

Dynamics of polygenic variability under stabilizing selection, recombination, and drift

SERGEY GAVRILETS^{1,2} AND ALAN HASTINGS^{1,3,4*}

¹Division of Environmental Studies, University of California, Davis, CA 95616

²N. I. Vavilov Institute of General Genetics, 3, Gubkin St., 117809 GSP-1, Moscow B-333, Russia

³Institute for Theoretical Dynamics, University of California, Davis, CA 95616

⁴Center for Population Biology, University of California, Davis, CA 95616

(Received 10 June, 1994 and in revised form 26 September, 1994)

Summary

We study the transient dynamics of the genotypic variance of an additive trait under stabilizing selection, recombination and random drift. We show how interaction of these factors determines the form and the rates of change of different components of the genotypic variance. Let V_g be the genic variance of the trait and C_L be the contribution of linkage disequilibrium to the genotypic variance. We demonstrate that the dynamics of the system on the plane (V_g, C_L) are typically characterized by a quick approach to a straight line with slow evolution along this line afterwards. We show that the number of loci, n , and the population size, N , affect the expected dynamics of V_g mainly through the ratio N/n . We use our analytical and numerical results in interpreting the published results of artificial stabilizing selection experiments. The analysis suggests that it is drift and not selection that most likely led to the reduction of genetic variability in most of these experiments. Even very strong stabilizing selection only slowly removes polygenic variability from populations.

1. Introduction

Stabilizing selection on polygenic characters has traditionally been considered as one of the core elements of modern evolutionary theory. Numerous models incorporating stabilizing selection have been introduced and analysed (for examples, see references in Weir *et al.* 1988). The effects of stabilizing selection have also been analysed experimentally (e.g. Falconer, 1957; Thoday, 1959; Prout, 1962; Gibson & Thoday, 1963; Scharloo, 1964; Scharloo *et al.* 1967; Tantaway & Tayel, 1970; Bos & Scharloo, 1973; Gibson & Bradley, 1974; Kaufman *et al.* 1977; Soliman, 1982). The main purpose of these studies was the analysis of the influence of stabilizing selection on different components of the phenotypic variability of a polygenic trait and the demonstration that stabilizing selection does decrease these components, as the classical theory predicts, or does not decrease them.

Unfortunately, direct comparison of theory with experiment is complicated. One difficulty is that while

most theoretical studies usually assume weak selection and focus on equilibrium behaviour, experiments use strong selection and display transient behaviour. Recent numerical and theoretical analyses (Gimelfarb, 1992; Gavrillets, 1993; Gavrillets & Hastings, 1993, 1994*a, c*) have shown that properties of multilocus systems under strong stabilizing selection are quite different from those under weak stabilizing selection. In particular, strong stabilizing selection can maintain genetic variability at many loci while weak selection cannot. A necessary condition for the former seems to be sufficiently tight linkage between the loci controlling the traits subject to stabilizing selection. Furthermore, even if the ultimate outcome of selection is complete elimination of genetic variability, this may take much longer than the time span of a typical experiment. At the end of the experiment one often observes only small (if any) reduction in the components of genetical variability. To understand the results of such experiments, one needs theoretical approaches for describing the transient dynamics. Several attempts to analyse the transient dynamics of components of the genetic variance under stabilizing selection have been made. Lewontin (1964) numerically studied a five-locus model of infinite populations. Bulmer (1971,

* Corresponding author: Alan Hastings, Division of Environmental Studies, University of California, Davis, CA 95616. FAX: (916) 752-3350. Phone: (916) 752-8116. e-mail: AMHASTINGS@UCDAVIS.EDU

1974) analytically considered a case of infinite populations with infinite number of unlinked loci. Keightley & Hill (1987, 1988) described an approach that can be used for analysing the case of finite populations with infinite number of unlinked or completely linked loci. Chevalet (1988) considered finite populations with finite number of unlinked loci. Gavrilets & Hastings (1994c) have analysed an infinite population with two arbitrary linked loci. In this paper we develop a new approach for studying transient dynamics of multi-locus systems. Using this approach we consider how long it takes to get to an equilibrium, how important different factors are during the transient period, and what the characteristics of the multilocus system during this period are.

Different factors can influence the transient dynamics under stabilizing selection. We shall concentrate on recombination and random drift. We will neglect mutation. Although it has been argued that mutation can contribute significantly to the response to artificial selection (Hill, 1982), mutation can be safely ignored in situations with high levels of initial genetic variability considered here. Another possible complication is the experimentally observed fact that selection influences not only the genotypic variance but also the microenvironmental component of the phenotypic variance (see references above). This effect has been analysed elsewhere (Gavrilets & Hastings, 1994b) and is neglected in this paper.

The purpose of this paper is to generalize our previous results in two ways to facilitate further comparison with data. The first is to allow the trait to be controlled by an arbitrary number of loci. Although a very small or infinitely large number of loci can sometimes account for observable variability in some quantitative traits, to have an approach without such limitations is, obviously, much more desirable. The second is to incorporate the effects of random genetic drift into the model. The latter is necessary for the following reason. The population size of experimental populations is usually small (Hill & Caballero, 1992). If there are n loci with two alleles each, then the number of possible gametes is 2^n . If $n = 1$ and, hence, there are only two possible gametes, then a population with $N = 100$ individuals may be considered as effectively infinite and a deterministic model can be used to understand population genetics processes. However, the simplification of ignoring stochasticity in a population of $N = 100$ is not valid if the number of loci $n = 10$, when there are more than one thousand possible gametes. This shows that random drift is increasingly important in multilocus systems.

In this paper, we present approximate formulae describing the dynamics of polygenic variability under stabilizing selection, recombination, and drift. To derive these formulae we utilize a heuristic approach for incorporating the effects of random drift, and a dynamical property of the system that has recently been discovered in numerical simulations. We use our

theoretical results in interpreting results of artificial stabilizing selection experiments. Surprisingly, the analysis suggests that it is drift and not selection that most likely led to the reduction of genetic variability in most of these experiments. Our analysis shows that even very strong stabilizing selection is inefficient in removing genetic variability from populations quickly. In multilocus systems subject to stabilizing selection random genetic drift seems to be a much more important factor in reducing genetic variability than selection.

2. Approximating dynamics of the genotypic variance

We consider an additive quantitative trait z determined by n diallelic loci. The standard model for an additive trait is

$$z = \sum_i (a_i + a'_i) + e, \tag{1}$$

where $a_i(a'_i)$ is the contribution of the i th locus from the paternal (maternal) gamete, and e is a random microenvironmental deviation whose distribution is independent of genotype and has zero mean and constant variance E . At the i th locus, let $\alpha_i > 0$ be half the difference between the contributions of two homozygotes to the trait and p_i be the frequency of the higher value allele. Denote by D_{ij} the pairwise linkage disequilibrium between the i th and j th loci. Under random mating the genotypic (additive) variance of the trait $G = V_g + C_L$ is the sum of two components: the genic variance,

$$V_g = \sum_i V_{g,i} \equiv \sum_i 2\alpha_i^2 p_i q_i, \tag{2a}$$

(where $q_i = 1 - p_i$) and the contribution of the linkage disequilibrium,

$$C_L = \sum_{i \neq j} C_{L,ij} \equiv \sum_{i \neq j} \alpha_i \alpha_j D_{ij}, \tag{2b}$$

where $V_{g,i}$ is the contribution of the i th locus to the genic variance and $C_{L,ij}$ is the contribution of the (i, j) th pair of loci to C_L . The phenotypic variance $P = G + E$, where E is the microenvironmental variance. We assume that the population is under stabilizing selection with the optimum value equal to zero. This selection can be described using quadratic or Gaussian fitness functions:

$$w(z) = 1 - sz^2, \tag{3a}$$

$$w(z) = \exp \left[-\frac{z^2}{2V_s} \right], \tag{3b}$$

where the parameters s and V_s measure the strength of selection. As we showed previously (Gavrilets & Hastings, 1994c), the function (3b) with $V_s = E$ also is a good approximation to the double truncation used in experiments (provided selection is at least moderately strong with the proportion selected $< 1/2$).

We shall consider a population of a finite size N with non-overlapping generations. The ultimate outcome of this model is a stochastic equilibrium between factors that eliminate genetic variability (e.g. random drift) and those that increase it (e.g. mutation). (Note that quadratic or Gaussian stabilizing selection on an additive trait does not necessarily eliminate genetic variability (Nagylaki, 1989; Gavrillets & Hastings, 1993, 1994a)). Such stochastic equilibria have been analysed previously (e.g. Bulmer, 1972; Keightley & Hill, 1988; Barton, 1989; Bürger *et al.* 1989). We shall assume that initially the population has a high enough level of genetic variability so that mutation can effectively be neglected. Here we are interested in the transient dynamics of the genotypic variance and its components in the first generations of selection when the genic variance is reduced to, say, one tenth of its initial level.

An exact analysis of the transient dynamics in this model seems impossible. Instead, in this section we present an approximate approach based on several simplifications. We start with a two-locus population of infinite size. We consider quadratic stabilizing selection (3a) assuming that the contributions α_i of the loci to the trait are equal. As was shown elsewhere (Gavrillets & Hastings, 1994c), in this case there are two different stages in the dynamics. During the first stage, which lasts for several generations, the mean value of the trait rapidly approaches the optimum and negative linkage disequilibrium is 'built up'. After that the linkage disequilibrium component C_L is approximately a constant proportion of the genic variance,

$$C_L(t) \approx \theta V_g(t), \tag{4}$$

where t is the generation number, and the dynamics of $V_g(t)$ can be approximated by a single differential equation

$$\frac{dV_g}{dt} = -\frac{SV_g(V_g^{\max} - V_g)}{\bar{w}} \tag{5}$$

(Gavrillets & Hastings, 1994c, eqn 12). Here $S = s(1 + \theta)$, $\bar{w} = 1 - SV_g$ is the mean fitness of the population, and V_g^{\max} is the maximum possible level of the genic variance. The parameter $\theta < 0$ depends on the intensity of selection and on the recombination rate, $\theta = r/s - \sqrt{[(r/s)^2 + 1]}$. Using as our justification extensive numerical simulations (described below), we begin by assuming that equation (4) with some value of θ holds for the multilocus model as well. Later we also give a simple intuitive explanation of why we expect this to be so. Using some additional simplifying assumptions (exactly stated in the Appendix) one can show that in the case of infinite populations with n 'equivalent' loci the dynamics of the genic variance are approximated by

$$\frac{dV_g}{dt} = -\frac{2SV_g(V_g^{\max} - V_g)}{n\bar{w}}. \tag{6}$$

Note that if $n = 2$, this equation reduces to (5). If the loci are different with respect to their contributions to the different components of the genotypic variance, (6) becomes

$$\frac{dV_g}{dt} = -\frac{2SV_g(V_g^{\max} - V_g)}{n_e\bar{w}}, \tag{7}$$

where n_e is an effective number of loci (defined exactly in the Appendix). If the loci are equivalent with respect to their contributions to different components of the genotypic variance, the effective number of loci is equal to the actual number of loci. In general, n_e changes with time. However, numerical simulations suggest (see the Appendix) that during the transient dynamics n_e can be considered as approximately constant.

Now we are in a position to incorporate the effects of the random drift into the model. According to a common method, the expected change in the genic variance as a result of sampling drift can be approximated by a differential equation

$$\frac{dV_g}{dt} = -\frac{1}{2N}V_g, \tag{8}$$

where N is the (effective) population size. This simple approximation has proved to be useful in many problems (e.g. Keightley & Hill, 1987, 1988). Combining this equation with (7) we finally get a single differential equation

$$\frac{dV_g}{dt} = -\frac{2SV_g(V_g^{\max} - V_g)}{n_e\bar{w}} - \frac{1}{2N}V_g, \tag{9a}$$

that together with relationship (4) completely describes the expected dynamics of the genic variance in finite populations under quadratic stabilizing selection and recombination. This equation can be represented as

$$\frac{dV_g}{d(t/N)} = -2\frac{NSV_g(V_g^{\max} - V_g)}{n_e\bar{w}} - \frac{1}{2}V_g. \tag{9b}$$

This means that if one scales the time in the units of the population size (i.e. expresses the time as t/N), then the influence of the population size and the effective number of loci on the behaviour of the genic variance is described by a single parameter N/n_e . This property of the dynamics was first noticed by Chevalet (1988) in numerical simulations. Further evidence of the generality of this effect is presented in the next section.

Equations (4) and (9) represent the main theoretical result of this paper. Equation (9) can be easily integrated. Using the obvious relations $C_L = \theta V_g$, $G = (1 + \theta)V_g$ and $\bar{w} = 1 - SV_g$ one can also use (9) for describing the dynamics of the contribution of linkage disequilibrium, of the genotypic variance and of the mean fitness of the population. One can also consider the relative importance of selection/recombination and drift in the dynamics of multilocus variability.

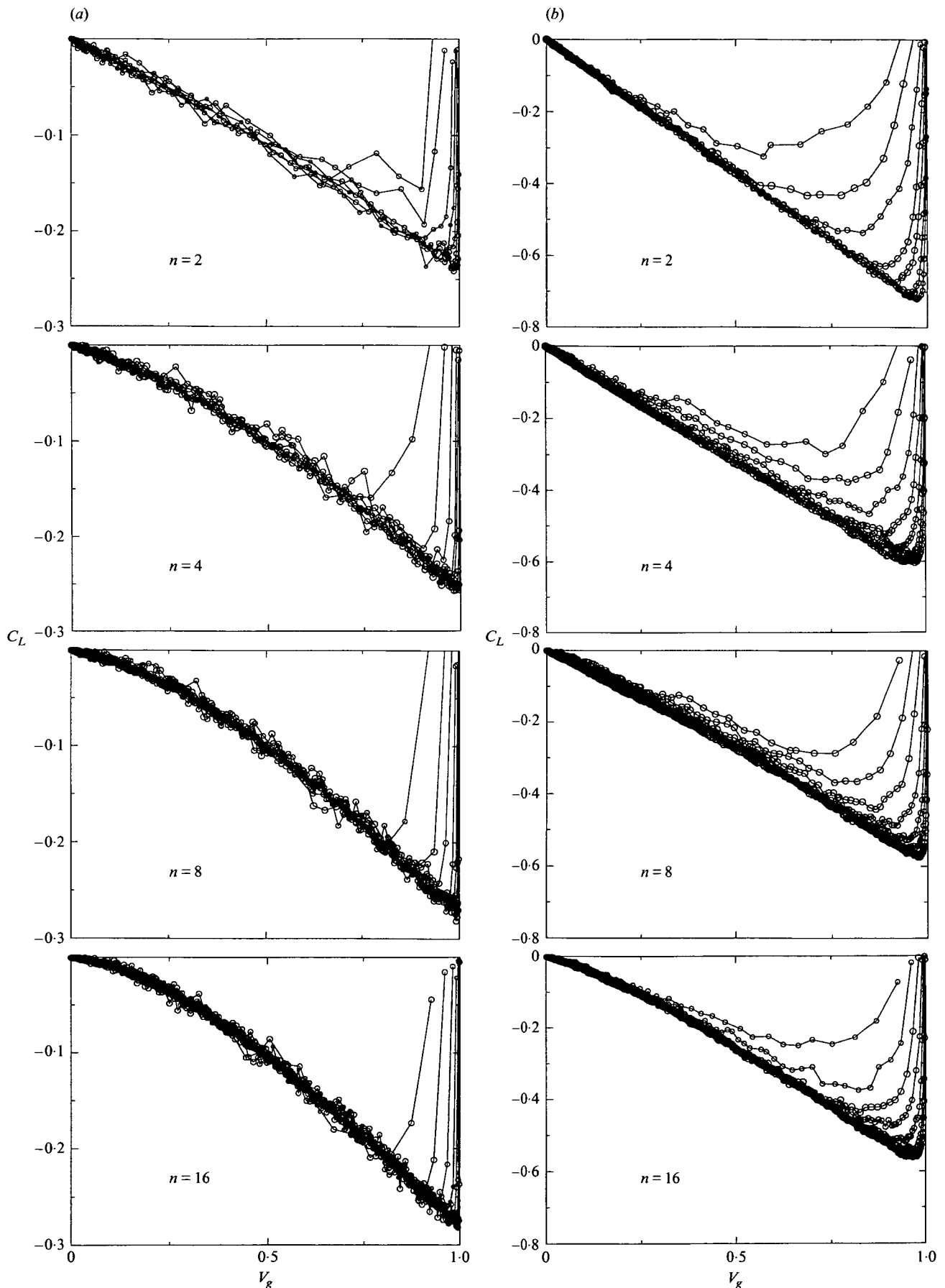


Fig. 1. The dynamics of the components V_g and C_L of the genotypic variance G under double truncation selection. The population size after selection $N_{\text{after}} = 8, 16, 32, 64, 126, 256$ and 512 , the population size before selection was four times its value before selection (the curve corresponding to a smaller population size always lies above). (a) Unlinked loci, $r_h = 0.5$, (b) $r_h = 0.1$.

The comparisons of predictions from equation (9) with simulation show that equation (9) allows to predict the time to change V_g from one value, say V_1 , to another, say V_2 , with an error within 20%–40% range depending on the recombination rate and V_1 and V_2 values. However, equation (9) was derived for quadratic stabilizing selection, which is quite different from the double truncation selection typically used in experiments looking at stabilizing selection. In the next section, we use numerical simulations to check if the properties of the transient dynamics of the genotypic variance under double truncation are similar to those under quadratic stabilizing selection.

3. Properties of the dynamics under double truncation stabilizing selection

In this section, we describe numerical results which show that the two properties of the transient dynamics, namely the approximate constancy of the C_L/V_g ratio and the fact that the influence of the population size and the number of loci on the behaviour of the genic variance is described by a single parameter N/n , hold under double truncation as used in artificial stabilizing selection experiments.

Our numerical results are based on a modified version of the program described in Hastings (1987). We considered $n = 2, 4, 8$ and 16 loci. Population size before selection is $N = 32, 64, 128, 256, 512, 1048$ and 2048. One fourth of individuals (with values of the trait closest to the mean) were selected. Micro-environmental variance E was chosen equal to the maximum possible level of the genic variability (at allele frequencies 0.5). The populations were started at linkage equilibrium (with all $D_{ij} = 0$). The contributions of loci to the trait were equal. Recombination rates between adjacent pairs of loci were identical. We considered four different recombination rates resulting in the mean harmonic recombination rates between pairs of loci, $r_h = 0.50, 0.25, 0.10$ and 0.025. In the two-locus case r_h is just the recombination rate. Number of runs for computing the 'average' dynamics were 100, 90, ..., 40 for $N = 32, 64, \dots, 2048$, respectively.

Figures 1*a, b* illustrate the 'average' dynamics of the components of the genotypic variance on the (V_g, C_L) plane for different numbers of loci, population sizes, and recombination rates. In each case the population starts at a vicinity of the point (1, 0), then it takes several generations to reach a straight line with evolution along this line towards the point (0, 0) afterwards. The time to reach the straight line increases with tighter linkage. The location of this line is controlled by the value of θ . This value is equal to the C_L/V_g ratio at an unstable polymorphic equilibrium with allele frequencies one half. In the two-locus case the value of θ can be derived from exact dynamic equations (Gavrilets & Hastings, 1994*c*). In principle,

one can try to solve the exact dynamic equations for equilibrium values of D_{ij} in the n -locus case. Instead we consider an alternative approach based on Bulmer's formula (Bulmer, 1974, eqn 13). Assuming that selection does not change V_g and using several heuristic approximations, Bulmer derived an equation that relates the equilibrium value of the contribution of linkage disequilibrium, C_L^* , with the genic variance, V_g , the strength of selection, and the mean harmonic recombination rate between pairs of loci, r_h :

$$C_L^*(V_g + C_L^* + E) = \frac{1}{2}\beta(V_g + C_L^*)^2/r_h. \quad (10)$$

Here $\beta = -2z_p\phi(z_p)/P$, where P is the proportion of individuals selected, z is the standard normal deviate corresponding to $(1+P)/2$, and $\phi(z)$ is the standard normal density function. To check this formula, Bulmer compared theoretical predictions with numerical values from several five- and six-locus models having unstable equilibria with allele frequencies equal to one half. His conclusion was that (10) predicts equilibrium values satisfactorily. Figures 1*a, b* can be used to compare the dynamics of V_g in models with different number of loci but with the same value of the mean harmonic recombination rate between pairs of loci. In all cases the proportion of individuals selected was $P = 0.25$. The latter gives $\beta \approx -0.97$. With $V_g = E = 1$ equation (10) produces $C_L^* = -0.29$ and -0.59 for $r_h = 0.5$ and 0.1 respectively. One can see that the trajectories corresponding to different numbers of loci are close enough to each other and that the straight lines along which the populations evolve cross the line $V_g = 1$ at points that are close to the values predicted from (10). The same is true for $r_h = 0.25$ and 0.01 (data are not presented).

Figure 2 shows that the curves representing the dynamics of $V_g(t/N)$ corresponding to different numbers of loci and population sizes but with the same value of the ratio N/n are very close.

4. Biological implications

An immediate effect of stabilizing selection on an additive trait is the buildup of linkage disequilibrium manifested in non-random combinations of alleles in gametes and a reduction of the genotypic variance. This effect was discussed in the classical literature (Fisher, 1930; Mather, 1941), first analysed in numerical simulations by Lewontin (1964) and described using an analytical model by Bulmer (1971, 1974). Stabilizing selection also influences allele frequencies. This process was first studied in the classical papers by Wright (1935) and Robertson (1956) with the conclusion that stabilizing selection invariably moves allele frequencies toward fixation. Subsequent work has shown, however, that stabilizing selection does not necessarily eliminate genetic variability (Gale & Kearsley, 1968; Kearsley & Gale, 1968; Nagylaki, 1989; Gimelfarb, 1992; Gavrilets &

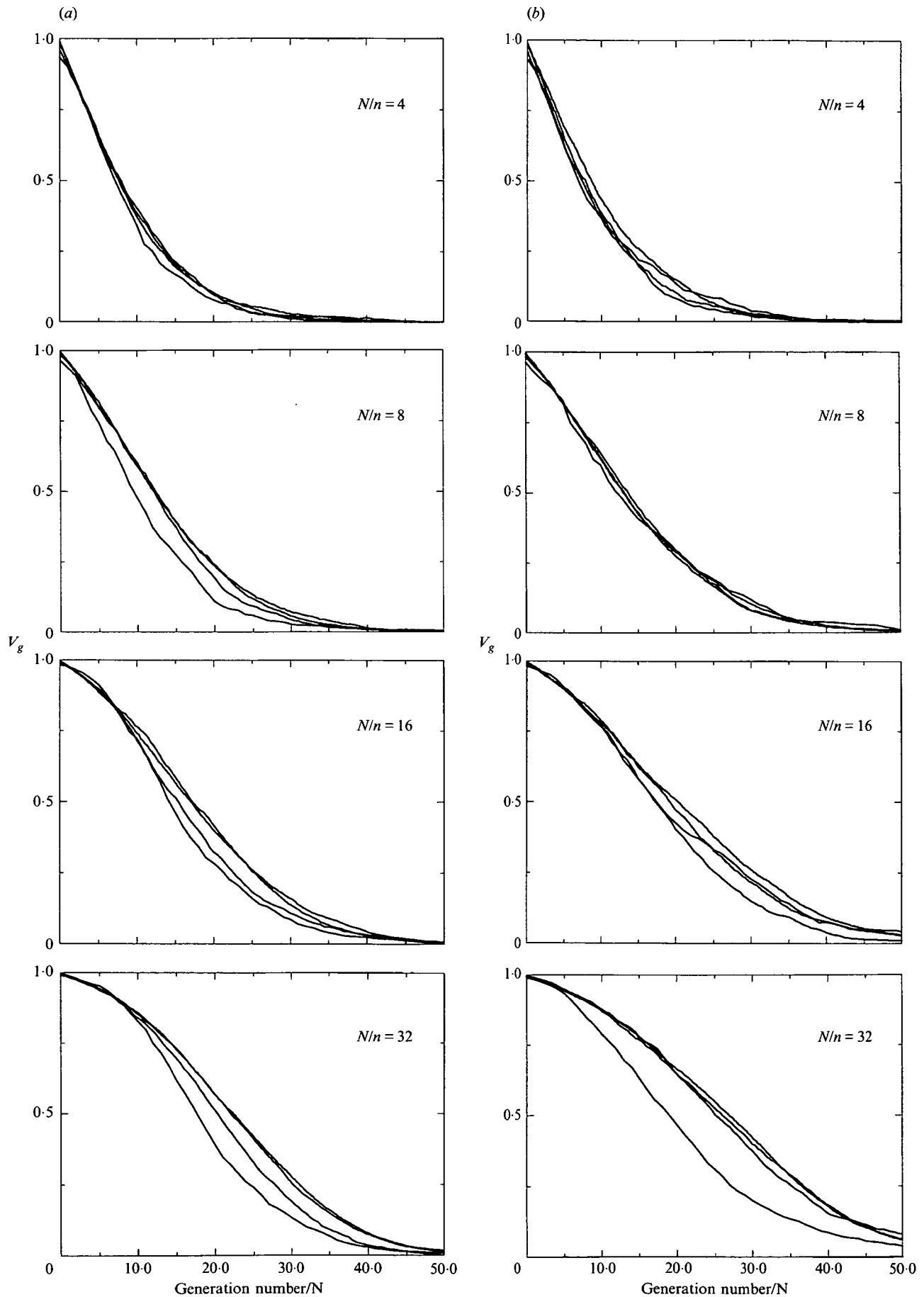


Fig. 2. The dynamics of the genic variance V_g in models with different numbers of loci, $n = 2, 4, 8, 16$, and population sizes, N , but with the same ratio N/n . The time (i.e. the generation number) is scaled in terms of the population size. (a) The case of unlinked loci, $r_h = 0.5$, (b) $r_h = 0.25$.

Hastings, 1993, 1994*a*). Even if the ultimate outcome of stabilizing selection is complete exhaustion of genetic variance, the rate of approach to fixation of alleles can be very slow and decreases with increase in linkage (Lewontin, 1964). Recently we presented an analytical treatment of these transient dynamics in two-locus models (Gavrilets & Hastings, 1994*c*). Here, we have begun the generalization of our approach to the more important case of multilocus populations with a finite size.

(i) *Properties of the transient dynamics*

We have analysed the transient dynamics of two components of the genotypic variance, namely the variance due to disequilibrium, C_L , and the genic variance, V_g . An interesting property of transient dynamics is the approximate constancy of the C_L/V_g ratio in time. This peculiarity allows us to incorporate all effects of linkage disequilibrium on the dynamics of the genic variance in a single parameter θ . A satisfactory approximation of θ is given by Bulmer's (1974) formulae with θ being a function of the mean harmonic recombination rate between pairs of loci and parameters characterizing intensity and form of stabilizing selection. In the two-locus case the approximate constancy of the C_L/V_g ratio in time has been proved analytically (Gavrilets & Hastings, 1994*c*). In the multilocus case this constancy was demonstrated numerically in this paper. An intuitive explanation of this effect for infinite populations is as follows. If selection is absent, any state of the population with linkage equilibrium represents an equilibrium. On the (V_g, C_L) plane the system has a line of equilibria described by the equation $C_L = 0$. If selection is very weak relative to recombination, the dynamics are characterized by quick reduction of disequilibrium with slow evolution of allele frequencies afterwards (Nagylaki, 1976, 1977, 1978, 1992 and 1993; Hoppensteadt, 1976). On the (V_g, C_L) plane such dynamics are represented by a quick movement of trajectories toward the straight line $C_L = 0$ with slow evolution along this line afterwards. If recombination is absent, any state of the population with no genotypic variance represents an equilibrium. On the (V_g, C_L) plane the system has a line of equilibria described by the equation $C_L = -V_g$. If recombination is very weak relative to selection, we expect that the dynamics on the (V_g, C_L) plane are represented by quick movement of trajectories toward the straight line $C_L = -V_g$ with slow evolution along this line afterwards. The fact that in two opposite extreme cases of weak and strong (relative to recombination) selection the dynamics on the (V_g, C_L) plane are qualitatively similar suggests that the same take place in 'intermediate' situations. Numerical simulations both confirm this expectation and demonstrate that the effect is preserved in finite populations.

We have presented further evidence of the generality of the effect discovered by Chevalet (1988) who noted that when time is scaled in the units of the population size N (i.e. expressed as the ratio t/N), the behaviour of genetic variance and its components depends mainly on N/n . Starting from the exact multilocus dynamics equations, we have shown analytically that this effect is present for quadratic stabilizing selection. For the double truncation selection scheme that is used in experiments, we have demonstrated numerically that the dynamics of genetic variance are determined mainly by N/n . This peculiarity of the system allows us to use two-locus dynamics equations in multilocus context.

Combining these properties of the dynamics of multilocus systems discovered in numerical simulations with a standard heuristic approach for incorporating the effects of random drift, we have derived approximate formulae describing the dynamics of polygenic variability under stabilizing selection, recombination, and drift. The resulting formulae give good quantitative agreement for the whole transient dynamics.

(ii) *Relative importance of selection and drift*

One can use the dynamic equation (9) for comparing the relative importance of recombination, selection and random drift in the transient dynamics of the genic variance V_g . A simple way is to compare the magnitudes of two terms in the right-hand part of (9) describing the contributions of selection/recombination and of random drift to the rate of change in V_g . This is done in Fig. 3, where these contributions are given as functions of the existing genic variance for four different numbers of loci ($n = 2, 4, 8$ and 16), a single population size $N = 32$, a single selection strength, $s = 0.1$, and two different recombination values ($r_h = 0.5$ and $r_h = 0.1$). One can see that different components dominate for different values of the genetic variance, V_g . Let us consider the cases of free recombination between two loci and among sixteen loci. We can see (Fig. 3*a*) that the contribution of the random drift to the rate of change of V_g is much bigger than that of selection/recombination in the case of 16 loci, but is much smaller than that of selection/recombination if there are only 2 loci. Thus the dynamics of the genic variance in a population of 32 individuals is effectively neutral, if there are 16 loci underlying the selected trait, and is effectively deterministic, in the case of two loci. If the loci are linked, then the linkage disequilibrium created by stabilizing selection further decreases the contribution of selection/recombination to the rate of change of V_g . As a result, the dynamics of the genic variance is effectively neutral in the case of 8 loci (Fig. 3*b*). The same is true with respect to the four locus model as long as $0.5 < V_g < 1$.

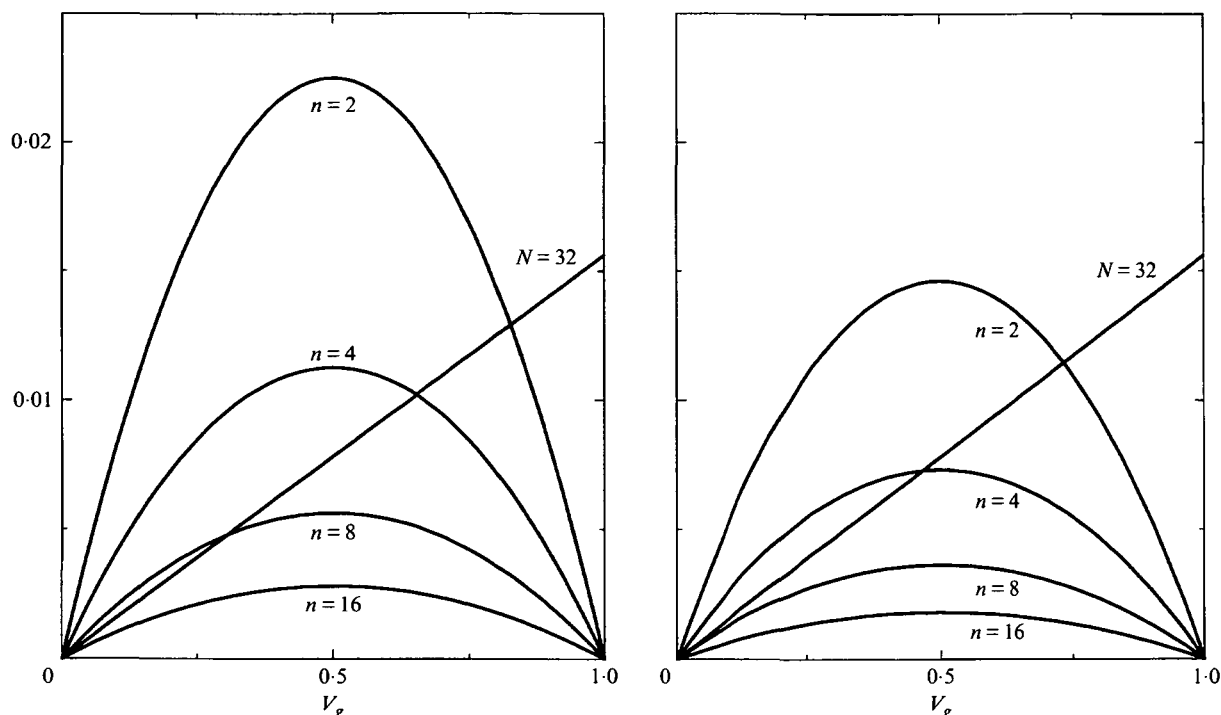


Fig. 3. The magnitudes of the two components of the rate of change of the genic variance (eqn (9)) as function of V_g . The straight line stands for the contribution of random drift for $N = 32$. The curves represent the contributions of selection/recombination for different number of loci, n . (a) The case of unlinked loci, $r_h = 0.5$. (b) $r_h = 0.25$.

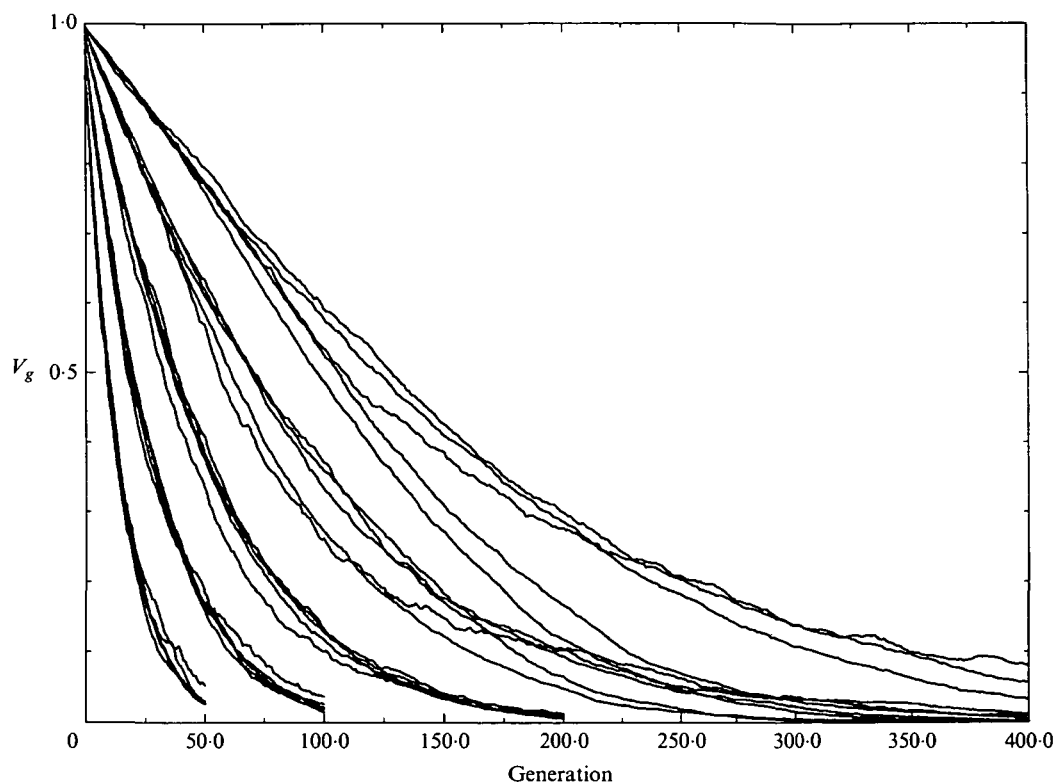


Fig. 4. The dynamics of the genic variance V_g in models with different population sizes after selection, N_{after} , and different recombination rates, r_h . Each set of trajectories corresponds to a specific value of N_{after} ($N_{\text{after}} = 8, 16, 32, 64$ and 128 with the set of trajectories corresponding to a bigger population size always lying above). $n = 16$. The five trajectories within each set correspond to $r_h = 0.5; 0.25; 0.1; 0.025$ and 0.01 .

A different illustration of this effect on the basis of the simulation model with double truncation described above is presented in Fig. 4. This Figure shows the

dynamics of the genic variance for sixteen-locus populations of different sizes. One can see that trajectories corresponding to different recombination

Table 1. Artificial stabilizing selection experiments. See text for discussion of role of drift and selection

Reference	Organism	Population size	Number of generations	Type
Falconer, 1957	<i>Drosophila</i>	40	14, 19	LS
Thoday, 1959	<i>Drosophila</i>	2	42	—
Prout, 1962	<i>Drosophila</i>	≈ 35	40	—
Gibson & Thoday, 1963	<i>Drosophila</i>	8	19	SL
Scharloo, 1964	<i>Drosophila</i>	8	13	SL
Scharloo <i>et al.</i> 1967	<i>Drosophila</i>	8	13, 38	SL
Tantaway & Tayel, 1970	<i>Drosophila</i>	18	10	LS
Bos & Scharloo, 1973	<i>Drosophila</i>	8	25	SL
Gibson & Bradley, 1974	<i>Drosophila</i>	8	40	SL
Kaufman <i>et al.</i> 1977	<i>Tribolium</i>	≈ 100	95	—
Soliman, 1982	<i>Tribolium</i>	30	7	LS

values are very close initially when the random drift dominates. The separation of these trajectories indicates the influence of selection and recombination. For $n = 16$, selection/recombination is practically unimportant for population sizes up to 64 individuals and it takes at least 50 generations for the effects of selection/recombination to show up.

(iii) Artificial stabilizing selection experiments

Of course our results do not lead to predictions of the dynamics of genetic variability in a single experiment where the outcome will depend on parameters (such as allelic contributions and recombination structure) and initial conditions (initial gamete frequencies) that are unknown. However, good quantitative agreement of theory and simulations in simplified models suggests that we may expect to get at least some qualitative understanding of stabilizing selection experiments. Table 1 summarizes the experimental design of most of the relevant experiments. Typically, the foundation stocks (presumably with large level of genetic variability) were derived from wild or large laboratory populations, number of replicates was small (2, 3 or 4), very strong selection (with 15%–30% selected) was applied, and reduction in the additive genetic variance G was measured. We shall exclude from consideration the experiment by Thoday (1959), where the population size was too small and that one by Kaufman *et al.* (1977), where both the population size and the duration of the experiment were large, but the intensity of selection seems to be small due to a specific 'within family' selection scheme used by these authors. We also exclude the experiment by Prout (1962), where the phenotypic variability increased (presumably due to changes in the environmental condition of the entire experiment). The remaining experiments can be roughly divided into two groups: one with (relatively) large population sizes and (relatively) short durations of experiment (marked LS

in Table 1) and another with (relatively) small population sizes and (relatively) long durations of experiment (marked SL in Table 1). The average population size and number of generations over the LS-experiments are $N = 30$ and $T = 12$, while those numbers for SL-experiments are 8 and 25.

The additive genetic variance G is reduced as a consequence of negative linkage disequilibrium generated by selection (Bulmer's effect) and changes in allele frequencies towards fixation caused by selection and random genetic drift. The first effect does not result in reduction of overall genetic variability, and G can be restored if selection ceases. In contrast, shifting the allele frequencies towards fixation results in permanent reduction of overall genetic variability.

If genetic drift is the only factor reducing the genic variance, V_g at the end of an experiment with $N = 30$ and $T = 12$ is expected to be 82% of its initial value, while at the end of an experiment with $N = 8$ and $T = 25$, V_g is expected to be 20% of its initial value. In very large populations the main factor reducing the genic variance is selection. In general, the degree of reduction of V_g resulting from selection only depends on three factors: the initial value of V_g relative to $V_{g,max}$; the number of loci n ; and linkage. Analysis of numerical data (described by Figs 1 and 2) for $N = 512$ shows that if n is not smaller than 4, then the change in V_g during 12 or even 25 generations of selection is not very large. The SL-experiments are expected to result in a drastic reduction of V_g with genetic drift being the major factor. The LS-experiments are expected to result in a small reduction of V_g with genetic drift and selection having approximately equal effect.

Given our analytical and numerical results, we conclude that the population sizes and the durations of these experiments were too small for the effects of stabilizing selection on the genetic variability (as measured by the genic variance) to show up. Some reduction of the additive genetic variance observed in

these experiments should be a result of the decrease in the genic variance caused by random drift (in SL-experiments) and of the build up of negative linkage disequilibrium (in LS-experiments).

(iv) *Evolutionary implications*

Our findings demonstrate that if there are several loci, even very strong stabilizing selection can require a very long time to reduce genetic variability (as measured by the genic variance) significantly. This conclusion has important implications for the maintenance of genetic variability. When thinking about natural systems, it is unreasonable to expect that environmental conditions and populations remain constant over time scales of hundreds of generations. Our results imply that over shorter time scales, strong stabilizing selection, which acts on fitness presumably controlled by a larger number of loci, will not have time to eliminate variability. In this case random drift can be more important in removing genetic variability.

We are grateful to Chuck Coxwell who did some of the numerical simulations. We thank Bob Costantino, Trudy Mackay and two anonymous reviewers for helpful comments on the manuscript. This work was supported by U.S. Public Health Service Grant R01 GM 32130 to A.H.

Appendix A

The exact equation for the dynamics of allele frequencies under quadratic stabilizing selection (3a) is

$$\Delta p_i = \left(\sum_k F_k D_{ik} + \sum_{k \neq h} c_{kh} D_{ikh} \right) / \bar{w}, \tag{A 1}$$

where \bar{w} is the mean fitness of the population. This equation was derived in (Zhivotovsky & Gavrilets, 1992) using results of Ewens & Thompson (1977) on marginal systems and induced fitnesses. Alternatively, (A 1) can be derived using an approach described in (Turelli & Barton, 1990; Barton & Turelli, 1991). In (A 1),

$$c_{kh} = -s\alpha_k \alpha_h, \tag{A 2}$$

$$F_k = -2s\alpha_k \bar{z} + s\alpha_k^2(p_k - q_k), \tag{A 3}$$

D_{ikh} is linkage disequilibrium among three loci, and \bar{z} is the mean value of the trait. We assume that $\bar{z} = 0$ and that linkage disequilibrium of the third order among different loci is zero. The first assumption is satisfied in artificial selection experiments when the optimum is close to the population mean. The second assumption seems to be plausible in the case of stabilizing selection provided a sufficiently high level of variability is still maintained in the population. Using these assumptions one can represent (A 1) in the form

$$\Delta p_i = s(\alpha_i^2 p_i q_i (p_i - q_i)(1 + \theta_i) + \mathcal{R}_i) / \bar{w}. \tag{A 4}$$

Here

$$\theta_i = C_{L,i} / V_{g,i} \tag{A 5}$$

is the ratio of the contributions of the i th locus to C_L and V_g , with $C_{L,i} = \sum_{k,k \neq i} 2\alpha_i \alpha_k D_{ik}$ and $V_{g,i} = 2\alpha_i^2 p_i q_i$. The remainder term $\mathcal{R}_i = \sum_{k,k \neq i} \alpha_k (\bar{z}_k + \bar{z}_i) D_{ik}$, where $\bar{z}_i = \alpha_i(p_i - q_i)$ is the contribution of the i th locus to the mean value of the trait $\bar{z} \equiv \sum \bar{z}_i$. If selection is very weak relative to recombination, linkage disequilibrium can be neglected and (A 4) simplifies to a well-known form

$$\Delta p_i \approx s\alpha_i^2 p_i q_i (p_i - q_i) / \bar{w}. \tag{A 6}$$

The change in the genic variance $V_{g,i}$ contributed by the i th locus is approximated as $\Delta V_{g,i} \approx 2\alpha_i^2 (q_i - p_i) \Delta p_i$ and is

$$\Delta V_{g,i} = -s(2\alpha_i^4 p_i q_i (p_i - q_i)^2 (1 + \theta_i) + 2\alpha_i \bar{z}_i \mathcal{R}_i) / \bar{w}. \tag{A 7}$$

Summing up over all loci and using the equality $(p_i - q_i)^2 = 1 - 4p_i q_i$, we get

$$\Delta V_g = -\frac{2s \sum V_{g,i} (V_{g,i}^{\max} - V_{g,i})(1 + \theta_i) + \mathcal{R}}{\bar{w}}, \tag{A 8}$$

where $V_{g,i}^{\max} = \alpha_i^2/2$ is the maximum possible level of the genic variability at the i th locus (at allele frequencies equal to one half), and $\mathcal{R} = s \sum 2\alpha_i \bar{z}_i \mathcal{R}_i$. Below we shall show that the remainder term \mathcal{R} can be neglected in comparison with the other sum. Using this and substituting the differential ratio dV_g/dt for the difference ratio $\Delta V_g/\Delta t$, one can approximate (A 8) as

$$\frac{dV_g}{dt} = -\frac{2s}{\bar{w}} \sum V_{g,i} (V_{g,i}^{\max} - V_{g,i})(1 + \theta_i). \tag{A 9}$$

If $V_{g,i} \approx V_{g,j}$, $V_{g,i}^{\max} \approx V_{g,j}^{\max}$ and $\theta_i \approx \theta_j$ for all pairs of loci i and j , we get equation

$$\frac{dV_g}{dt} = -\frac{2s V_g (V_g^{\max} - V_g)(1 + \theta)}{n \bar{w}}, \tag{A 10}$$

that approximates the dynamics of the genic variance in the multilocus case. Gavrilets & Hastings (1994c) have shown that in the two-locus case with equal loci equation (A 10) adequately describes dynamics after some short time interval. If the loci are different with respect to their contributions to the components of the genotypic variance, one can introduce the effective number of loci n_e

$$n_e = \frac{V_g (V_g^{\max} - V_g)(1 + \theta)}{\sum V_{g,i} (V_{g,i}^{\max} - V_{g,i})(1 + \theta_i)}, \tag{A 11}$$

and represent (A 9) in the form (7). In general, n_e changes with time and can be both smaller or bigger than n . However, numerical iterations of (A 6) for different number of loci have suggested that during the transient dynamics (when the genic variance is reduced to about one tenth of its initial level) the

changes in n_e are not very big. For example, for $n = 4$, depending on the initial conditions and the differences in the locus contributions, n_e stays between 3 and 6, while for $n = 10$, n_e stays between 7 and 14.

To show that the term \mathcal{R} can be neglected, let us represent it as

$$\mathcal{R} = s \sum_{i,j,k} \bar{z}_i \bar{z}_j C_{L,ik}, \quad (\text{A } 12)$$

where $C_{L,ik} = 2\alpha_i \alpha_k D_{ik}$ is the contribution of the (ik)th pair of the loci to C_L and the sum is over all different loci i, j and k . To derive (A 12), we have used the equality $\bar{z}_i = -\sum_{k, k \neq i} \bar{z}_k$ that is valid provided $\bar{z} = 0$. We expect that all $C_{L,ik}$ values are small and negative, and \bar{z}_i values are both positive and negative and small. This suggests that positive and negative terms of the form $\bar{z}_i \bar{z}_j C_{L,ik}$ will cancel each other and the resulting sum will be smaller in absolute value than the first sum in the right-hand side of (A 8) where all elements have the same sign.

References

- Barton, N. H. (1989). The divergence of a polygenic systems subject to stabilizing selection, mutation and drift. *Genetical Research* **54**, 59–77.
- Barton, N. H. & Turelli, M. (1991). Natural and sexual selection on many loci. *Genetics* **127**, 229–255.
- Bos, M. & Scharloo, W. (1973). The effects of disruptive and stabilizing selection on body size in *Drosophila melanogaster*. *Genetics* **75**, 679–693.
- Bulmer, M. (1971). The effects of selection on genetic variability. *American Naturalist* **105**, 210–211.
- Bulmer, M. (1972). The genetic variability of polygenic characters under optimizing selection, mutation and drift. *Genetical Research* **19**, 17–25.
- Bulmer, M. (1974). Linkage disequilibrium and genetic variability. *Genetical Research* **23**, 281–289.
- Bulmer, M. (1980). *The Mathematical Theory of Quantitative Genetics*. Oxford: Clarendon Press.
- Bürger, R., Wagner, G. P. & Stettinger, F. (1989). How much heritable variation can be maintained in finite populations by mutation-selection balance? *Evolution* **43**, 1748–1766.
- Chevalet, C. (1988). Control of genetic drift in selected populations. In *Proceedings of the Second International Conference on Quantitative Genetics* (ed. B. S. Weir, E. J. Eisen, M. M. Godman and G. Namkong), pp. 379–394. Sunderland, MA: Sinauer.
- Ewens, W. E. & Thomson, G. (1977). Properties of equilibria in multilocus genetic systems. *Genetics* **87**, 809–819.
- Fisher, R. A. (1930). *The Genetical Theory of Natural Selection*. Oxford: Clarendon Press.
- Falconer, D. S. (1957). Selection for phenotypic intermediates in *Drosophila*. *Journal of Genetics* **55**, 551–561.
- Gale, J. S. & Kearsley, M. J. (1968). Stable equilibria under stabilizing selection in the absence of dominance. *Heredity* **23**, 553–561.
- Gavrilets, S. (1993). Equilibria in an epistatic viability model under arbitrary strength of selection. *Journal of Mathematical Biology* **31**, 397–410.
- Gavrilets, S. & Hastings, A. (1993). Maintenance of genetic variability under strong stabilizing selection: a two locus model. *Genetics* **134**, 377–386.
- Gavrilets, S. & Hastings, A. (1994a). Maintenance of multilocus variability under strong stabilizing selection. *Journal of Mathematical Biology* **32**, 287–302.
- Gavrilets, S. & Hastings, A. (1994b). A quantitative genetic model for selection on developmental noise. *Evolution* (in the press).
- Gavrilets, S. & Hastings, A. (1994c). Dynamics of genetic variability in two-locus models of stabilizing selection. *Genetics* **138**, 519–532.
- Gibson, J. B. & Thoday, M. M. (1963). Effects of disruptive selection. VIII. Imposed quasi-random mating. *Heredity* **18**, 513–524.
- Gibson, J. B. & Bradley, B. P. (1974). Stabilizing selection in constant and fluctuating environments. *Heredity* **33**, 293–302.
- Gimelfarb, A. (1992). Pleiotropy and multilocus polymorphism. *Genetics* **130**, 223–227.
- Hastings, A. (1987). Substitution rates under stabilizing selection. *Genetics* **116**, 479–486.
- Hill, W. G. (1982). Predictions of response to artificial selection from new mutations. *Genetical Research* **40**, 255–278.
- Hill, W. G. & Caballero, A. (1992). Artificial selection experiments. *Annual Review of Ecology and Systematics* **23**, 287–310.
- Hoppenstead, F. C. (1976). A slow selection analysis of two locus, two allele traits. *Theoretical Population Biology* **9**, 68–81.
- Kaufman, P. F., Enfield, F. D. & Comstock, R. E. (1977). Stabilizing selection for pupa weight in *Tribolium castaneum*. *Genetics* **87**, 327–341.
- Kearsley, M. J. & Gale, J. S. (1968). Stabilizing selection in the absence of dominance: an additional note. *Heredity* **23**, 617–620.
- Keightley, P. D. & Hill, W. G. (1987). Directional selection and variation in finite populations. *Genetics* **117**, 573–582.
- Keightley, P. D. & Hill, W. G. (1988). Quantitative genetic variability maintained by mutation-stabilizing selection balance in finite populations. *Genetical Research* **52**, 33–43.
- Lewontin, R. C. (1964). The interaction of selection and linkage. II. Optimal model. *Genetics* **50**, 757–782.
- Mather, K. (1941). Variation and selection of polygenic characters. *Journal of Genetics* **41**, 159–193.
- Nagylaki, T. (1976). The evolution of one- and two-locus systems. *Genetics* **83**, 583–600.
- Nagylaki, T. (1977). The evolution of one- and two-locus systems. II. *Genetics* **85**, 347–354.
- Nagylaki, T. (1978). *Selection in one- and two-Locus Systems*. Berlin: Springer.
- Nagylaki, T. (1989). The maintenance of genetic variability in two-locus models of stabilizing selection. *Genetics* **122**, 235–248.
- Nagylaki, T. (1992). *Introduction to Theoretical Population Biology*. Berlin, Heidelberg, N.Y.: Springer-Verlag.
- Nagylaki, T. (1993). The evolution of multilocus systems under weak selection. *Genetics* **134**, 627–647.
- Prout, T. (1962). The effect of stabilizing selection on the time of development in *Drosophila melanogaster*. *Genetical Research* **3**, 364–382.
- Robertson, A. (1956). The effect of selection against extreme deviants based on deviation or on homozygosity. *Journal of Genetics* **54**, 236–248.
- Scharloo, W. (1964). The effect of disruptive and stabilizing selection on the expression of a *cubitus interruptus* mutant in *Drosophila*. *Genetics* **50**, 553–562.
- Scharloo, W., Hoogmoed, M. S. & Ter Kuile, A. (1967). Stabilizing and disruptive selection on a mutant character in *Drosophila*. I. The phenotypic variance and its components. *Genetics* **56**, 709–726.
- Soliman, M. H. (1982). Directional and stabilizing selection

- for developmental time and correlated response in reproductive fitness in *Tribolium castaneum*. *Theoretical and Applied Genetics* **63**, 111–116.
- Tantaway, A. O. & Tayel, A. A. (1970). Studies on natural population of *Drosophila*. X. Effects of disruptive and stabilizing selection on wing length and correlated response in *Drosophila melanogaster*. *Genetics* **65**, 121–132.
- Thoday, J. M. (1959). Effects of disruptive selection. I. Genetic flexibility. *Heredity* **13**, 187–203.
- Turelli, M. & Barton, N. H. (1990). Dynamics of polygenic characters under selection. *Theoretical Population Biology* **38**, 1–57.
- Weir, B. S., Eisen, E. J., Godman, M. M. & Namkong, G. (eds) (1988). *Proceedings of the Second International Conference on Quantitative Genetics*. Sunderland, MA: Sinauer.
- Wright, S. (1935). Evolution in populations in approximate equilibrium. *Journal of Genetics* **30**, 257–266.
- Zhivotovsky, L. & Gavrilets, S. (1992). Quantitative variability and multilocus polymorphism under epistatic selection. *Theoretical Population Biology* **42**, 254–283.