluctuate considerably above and below the genic variance due to departures from Hardy–Weinberg and in particular from linkage equilibrium. We shall then consider the effects of disruptive, stabilizing and directional selection; it will be shown that these can best be understood by considering separately the stability of equilibria under selection, which determines the genic variance, and the generation of linkage disequilibrium which determines C_L.

2. DESCRIPTION OF THE SIMULATIONS

The metric character studied was assumed to be determined by 12 loci each with two alleles, + and −, contributing 0 and 1 to the character without dominance or epistasis. The genotypic value was thus the number of + alleles, ranging from 0 to 24. The phenotypic value was obtained by adding a normally distributed environmental component with zero mean and variance 4.

Two alternative assumptions about linkage were made. In one set of simulations (called mouse simulations) the 12 loci were all assumed to be on different chromo-

somes and hence to segregate independently. In the other set of simulations (called *Drosophila* simulations) the loci were assumed to be arranged in groups of 4 on 3 chromosomes; non-homologous chromosomes were assumed to segregate independently in both sexes, and the recombination fraction between adjacent loci on the same chromosome was taken to be 0·1 in females (with no interference) and zero in males.

In each generation 100 individuals of each sex were selected (either at random, or with directional, disruptive or stabilizing selection) to be the parents of the next generation. These selected individuals were then paired, in most experiments at random but in some cases assortatively, and each pair was programmed to produce 10 offspring (with one exception to be described shortly); from the 1000 offspring produced 500 were chosen at random to be male and 500 female. (An appropriate random number generator was used to choose individuals at random, to simulate recombination and segregation, and to generate normal deviates.)

Several different selection procedures were used. In the *control simulations* 100 males and 100 females were chosen at random from the 500 individuals of each sex in the population. Two types of *stabilizing selection* were used. In the first type (*rank selection*) the 500 individuals of each sex were arranged in rank order by their phenotypic value and the middle 100 were selected; in the second (*value selection*) the 100 individuals of each sex with phenotypic values nearest 12 were selected. Similarly two types of *disruptive selection* were used. In the first (*rank selection*) the 500 individuals of each sex were arranged in rank order and the largest 50 and the smallest 50 were selected; in the second (*value selection*) the 100 individuals of each sex with phenotypic values furthest from 12 were selected. In *directional selection* the largest 100 individuals of each sex were selected. All these procedures used a 20 % selection rate. The effect of a 50 % selection rate under disruptive selection was also investigated. In these simulations the largest 50 and the smallest 50 individuals of each sex were selected, but each pair was programmed to produce 4 instead of 10 offspring, so that the population before selection contained 200 males and 200 females. The different selection procedures are summarized in Table 1.

In most simulations the selected males were mated at random with the selected females, but the effects of positive and negative assortative mating were also investigated under disruptive rank selection. In positive assortative mating the 50 largest males were mated at random with the 50 large females, and the small males with the small females; in negative assortative mating the 50 large males were mated at random with the 50 small females and vice versa.

The simulations were started by selecting 500 individuals of each sex (200 under 50% selection) from a very large initial population in perfect Hardy–Weinberg and linkage equilibrium. To investigate the effect of different initial gene frequencies, the frequency of the + allele in the initial population was taken to be 0·25 at 3 loci, 0·5 at 6 loci and 0·75 at the remaining 3 loci. In the Drosophila simulations the + allele frequencies at successive positions on each chromosome were 0·25, 0·5, 0·75 and 0·5 respectively. The simulations were continued for 20, or sometimes 30, generations.

The three components of variance defined in (2) were calculated and printed in each generation (before selection) together with the distribution of genotypic values and the gene frequencies at the 12 loci. The phenotypic mean and variance both before and after selection were also calculated and printed in each generation.

Table 1. *Description of the selection procedures*

Type of selection	Individuals of each sex selected	Number of each sex before selection
Control	100 at random	500
Stabilizing rank	Middle 100 in rank order	500
Stabilizing value	100 with values nearest 12	500
Disruptive rank	Largest 50 and smallest 50	500
Disruptive rank (50%)	Largest 50 and smallest 50	200
Disruptive value	100 with values furthest from 12	500
Directional	Largest 100	500
Type of mating		
Random	100 selected males mated at random with 100 selected females	
Positive assortative*	Largest 50 males mated at random with largest 50 females and smallest 50 males with smallest 50 females	
Negative assortative*	Largest 50 males mated at random with smallest 50 females and vice versa	

* Disruptive rank selection only.

3. CONTROL SIMULATIONS

In these simulations 100 males and 100 females were chosen at random in each generation from the 500 individuals of each sex in the population and were then paired at random. The results show what is likely to happen in the absence of selection. Two mouse and two Drosophila simulations were done, each lasting for 20 generations. The effective population size can be calculated as $N = 4 \times 200/(2+\sigma^2)$, where σ^2 is the variance of the number of offspring which survive to breed; this is the binomial variance, $\sigma^2 = 10 \times 0{\cdot}2 \times 0{\cdot}8 = 1{\cdot}6$, since each of the 10 offspring has a survival probability of 0·2. Hence $N = 222$.

The genotypic and genic variances in the first mouse simulation are shown in Fig. 1. It can be seen that the genic variance declines slowly but steadily, presumably because of genetic drift. Table 2 shows the genic variance after 20 generations in all four experiments. The genic variance should decrease through drift by a fraction $1/2N$ per generation, or by 4·5 % in 20 generations. It will be seen from Table 2 that the observed genic variance after 20 generations averaged over the four simulations is in excellent agreement with its predicted value.

It can also be seen from Fig. 1 that the genotypic variance fluctuates quite

noticeably above and below the genic variance, presumably due to departures from Hardy–Weinberg and linkage equilibrium. The components C_{HW} and C_L which measure these effects have been computed in each generation; the mean, standard deviation and first serial correlation coefficient have been calculated

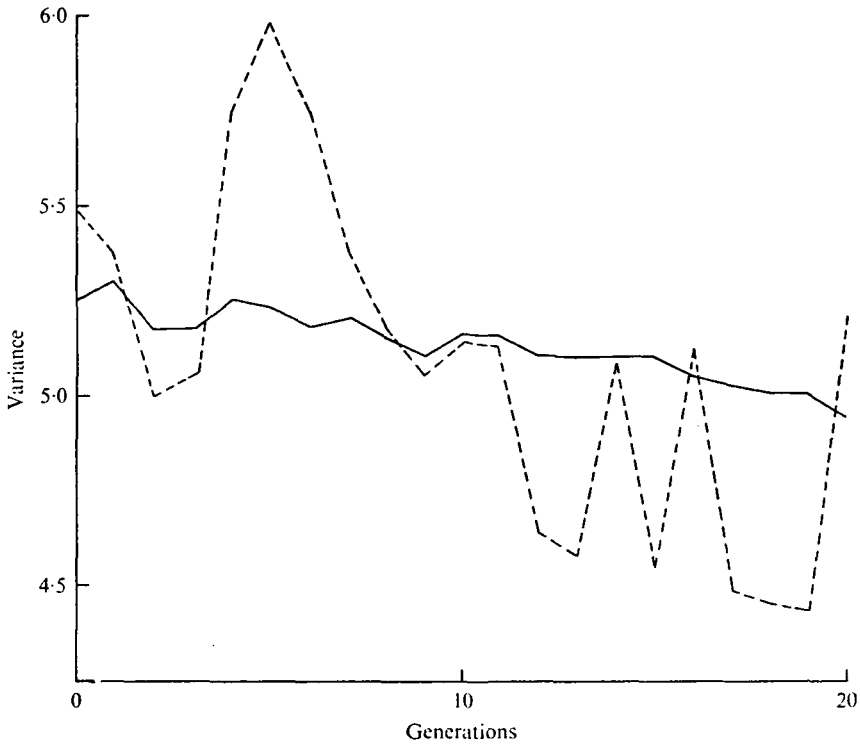

Fig. 1. Genic variance (solid) and genotypic variance (dotted) for the first mouse control simulation.

Table 2. *The genic variance after* 20 *generations with no selection*

Mouse 1	4·92
Mouse 2	4·99
Drosophila 1	4·90
Drosophila 2	5·19
Average	5·00
Variance in initial population	5·25
Predicted variance after 20 generations	5·01

from the values for each component in each of the four experiments for generations 6–20 and the results are shown in Table 3. (In interpreting these results it should be remembered that the average genic variance was about 5·1.) The mean values are in both cases nearly zero as it to be expected with no selection under random mating. The standard deviation of C_L is considerably larger than that of C_{HW}, which shows that most of the fluctuations in the genotypic variance are due to randomly generated linkage disequilibrium. It will also be seen that there is a positive correlation between successive values of C_L. The practical importance of

this is that observations on the genotypic variance in successive years are not independent of one another, so that more observations are required to obtain a given accuracy than if they were independent.

Table 3. *Departures from Hardy–Weinberg and linkage equilibrium with no selection*

	C_{HW}			C_L		
	Mean	S.D.	r_1	Mean	S.D.	r_1
Mouse 1	0·00	0·11	−0·04	−0·10	0·43	0·31
Mouse 2	−0·04	0·10	−0·19	−0·02	0·45	0·62
Mouse (predicted)	−0·01	0·10	0·00	0·00	0·38	0·50
Drosophila 1	−0·03	0·10	+0·10	0·48	0·63	0·64
Drosophila 2	−0·02	0·09	−0·27	−0·11	0·49	0·50
Drosophila (predicted)	−0·01	0·10	0·00	0·00	0·58	>0·50

The predicted values in Table 3 were obtained as follows. It can be shown by standard techniques that

$$\left.\begin{aligned} E(C_{HW}) &= -V_g/2N, \\ \text{Var}\,(C_{HW}) &\simeq 4\sum_i p_i^2(1-p_i)^2/N. \end{aligned}\right\} \tag{3}$$

If the gene frequencies are all the same the variance simplifies to

$$\text{Var}\,(C_{HW}) \simeq V_g^2/nN \tag{4}$$

in the general case with n loci; this expression gives a good approximation provided that the gene frequencies are not too different. The mean and standard deviation of C_{HW} can therefore be predicted by putting $V_g = 5{\cdot}1$ in (3) and (4). There is good agreement between the observed and predicted values.

We turn now to the distribution of C_L. Let D_{ij} be the gametic phase disequilibrium between loci i and j defined as the determinant of the gametic frequencies, and let c_{ij} be the recombination fraction between these loci. (In the Drosophila simulations c_{ij} will be taken as the average of the recombination fractions in the two sexes.) Provided that $Nc_{ij} \gg 1$ (which is satisfied here) it follows from results of Hill & Robertson (1968) that

$$\left.\begin{aligned} E(D_{ij}) &= 0 \\ \text{Var}\,(D_{ij}) &\simeq p_i(1-p_i)p_j(1-p_j)/2Nc_{ij}(2-c_{ij}). \end{aligned}\right\} \tag{5}$$

It has also been shown by Hill (1974) that covariances between disequilibria at different pairs of loci are zero. These results were obtained under a rather unrealistic haploid model in which no random variation was permitted in obtaining the zygotic from the gametic frequencies. It follows that $\text{Cov}\,(i,j) = 2D_{ij}$ exactly under this model, so that

$$\left.\begin{aligned} E(C_L) &= 0 \\ \text{Var}\,(C_L) &\simeq 8\sum_{i<j} p_i(1-p_i)p_j(1-p_j)/Nc_{ij}(2-c_{ij}). \end{aligned}\right\} \tag{6}$$

https://doi.org/10.1017/S0016672300016797 Published online by Cambridge University Press

A good approximation to the variance, obtained by assuming all the gene frequencies to be the same is given by

$$\text{Var}\,(C_L) \simeq (n-1)V_g^2/nN\alpha, \tag{7}$$

where α is the harmonic mean of the quantities $c_{ij}(2-c_{ij})$, which is equal to 0·75 for the mouse and 0·323 for the Drosophila simulation. The observed standard deviation of C_L is in good agreement with the value predicted from (7). This theoretical value should, however, be regarded with some caution because of the unrealistic model under which it was obtained.

It will be seen from (4) and (7) that the standard deviation of C_L does not decline, relative to V_g, as the number of loci increases whereas that of C_{HW} does; linkage disequilibrium is thus the dominant source of random fluctuations in the genotypic variance of a character determined by many loci. The reason for this is that n loci contribute to C_{HW} whereas $\frac{1}{2}n(n-1)$ pairs of loci contribute to C_L.

Successive values of C_{HW} will be independent of one another but successive values of C_L will be correlated, because the correlation between successive values of D_{ij} is $(1-c_{ij})$. The correlation between successive values of C_L should clearly be $\frac{1}{2}$ in the absence of linkage and greater than $\frac{1}{2}$ in the presence of linkage; this expectation is confirmed by the observed results.

4. THE STABILITY OF EQUILIBRIA UNDER SELECTION

The genic variance depends only on gene frequencies; it will therefore be affected by any directional changes in gene frequencies or by their movement towards (or away from) stable (or unstable) equilibria. In this section we shall investigate how stability or instability affects the genic variance under stabilizing and disruptive selection.

The problem of determining the stability of equilibria has been investigated elsewhere (Bulmer, 1971*a*, 1974*a*) under a rather general model of selection acting on a character determined by n loci with equal effects without dominance and under random mating. Suppose that there is an equilibrium with all the gene frequencies taking some common value, $\hat{p}$ (which is $\frac{1}{2}$ under the models considered here). The stability of this equilibrium depends on two quantities which will here be called β and γ and which are defined as

$$\left.\begin{aligned}\beta &= (V^*-V)/V,\\ \gamma &= \text{Cov}\,(y, \partial W/\partial M).\end{aligned}\right\} \tag{8}$$

In these equations y is the phenotypic value of the character which is assumed to be normally distributed with mean M and (phenotypic) variance V before selection; W is the absolute fitness of an individual with value y, and the effect of selection is to change the phenotypic variance from V to V^*. Thus β will be negative under stabilizing and positive under disruptive selection. γ is rather more difficult to interpret, but seems to be related to the 'softness' or 'hardness' of selection.

If these quantities are evaluated at the equilibrium then it was shown that the following recurrence relations are approximately true near the equilibrium:

$$\left.\begin{aligned}[p_i(t+1)-\bar{p}(t+1)] &= \left[1-\frac{h^2\beta}{2n}\right][p_i(t)-\bar{p}(t)],\\ [\bar{p}(t+1)-\hat{p}] &= [1+h^2(\beta+\gamma)]\,[\bar{p}(t)-\hat{p}].\end{aligned}\right\} \tag{9}$$

In these equations $p_i(t)$ and $\bar{p}(t)$ denote the gene frequency at the ith locus and the average gene frequency respectively in generation t and h^2 is the heritability. Thus the quantities $(p_i-\bar{p})$ are stable or unstable according as β is positive or negative; we would therefore expect the variance of the gene frequencies, defined as

$$\sigma^2(p) = \sum_i (p_i-\bar{p})^2/n,$$

to decrease towards zero if β is positive and to increase if β is negative. This kind of stability will be called *stability in variance*. On the other hand, $(\bar{p}-\hat{p})$ is stable or unstable according as $\beta+\gamma$ is negative or positive; this will be called *stability in mean*.

Table 4. *Observed and predicted values of* β

Type of selection	Observed β	Predicted β
Stabilizing rank	−0·9769	−0·9785
Stabilizing value	−0·9785	−0·9785
Disruptive rank	2·17	2·25
Disruptive rank (50%)	0·84	0·86
Disruptive value	—	2·25

If the phenotype value is normally distributed then β can be evaluated from the following formulae under the forms of truncation selection used here:

$$\beta = -2z\phi(z)/P \text{ under stabilizing selection,} \tag{10a}$$

$$\beta = 2z\phi(z)/P \text{ under disruptive selection,} \tag{10b}$$

$$\beta = -\frac{\phi(z)}{P}\left[\frac{\phi(z)}{P}-z\right] \text{ under directional selections.} \tag{10c}$$

In these formulae P is the proportion of animals selected, taken from the middle of the distribution under stabilizing selection, taken equally from the two tails of the distribution under disruptive selection and taken from one tail under directional selection (though there is no equilibrium in the last case); z is the standard normal deviate corresponding to $\frac{1}{2}(1+P)$ in (10a), to $(1-\frac{1}{2}P)$ in (10b) and to $(1-P)$ in (10c); and $\phi(z)$ is the standard normal density function.

Table 4 compares the values of β predicted from (10) with the values calculated by using the definition in (8) from the observed phenotypic variances in the simulations before and after selection. The first ten (or sometimes fifteen) generations during which linkage disequilibrium was building up were excluded; no value was calculated for the disruptive value simulation in which the gene frequencies moved rapidly to fixation. The values for the mouse and Drosophila

simulations were similar in all cases and were averaged. It will be seen that the observed value is very close to the predicted value under stabilizing selection, but is slightly less than the predicted value under disruptive selection; this can be attributed to departures from normality in the latter case caused by the high degree of linkage disequilibrium.

Table 5. *Predicted behaviour under different types of selection*

Type of selection	β	$\beta+\gamma$	Stability in variance	Stability in mean	$\sigma^2(p)$	$\bar{p}$
Stabilizing rank	<0	0	Unstable	Neutral	Increase	Random
Stabilizing value	<0	<0	Unstable	Stable	Increase	Near $\frac{1}{2}$
Disruptive rank	>0	0	Stable	Neutral	Decrease	Random
Disruptive value	>0	>0	Stable	Unstable	Decrease	Near 0 or 1

Table 6. *Observed stability behaviour after* 20 *generations of selection*

Type of selection	Type of simulation	Type of mating	$\sigma^2(p)$	$\bar{p}$	V_g
Control*	Mouse	Random	0·043	0·528	4·96
Control*	Drosophila	Random	0·040	0·521	5·04
Stabilizing rank	Mouse	Random	0·060	0·528	4·53
Stabilizing rank	Drosophila	Random	0·059	0·492	4·58
Stabilizing value	Mouse	Random	0·057	0·499	4·63
Stabilizing value	Drosophila	Random	0·083	0·503	4·01
Disruptive rank	Mouse	Random	0·000	0·500	6·00
Disruptive rank	Drosophila	Random	0·000	0·500	6·00
Disruptive rank	Mouse	Positive	0·000	0·500	6·00
Disruptive rank	Drosophila	Positive	0·000	0·500	6·00
Disruptive rank	Mouse	Negative	0·016	0·499	5·62
Disruptive rank	Drosophila	Negative	0·014	0·511	5·66
Disruptive rank (50%)	Mouse	Random	0·010	0·455	5·70
Disruptive rank (50%)	Drosophila	Random	0·009	0·549	5·72
Disruptive value	Mouse	Random	0·000	0·000	0·00
Disruptive value	Drosophila	Random	0·000	1·000	0·00
Value in initial population			0·031	0·500	5·25

* Average of two simulations.

It is also quite easy to show under the assumption of normality that $\gamma = \beta$ under both types of rank selection, while $\gamma = 0$ under both types of value selection. The predicted behaviour under different types of selection is summarized in Table 5, and the observed behaviour after 20 generations of selection is shown in Table 6. The observed behaviour can be characterized by the variance of the gene frequencies, $\sigma^2(p)$, and the mean gene frequency, $\bar{p}$, but it is also of interest to consider the genic variance which is related to these two quantities by the formula

$$V_g = 2n[\bar{p}(1-\bar{p}) - \sigma^2(p)]. \tag{11}$$

It will be seen that the predictions in Table 5 are in most cases fulfilled in

Table 6. The variance in the gene frequencies, $\sigma^2(p)$, has increased considerably more under both types of stabilizing selection than in the control simulations and has decreased under both types of disruptive selection. The mean gene frequency, $\bar{p}$, has moved a small distance, apparently at random, under both stabilizing rank and disruptive rank (50 %) selection, has remained very close to $\frac{1}{2}$ under stabilizing value selection and has been fixed at one or other of the boundaries under disruptive value selection. There is, however, one exception. Under disruptive rank selection (with 20 % selection) with either random or positive assortative mating the mean gene frequency was exactly $\frac{1}{2}$ after 20 generations and was clearly positively rather than neutrally stable. It will be shown in the next section that linkage disequilibrium becomes so large in these simulations that the system behaves as if it is controlled by a single locus with two alleles having values 0 and 12. The stability analysis outlined above is inappropriate since it assumes that the system behaves in a polygenic manner. It is easily verified that in a single locus situation in which equal numbers of both homozygotes are selected and mated either at random or with positive assortative mating, there is a stable equilibrium with a gene frequency of $\frac{1}{2}$.

It is fairly easy to see intuitively the reason for stability or instability in mean. Under stabilizing value selection in which individuals with phenotypic values nearest 12 are selected, an external optimum is imposed on the system, whereas no preference is given to any particular value under stabilizing rank selection. Under disruptive value selection in which individuals with phenotypic values furthest from 12 are selected, the system will clearly be repelled from this value whereas no disadvantage is given to any particular value under disruptive rank selection. In fact disruptive value selection behaves, after the first few generations, like directional selection in which the direction is determined by random events during the first few generations.

The main weakness of the above theory is that the critical equations (8) defining stability were derived under the assumption that there is no linkage disequilibrium. It seems reasonable to suppose, however, that they will remain approximately true as linkage disequilibrium builds up under selection (if linkage is not too strong) provided that h^2 is interpreted as the ratio V_G/V which allows for the effect of disequilibrium on the apparent heritability and provided that allowance is also made for any change in the quantities β and γ; under truncation selection these dimensionless quantities depend only on the shape of the phenotypic distribution and so change little as demonstrated for β in Table 4.

To test the validity of (8) I have calculated the observed regression coefficients

$$\left.\begin{aligned} b_1 &= \frac{\sum_{i,t} (d_i(t+1) - d_i(t))\, d_i(t)}{\sum_{i,t} d_i^2(t)}, \\ b_2 &= \frac{\sum_{t} (\bar{p}(t+1) - \bar{p}(t))\, (\bar{p}(t) - \tfrac{1}{2})}{\sum (\bar{p}(t) - \tfrac{1}{2})^2}. \end{aligned}\right\} \tag{12}$$

https://doi.org/10.1017/S0016672300016797 Published online by Cambridge University Press

In these equations $d_i(t) = p_i(t) - \bar{p}(t)$; the regression lines have been constrained to pass through the origin. If (9) is valid then b_1 is an estimate of $-h^2\beta/24$ and b_2 is an estimate of $h^2(\beta+\gamma)$. The results are shown in Table 7. The disruptive value simulations, which moved very rapidly towards fixation, have been excluded, as have the simulations with non-random mating. Only the first ten generations of the disruptive rank simulations have been used, before they began to behave like a single locus model. The observed values of b_2 are in reasonable agreement with their predicted values. The observed values of b_1, though always of the correct sign, tend to be smaller than expected under stabilizing selection and larger than expected, in absolute value, under disruptive selection. They are nevertheless of the right order of magnitude, and it can be concluded that the stability equations (9) remain at least qualitatively valid when linkage disequilibrium becomes established.

Table 7. *Observed and predicted values of stability coefficients*

Type of selection	Type of simulation	$b_1 \pm$ S.E.	$-h^2\beta/24$	$b_2 \pm$ S.E.	$h^2(\beta+\gamma)$
Stabilizing rank	Mouse	$0{\cdot}012 \pm 0{\cdot}006$	0·019	$0{\cdot}03 \pm 0{\cdot}05$	0
Stabilizing rank	Drosophila	$0{\cdot}009 \pm 0{\cdot}005$	0·016	$-0{\cdot}09 \pm 0{\cdot}08$	0
Stabilizing value	Mouse	$0{\cdot}010 \pm 0{\cdot}007$	0·019	$-0{\cdot}46 \pm 0{\cdot}18$	−0·47
Stabilizing value	Drosophila	$0{\cdot}011 \pm 0{\cdot}004$	0·016	$-0{\cdot}31 \pm 0{\cdot}13$	−0·39
Disruptive rank	Mouse	$-0{\cdot}116 \pm 0{\cdot}015$	−0·082	$-0{\cdot}22 \pm 0{\cdot}18$	0
Disruptive rank	Drosophila	$-0{\cdot}104 \pm 0{\cdot}015$	−0·082	$-0{\cdot}03 \pm 0{\cdot}13$	0
Disruptive rank (50%)	Mouse	$-0{\cdot}044 \pm 0{\cdot}009$	−0·028	$0{\cdot}00 \pm 0{\cdot}04$	0
Disruptive rank (50%)	Drosophila	$-0{\cdot}039 \pm 0{\cdot}008$	−0·029	$-0{\cdot}04 \pm 0{\cdot}05$	0

5. THE GENERATION OF LINKAGE DISEQUILIBRIUM UNDER SELECTION

We turn now to the effect of selection in generating linkage disequilibrium and so causing the mean value of C_L to depart from zero. Figs. 2 and 3 show the genic and genotypic variances in the mouse stabilizing and disruptive rank simulations (with random mating). The genotypic variance fluctuates considerably as in the control simulation shown in Fig. 1 (though the difference in scale between Figs. 1, 2 and 3 should be noted), but about a mean level which is below the genic variance under stabilizing selection and above the genic variance under disruptive selection. Under random mating the difference in mean between the genotypic variances is due entirely to a negative (or positive) value of C_L (see Table 8) which reflects the generation of negative (or positive) gametic phase disequilibrium under stabilizing (or disruptive) selection.

This phenomenon has been considered theoretically elsewhere (Bulmer, 1971*b*, 1974*b*), and it was shown that, when a steady state is reached after a number of generations of selection, C_L attains a mean value which satisfies the equation

$$C_L(V_g + V_e + C_L) = \tfrac{1}{2}\beta(V_a + C_L)^2/H. \tag{13}$$

In this equation V_g, V_e and V_a are respectively the genic variance, the environmental variance and the additive contribution to the genic variance, H is the

harmonic means of the recombination fractions, and β is defined in (8). Under truncation selection β is a known constant (or very nearly so, see equation 10 and Table 4), so that (13) can be solved as a quadratic equation in C_L. The appropriate solution is the smallest root (in absolute value) which has the same sign as β.

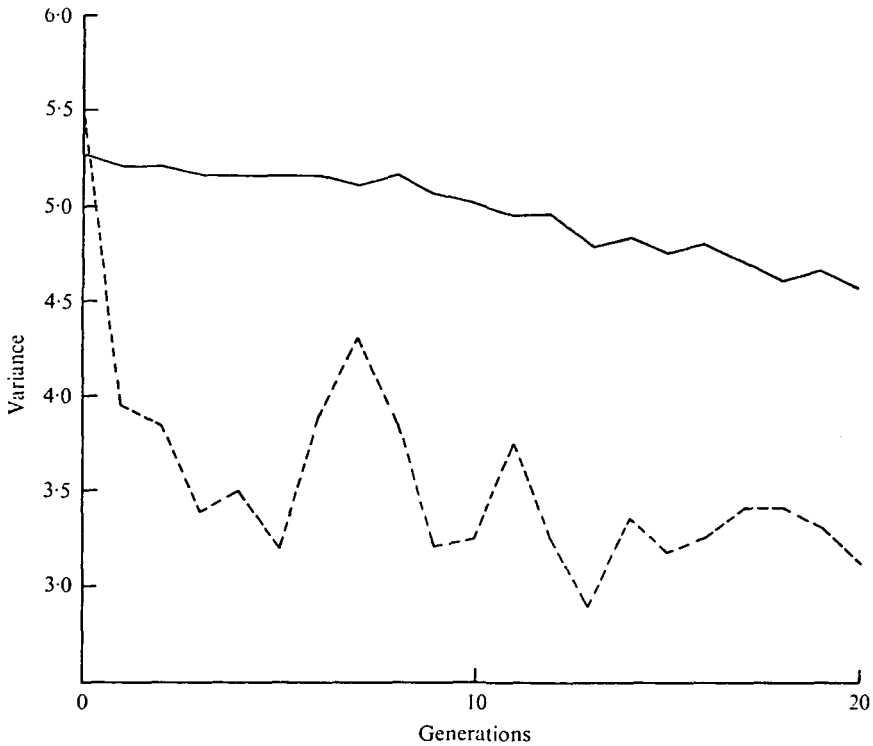

Fig. 2. Genic variance (solid) and genotypic variance (dotted) for the mouse stabilizing rank simulation.

When β is large (that is to say, under strong disruptive selection) there may be no real positive root; the conditions for this to occur are that

$$\left.\begin{aligned} &\text{(i) } \beta > 2H \\ \text{and} \quad &\text{(ii) } V_a/(V_g+V_e) > \min\,(H/\beta, \tfrac{1}{2}-\tfrac{1}{2}(1-2H/\beta)^{\frac{1}{2}}). \end{aligned}\right\} \tag{14}$$

In these circumstances the theoretical treatment based on an effectively infinite number of loci predicts that C_L will increase without bound, though it must also be remembered that the theory is only valid in the presence of linkage (that is to say with $H < \frac{1}{2}$) provided that the individual linkage disequilibria are small at nearly all pairs of loci. It can be concluded that if (14) remains true when H is replaced by $\frac{1}{2}$ then C_L will increase up to its maximum possible value. However, if (14) ceases to hold when H is replaced by $\frac{1}{2}$, it can only be concluded that C_L will become larger than the value predicted from (13) with $H = \frac{1}{2}$. Under disruptive truncation selection it is therefore critical whether β is smaller or larger than 1, which corresponds to whether the proportion of animals selected is larger or smaller

than 45%; if the proportion selected is less than 45% it is predicted that the genotypic variance will increase up to its maximum possible value under random mating, provided that the initial heritability is not too small.

Table 8 shows the observed breakdown of the genotypic variance averaged over

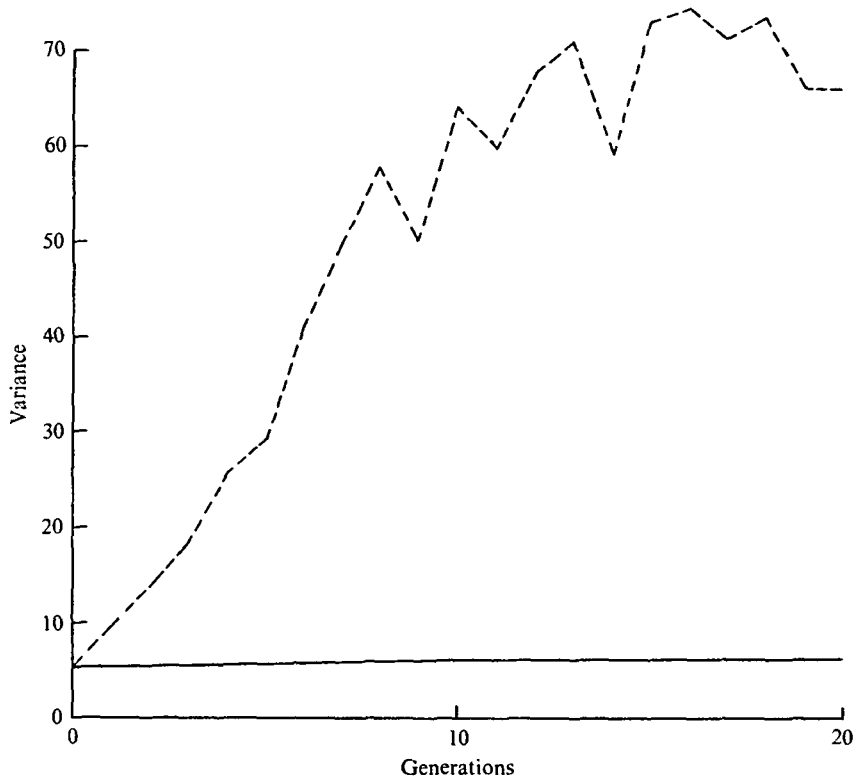

Fig. 3. Genic variance (solid) and genotypic variance (dotted) for the mouse disruptive rank simulation.

the last part of each simulation after an approximate steady state had been reached. As expected C_{HW} is nearly zero except under assortative mating, but C_L departs markedly from zero. The predicted values of C_L were calculated from (13) after substituting the observed value of β from Table 4 and putting $V_a = V_0$, $V_e = 4$ and $H = 0{\cdot}5$ for the mouse and $H = 0{\cdot}18$ for the Drosophila simulations. Agreement with the observed values is good. In particular it will be seen that C_L attains nearly its maximum possible value under disruptive selection with 20% selected but not with 50% selected. In fact in the former case with random mating the system moved to a steady state in which all the gene frequencies were exactly $\frac{1}{2}$ (see Table 6) and in which linkage disequilibrium was so strong that, of the 25 genotypic values between 0 and 24 which can occur, only three were in fact observed (0, 12, 24); the relative frequencies of these values were approximately 0·25, 0·5 and 0·25, though small fluctuations occurred from generation to generation (subject to the first and last frequencies being equal), presumably

due to random variation in the frequencies of different mating types under random mating. The system thus behaves as if it is controlled by a single locus with a gene frequency of $\frac{1}{2}$ with two alleles having values 0 and 12. Under disruptive rank selection with positive assortative mating the system moved to a steady state in which only the two extreme genotypic values, 0 and 24, occurred in exactly equal frequencies.

Table 8. *Components of variance under selection*

Type of selection	Type of simulation	Type of mating	Generations included	V_g	C_{HW}	C_L	Predicted C_L
Stabilizing rank	Mouse	Random	11–20	4·75	−0·04	−1·42	−1·46
Stabilizing rank	Drosophila	Random	21–30	4·41	+0·01	−2·20	−2·18
Stabilizing value	Mouse	Random	11–20	4·95	−0·05	−1·54	−1·54
Stabilizing value	Drosophila	Random	21–30	3·83	+0·01	−1·78	−1·82
Disruptive rank	Mouse	Random	13–20	6·00	−0·06	+64·59	+66*
Disruptive rank	Drosophila	Random	13–20	6·00	−0·26	+62·75	+66*
Disruptive rank	Mouse	Positive	13–20	6·00	+6·00	+132·00	—
Disruptive rank	Drosophila	Positive	13–20	6·00	+6·00	+132·00	—
Disruptive rank	Mouse	Negative	11–20	5·60	−0·28	−1·16	—
Disruptive rank	Drosophila	Negative	11–20	5·50	−0·51	+1·01	—
Disruptive rank (50%)	Mouse	Random	15–30	5·73	+0·02	+10·82	+12·64
Disruptive rank (50%)	Drosophila	Random	20–30	5·84	−0·01	+19·46	>13·29†
Value before selection				5·25	0	0	

* Maximum possible value under random mating.
† Note that $2H < \beta < 1$.

Table 9. *Components of variance and heritability under directional selection*

	Mouse	Drosophila
V_g	3·83	3·92
C_{HW}	0·00	−0·08
C_L	−0·86	−0·77
Selection differential, S	3·70	3·64
Response to selection, R	1·60	1·54
Realized heritability $= R/S$	0·43	0·42
Theoretical heritability $= V_G/(V_G+4)$	0·43	0·43

Under directional selection the gene frequencies went to fixation after 13 generations in the mouse and 16 generations in Drosophila. The gene frequencies moved too rapidly for a steady state to be reached, but negative values of C_L were generated, as expected since β was negative (see equation 10*c*). Table 9 shows the components of variance averaged over the first five generations, before fixation was approached, together with the selection differential, the response to selection and the realized heritability as defined by Falconer (1960) and the theoretical heritability calculated from the genotypic variance. The realized and theoretical

heritabilites are in good agreement, which shows that C_L is to be regarded as a contribution to the additive genetic variance, as postulated elsewhere (Bulmer, 1971*b*).

6. DISCUSSION

The purpose of this paper has been to illustrate and to validate some previous theoretical work on the effect of selection on genetic variability. The main conclusions can be summarized as follows: (1) It is important to distinguish changes in the genotypic variance due to changes in gene frequencies from those due to linkage disequilibrium (and possibly to departures from Hardy–Weinberg equilibrium). To this end it is convenient to partition the genotypic variance into three components representing these three factors (equation 1). (2) Under random mating in finite populations considerable fluctuation may occur in the genotypic variance due mainly to random departures from linkage equilibrium (§3). (3) Equations have been derived previously by classical stability analysis (see (9)) which describe approximately the behaviour of gene frequencies under non-directional selection. These equations are qualitatively correct but they are quantitatively inaccurate since they ignore the effect of linkage disequilibrium (§4). (4) An equation has also been derived previously (see (13)) which predicts the amount of linkage disequilibrium which will be generated under selection. This equation seems to give reasonably accurate predictions (§5).

The above results have applications under both artificial and natural selection. In artificial experiments on stabilizing and disruptive selection the population size is often very small, so that random fluctuations in the genotypic variance due to randomly generated linkage disequilibrium may obscure any changes in the mean level of the variance. For example, Falconer (1957) has reported an experiment on stabilizing selection on abdominal bristle number in *Drosophila melanogaster*, and he concluded that in the second experiment 'the selection of intermediates produced no detectable effect on the variance'. This experiment has been re-analysed, and the results are shown in Table 10. The quantities shown are the variance of the sum (σ_S^2) and of the difference (σ_D^2) of the numbers of bristles on the two sides of the body; σ_D^2 is a measure of environmental variance and ($\sigma_S^2 - \sigma_D^2$) of genetic variance. These variances have been averaged over both sexes, over two replicate lines and over 6 generations in order to eliminate as much random error as possible. There seems to be good evidence of a reduction in the genetic variance in the selected lines. With an initial heritability of 0·5, with 50 % nearest the mean selected, and with a harmonic mean recombination fraction of 0·07 (which is the median prediction for *Drosophila* given by Bulmer, 1974*b*), it is predicted from (13) that the phenotypic variance should be reduced to 60 % of its initial value under selection; this is in good agreement with the observed reduction in the variance.

In artificial experiments it seems likely that the effects of linkage disequilibrium will be more important than the slower and less dramatic effects due to changes in gene frequencies. It must be remembered, however, that many important factors

(such as natural selection against extreme deviants and factors connected with developmental flexibility and genetic homeostasis) have been ignored in the models considered here. The theory presented here should be interpreted as a prediction of what will happen to the genetic variance under the simplest possible model in the absence of any disturbing factors; it is analogous to the standard theory of the response of the genetic mean under directional selection in the absence of disturbing factors.

Under the longer time scale of natural selection it may well be that changes in gene frequencies under the forces described in § 4 will attain a greater importance. In particular it was shown that genetic variability is maintained under disruptive selection ($\beta > 0$) whereas it is eliminated under stabilizing selection ($\beta < 0$). Under some types of selection β may be negative when the phenotypic variance is large and positive when it is small; in particular this is likely to happen if selection for an optimal value is balanced by differential competition between

Table 10. *Re-analysis of Falconer's* (1957) *data on stabilizing selection for abdominal bristle number in* Drosophila melanogaster

	σ_S^2		σ_D^2	
	Control lines	Selected lines	Control lines	Selected lines
Generations 0–6	10·1	9·2	4·8	4·8
Generations 14–19	10·4	6·5	3·9	4·2

individuals through resource specialization. This model is considered in more detail elsewhere (Bulmer, 1974*a*). Under these circumstances variability will be eliminated when the variance is high, but there is an opportunity for fresh genetic variability to be created when the variance is low. Thus the population will tend to evolve by shifting its variance until $\beta \simeq 0$, that is to say until the direct effect of selection on the phenotypic variance is minimized.

It is tentatively concluded that a population will at first respond to a change in the intensity of selection by a change in the linkage disequilibrium component, C_L, but that this will gradually be replaced by a change in the genic variance, V_g. It should be noted that Roughgarden's (1972) model of the evolution of niche width in sexual populations (see also Slatkin, 1970) deals only with changes in variance due to linkage disequilibrium. Furthermore, the results of these authors are valid only in the special case when the heritability is 1 and when there is no linkage, though their treatment can be generalized to an arbitrary heritability without difficulty.

REFERENCES

BULMER, M. G. (1971*a*). The stability of equilibria under selection. *Heredity* **27**, 157–162.
BULMER, M. G. (1971*b*). The effect of selection on genetic variability. *American Naturalist* **105**, 201–211.
BULMER, M. G. (1974*a*). Density-dependent selection and character displacement. *American Naturalist* **108**, 45–58.
BULMER, M. G. (1974*b*). Linkage disequilibrium and genetic variability. *Genet. Res.* **23**, 281–289.
FALCONER, D. S. (1957). Selection for phenotypic intermediates in *Drosophila*. *J. Genet.* **55**, 551–561.
FALCONER, D. S. (1960). *Introduction to Quantitative Genetics*. Edinburgh: Oliver and Boyd.
HILL, W. G. (1974). Disequilibrium among several linked neutral genes in finite populations. II. Variances and covariances of disequilibria. *Theor. Pop. Biol.* **6**, 184–198.
HILL, W. G. & ROBERTSON, A. (1968). Linkage disequilibrium in finite populations. *Theor. Appl. Genet.* **38**, 226–231.
ROUGHGARDEN, J. (1972). Evolution of niche width. *American Naturalist* **106**, 683–718.
SLATKIN, M. (1970). Selection and polygenic characters. *Proc. Nat. Acad. Sci.* **66**, 87–93.