Genet. Res., Camb. (1998), **71**, *pp.* 161–170. *With 2 figures.* Printed in the United Kingdom © 1998 Cambridge University Press

Linkage and the maintenance of variation for quantitative traits by mutation–selection balance: an infinitesimal model

ENRIQUE SANTIAGO*
Departamento de Biología Funcional, Universidad de Oviedo, 33071 Oviedo, Spain

(*Received 1 August 1997 and in revised form 24 October 1997 and 9 December 1997*)

Summary

The infinitesimal model is extended to cover linkage in finite populations. General equations to predict the dynamics of the genetic variation under the joint effects of mutation, selection and drift are derived. Under truncation and stabilizing selection, the quadratic equations for the asymptotic genetic variance (V_G) are respectively

$$V_G^2(1+kS)+V_G(V_e-2N_e V_m)-2N_e V_m V_e=0$$

and

$$V_G^2(1+S)+V_G(V_e+\gamma-2N_e V_m)-2N_e V_m(V_e+\gamma)=0,$$

where N_e is the effective population size, V_m is the mutational variance, V_e is the environmental variance, γ is the parameter that measures the spread of fitness around the optimum under stabilizing selection, k is equal to $i(i-x)$ where i is the selection intensity and x is the cut-off point under truncation selection. The term S is a function of the number of chromosomes (v) and the average chromosome length (l):

$$S \approx \frac{v-1}{v}+\frac{1}{vl}\ln(N_e l+1).$$

These predictions are accurate when compared with results of simulations of small populations unless the number of genes is small. The infinitesimal model reduces to the continuum of alleles model if there is no recombination between homologous chromosomes.

1. Introduction

The classical infinitesimal model (Fisher, 1918) assumes a very large number of unlinked loci with additive effects. The reduction of genetic variance under truncation or stabilizing selection in this model is due entirely to the build-up of negative linkage disequilibrium among the loci (Bulmer, 1971). In an infinite population with no mutational input of variation, the independent segregation of the loci removes the disequilibrium very quickly and the genetic variance decreases to its asymptotic value in a few generations (see Bulmer, 1980). In the presence of linkage, the reduction of the variance is expected to be larger, as the disequilibrium is removed by recombination at a slow rate. Bulmer (1974, 1976) obtained a general formula to predict the asymptotic genetic variance in a model with an arbitrary number of linked loci. The formula is a function of the harmonic mean of the recombination frequencies between all the possible pairs of loci. When the number of loci tends to infinity, the recombination frequencies between adjacent loci tend to 0 and the harmonic mean of the recombination frequencies and the asymptotic genetic variance also drop to 0, independently of how much variability was present at the beginning. Linkage without any input of variation in an infinitesimal model ultimately leads to the hiding of all the original

* e-mail: esr@sauron.quimica.uniovi.es.

https://doi.org/10.1017/S0016672398003231 Published online by Cambridge University Press

genetic variance behind negative covariances between neighbouring loci when selection is acting. A non-zero equilibrium variance is expected if the reduction by selection is balanced by mutation. Additionally, finite population size interacts with selection making the genes drift in the population for not too long and eroding the creation of strong negative linkage disequilibrium.

Although linkage is a natural extension of the infinitesimal model, there is no simple prediction of the evolution of the genetic variance under linkage. The behaviour of the model under truncation selection was explored by simulation and numerical computation by Keightley & Hill (1987). They concluded that the infinitesimal model is a poor predictor for complete linkage and the effect of linkage is eliminated by a few crossovers per chromosome. A great deal of work has been undertaken on models dealing with a finite number of linked loci. Under these models, the amount of expressed genetic variance at equilibrium is dependent on the rate of loss of variation at individual loci. Most of them are concerned with the problem of the maintenance of genetic variation in infinite populations (Lande, 1976; Turelli, 1984; Burger, 1989, 1993). The general conclusion seems to be that linkage disequilibrium can be ignored in discussing the amount of variability maintained by the balance between selection and mutation unless linkage is tight. However, short-term changes in the genetic variance would be mainly dependent on the dynamics of linkage disequilibrium (Barton & Turelli, 1989).

In this paper we extend the infinitesimal model to cover linkage in a genetic system with an arbitrary number of chromosomes. We follow a derivation similar to that used by Bulmer (1980, p. 158) for an arbitrary number of linked loci. General equations are obtained to predict the evolution of the genetic variance in finite populations under selection, the asymptotic variance at equilibrium and the distribution of the genetic variability along the chromosome. Following Turelli & Barton (1994), linkage disequilibria of order three and higher can be neglected even under tight linkage and the normal distribution of breeding values is assumed over the whole selection process.

2. The general infinitesimal model

(i) *No mutation and infinite population size*

Imagine a monoecious diploid population of an infinite number of individuals, which mate at random in non-overlapping generations. Every individual is considered to be made up of two gametes (paternal and maternal) coming from its parents. Each gamete has a haploid number of chromosomes v, each l morgans long. It is assumed that there is no interference at crossing over. Therefore, the relation between the recombination fraction r and the map distance x is given by the mapping function of Haldane (1919):

$$r = \tfrac{1}{2}(1 - e^{-2x}).$$

The quantitative trait is determined by a very large number of additive loci with equal effects. The density of the loci on the chromosomes is proportional to the map distance.

Genetic and environmental effects are independent and normally distributed at each generation; therefore the phenotypic variance before selection in generation 0 (V_{p_0}) is the sum of the genetic variance in generation 0 (V_{G0}) and the environmental variance (V_e), which is assumed to be constant over generations. There is no linkage disequilibrium assumed at this generation. The gametic variance is $V_{g0} = V_{G0}/2$, as the breeding value of every individual is the sum of the effects of both gametes. Phenotypic selection (e.g. truncation, stabilizing) is carried out on the trait every generation. After the first round of selection, the remaining genetic variance of selected individuals in generation 0 is

$$V_{G0}* = V_{G0}\, G_0.$$

The parameter G_0 is generally dependent on the amount and distribution of genetic variation before selection as well as on the particular selective method. It will remain indeterminate for the moment, but approximations for particular selective modes will be given below.

In the infinitesimal model, the change in the genetic variance by selection from V_{G0} to $V_{G0}*$ is entirely due to covariances between pairs of loci. As all the loci are considered to contribute to the genetic variance by the same amount, these covariances are identical between any pair of loci. There are two different types of covariances: covariances between loci within gametes (i.e. linkage disequilibrium) versus covariances between loci from the two different gametes of the same individual. Each type of covariance accounts for one-half of the change in the genetic variance as, if one takes two genes at random from the same individual, the probabilities of being in the same gamete or in different gametes are nearly identical if the number of loci is very large. Therefore, the reduction of the gametic variance from V_{g0} to $V_{g0}*$ is one-half of the reduction of the genetic variance of individuals:

$$V_{g0}* = V_{g0}\left(1 - \frac{1-G_0}{2}\right).$$

Covariances between loci from different gametes generate an overall covariance between the genetic values of both gametes (Cov_{g0}) within individuals, which accounts for the other half of the reduction in the genetic variance:

https://doi.org/10.1017/S0016672398003231 Published online by Cambridge University Press

$$V_{G0}* = 2V_{g0}\,G_0 = 2V_{g0}\left(1-\frac{1-G_0}{2}\right)+2\,\mathrm{Cov}_{g0};$$

therefore

$$\mathrm{Cov}_{g0} = -V_{g0}\frac{1-G_0}{2},$$

Now, we will focus on the same problem from the point of view of particular loci. If there are n loci determining the character, the genic variance contributed by a haploid locus (v) will be a fraction $1/n$ of the variance of gametes ($v = V_{g0}/n$). In the infinitesimal model, the change in the genic variance (excluding linkage disequilibrium) of a particular locus is negligible because selection acts on phenotypes and the phenotypic variance is much larger than v. Therefore, we will consider v to be constant over all the selection process.

After the first round of selection in generation 0, the covariance generated by selection is evenly distributed among all pairs of loci, either on the same or different chromosomes. As there are n^2 different pairs of loci from different gametes, the covariance between any pair of loci (cov_0) is a fraction $1/n^2$ of the total covariance between both gametes (Cov_{g0}):

$$\mathrm{cov}_0 = \mathrm{Cov}_{g0}\frac{1}{n^2} = -V_{g0}\frac{1-G_0}{2}\frac{1}{n^2}.$$

In the initial generation, recombination does not remove covariances between loci because their values are identical in coupling and repulsion. Therefore in the next generation (generation 1), the gametic variance before selection is the same as $V_{g0}*$:

$$V_{g1} = V_{g0}* = nv+n(n-1)\,\mathrm{cov}_0 \approx V_{g0}-V_{g0}\frac{1-G_0}{2},$$

and the genetic variance between individuals before selection is $V_{G1} = 2V_{g1}$.

Selection in generation 1 changes the genetic variance from V_{G1} to $V_{G1}*$ because new covariances (cov_1) are generated between pairs of loci. Again, the new covariances are assumed to be the same for any pair of loci in coupling or repulsion as the number of loci is very large and the marginal genetic variance (i.e. including linkage disequilibrium with all the other loci) is approximately the same for any locus. This is not strictly true because the different sites on a linear chromosome are not equivalent, but the assumption is not far away from the real expectation, as will be shown in Section 3(i). Thus

$$\mathrm{cov}_1 = \mathrm{Cov}_{g1}\frac{1}{n^2} = -V_{g1}\frac{1-G_1}{2}\frac{1}{n^2},$$

where $G_1 = V_{G1}*/V_{G1}$.

After selection in generation 1, the total covariance between any pair of loci within gametes (i.e. the linkage disequilibrium) will be approximately the sum of the old and the new covariances ($\mathrm{cov}_0+\mathrm{cov}_1$) as their values remain very small if the number of loci is large, but the total covariance between loci from different gametes in the same individual will be only cov_1 as segregation and random combination of gametes eliminate the old covariances between gametes.

Recombination removes a proportion of the old covariances between loci within the same gamete, and the remaining old covariance between two particular loci, which have a recombination frequency r, is

$$\mathrm{cov}_0' = \mathrm{cov}_0(1-r).$$

The new covariance (i.e. cov_1) is unaffected by recombination as its value is the same in coupling and repulsion. For two loci at random, the probability of being on different chromosomes is $(v-1)/v$; in this case $r = 1/2$. If both loci are on the same chromosome (probability $= 1/v$), the recombination fraction will depend on the genetic distance between them. The distribution of distances between two random sites in the same chromosome follows a triangular distribution with the highest frequency corresponding to closely linked loci and the lowest frequency corresponding to the maximum distance l. In other words, if one imagines 10^3 loci on a linear chromosome, there are 999 different pairs of adjacent loci and the frequency decreases linearly with distance to only one pair of loci, which have a distance of l morgans between them. Integrating over all the possible pairs of loci on the same or different chromosomes, the expected value of the old covariance, which still remains after recombination in meiosis of parents of generation 1, will be

$$E[\mathrm{cov}_0'] = \mathrm{cov}_0\left[\frac{v-1}{v}\frac{1}{2}+\frac{1}{v}\frac{1}{l/2}\int_0^l\frac{l-x}{l}\frac{1+e^{-2x}}{2}dx\right]$$

$$= \mathrm{cov}_0 f_1 = -V_{g0}\frac{1-G_0}{2}\frac{1}{n^2}f_1.$$

The term f_1 is the expected value of the remaining proportion of the covariance cov_0 between two random loci after one round of recombination. It could also be understood as the average remaining proportion of cov_0 over all the pairs of loci.

Therefore, in generation 2, the genetic variance between gametes before selection will be the sum of the genic variances of all the n haploid loci (excluding linkage disequilibrium) and the effect of linkage disequilibrium accumulated in generations 1 and 0:

$$V_{g2} = nv+n(n-1)\left(\mathrm{cov}_1+E[\mathrm{cov}_0']\right)$$

$$\approx V_{g0}-V_{g1}\frac{1-G_1}{2}-V_{g0}\frac{1-G_0}{2}f_1.$$

https://doi.org/10.1017/S0016672398003231 Published online by Cambridge University Press

Selection in generation 2 generates new covariances between loci. Again, the value of these covariances is considered to be the same between any pair of loci:

$$\mathrm{cov}_2 = \mathrm{Cov}_{g2}\frac{1}{n^2} = -V_{g2}\frac{1-G_2}{2}\frac{1}{n^2}.$$

The total covariance between any pair of loci within gametes will be the sum of the old covariances and the new covariance ($\mathrm{cov}_0' + \mathrm{cov}_1 + \mathrm{cov}_2$) and the covariance between loci from different gametes in the same individual will be cov_2.

After recombination with frequency r between two loci, the remaining proportion in generation 2 of the old covariance generated by selection in generation 0 is

$$\mathrm{cov}_0'' = \mathrm{cov}_0'(1-r) = \mathrm{cov}_0(1-r)^2,$$

and the remaining proportion of the covariance generated in the previous generation by selection is

$$\mathrm{cov}_1' = \mathrm{cov}_1(1-r).$$

Averaging over all possible pairs of loci, the expected value of both covariances over all the pairs of loci will be

$$E[\mathrm{cov}_0''] = \mathrm{cov}_0\left(\frac{v-1}{v}\frac{1}{2}^2 + \frac{1}{v}\frac{1}{l/2}\int_0^l \frac{l-x}{l}\frac{1+e^{-2x}}{2}^2 dx\right)$$

$$= \mathrm{cov}_0 f_2 = -V_{g0}\frac{1-G_0}{2}\frac{1}{n^2}f_2,$$

$$E[\mathrm{cov}_1'] = \mathrm{cov}_1\left(\frac{v-1}{v}\frac{1}{2} + \frac{1}{v}\frac{1}{l/2}\int_0^l \frac{l-x}{l}\frac{1+e^{-2x}}{2}dx\right)$$

$$= \mathrm{cov}_1 f_1 = -V_{g1}\frac{1-G_1}{2}\frac{1}{n^2}f_1.$$

The new covariance (i.e. cov_2) is unaffected by recombination as its value is the same in coupling and repulsion.

The generalization of this process is obvious. For any generation $t+1$ the genetic variance of gametes before selection is equal to the original genetic variance plus the cumulative effect of linkage disequilibrium, that is the sum of all the remaining fractions of covariances between all the pairs of loci of the same gamete that originated in previous generations:

$$V_{g(t+1)} = nv + \sum_{i=0}^{t} n(n-1)\,\mathrm{cov}_i f_{(t-i)}$$

$$\approx V_{g0} - \sum_{i=0}^{t} V_{gi}\frac{1-G_i}{2}f_{(t-i)}.$$

The same equation holds for the genetic variance of individuals because it is twice the variance of gametes before selection:

$$V_{G(t+1)} \approx V_{G0} - \sum_{i=0}^{t} V_{Gi}\frac{1-G_i}{2}f_{(t-i)}, \qquad (1)$$

where f_y represents the remaining proportion of a covariance y generations after it was generated:

$$f_y = \frac{v-1}{v}\frac{1}{2}^y + \frac{1}{v}\frac{1}{l/2}\int_0^l \frac{l-x}{l}\frac{1+e^{-2x}}{2}^y dx.$$

This parameter is equivalent to the expected proportion of the donor genome in a backcrossing programme using a marker gene when the location of the marker is unknown. The problem was analysed by Stam & Zeven (1981). They found that the solution given by Hanson (1959) for the size of the donor chromosome segment, which is adjacent to a central marker, is an excellent approximation to f_y:

$$f_y \approx \frac{v-1}{v}\frac{1}{2}^y + 2\frac{1-e^{-yl/2}}{vly}, \quad \text{with } f_0 = 1.$$

The iterative formula (1) allows the computation of the genetic variance at any generation. After an infinite number of generations, all the genetic variance will be depleted under directional or stabilizing selection unless the number of chromosomes is infinite. But it will be shown that the reduction is quite slow after an initial drop during a few generations.

(ii) *Mutation and finite population size*

To clarify some of the arguments, the previous model will be completed assuming a very large number of loci that eventually can mutate, thereby increasing or decreasing the value of the character by one unit with identical probability. In a finite population, only some of the loci are segregating and the probability of mutation of a previously segregating locus is negligible. As the number of segregating loci is considered to be large, the density of variability over all the chromosome regions is assumed to be approximately uniform if the population size is not small.

Both the genetic variance at particular loci and the linkage disequilibrium are expected to be reduced by drift at a rate of $1/(2N_e)$ per generation, where N_e is the effective population size. Mutation introduces new variation in non-segregating loci, increasing the genetic variance of individuals by V_m each generation (and by $V_m/2$ for gametes). The inclusion of both processes in the recursive equation (1) is straightforward. The genetic variance at any generation is the sum of three terms:

$$V_{G(t+1)} = V_{G0}\left(1-\frac{1}{2N_e}\right)^{t+1} + \sum_{i=0}^{t+1} V_m\left(1-\frac{1}{2N_e}\right)^{i}$$

$$-\sum_{i=0}^{t} V_{Gi}\frac{1-G_i}{2}f_{(t-1)}\left(1-\frac{1}{2N_e}\right)^{t-i+1}. \qquad (2)$$

The first term is the remaining fraction of the original genic variance (the first term in (1)) after the decay of variability caused by drift during t generations (no linkage disequilibrium is assumed by this term).

https://doi.org/10.1017/S0016672398003231 Published online by Cambridge University Press

The second term is the new genic variance (assuming linkage equilibrium) originated by mutation over generations. This term is also reduced by drift each generation, but as the variance is generated across generations, the expected value at any generation is the sum of the geometric series

$$V_m + V_m(1-1/2N_e) + V_m(1-1/2N_e)^2 + \cdots + V_m(1-1/2N_e)^{t+1},$$

where the first element of the series is the variation generated by mutation in the actual generation, the second element is the remaining fraction of the variation generated by mutation one generation ago, and so on. Together, the first and the second terms account for the standing genetic variance. The third term corresponds to the cumulative covariances, i.e. linkage disequilibrium, generated over generations (the second term in (1)). These covariances are also reduced by drift at a rate $1/(2N_e)$ every generation. After an infinite number of generations, the parameters V_G and G will reach asymptotic values and all the original genetic variance (the first term in (2)) and the mutational variability generated in the initial generations (when the G_i values were different due to the initial change of V_G) will disappear due to drift. Equation (2) reduces to:

$$V_G = 2N_e V_m - V_G(1-G)S, \tag{3}$$

where

$$S = \frac{1}{2}\sum_{i=0}^{\infty} f_i \left(1-\frac{1}{2N_e}\right)^i.$$

The first term in (3) is the standing genetic variance (i.e. assuming linkage equilibrium) and the second term accounts for the cumulative effect of linkage disequilibrium in finite populations. After some algebra and applying Hanson's simplification and the approximation $(1-1/(2N_e))^i \approx e^{-i/(2N_e)}$, S reduces to

$$S \approx \frac{v-1}{v} + \frac{1}{vl}\ln(N_e l + 1).$$

This equation should be used for values of l between 0 and 3 morgans, as Hanson's approximation is acceptable only within this range. For l values higher than 3, the exact solution given by Stam & Zeven (1981) should be applied. S is equal to 1 when the number of chromosomes is infinite, which is the case of free recombination. When there is a single chromosome, S converges towards N_e as l tends to 0.

It is worth noting that the mutational variance at equilibrium is equal to a fraction $1/(2N_e)$ of the total standing genic variance. From (3),

$$V_m = \frac{1}{2N_e}[V_G + V_G(1-G)S].$$

The two terms of the sum between brackets are the expressed genetic variance and the genetic variance hidden behind linkage disequilibrium respectively.

(iii) *The distribution of the genetic variance along the chromosome*

In the previous sections, we assumed that all the loci contributed the same marginal genetic variance (including linkage disequilibrium). This is not strictly true when loci are on a linear chromosome. The effect of linkage is expected to be smaller in the telomere than in the middle of the chromosome as the number of genes that are closely linked to a particular locus at the tip of the telomere is a half of that of a locus in the middle of the chromosome. In this section we will approximate the expected value of the ratio between the marginal genetic variance and the genic variance contributed by loci along the chromosome. As will be shown in the next section, the average differences between sites is small and it is not expected that the assumption of equivalence of all the loci will cause a significant deviation in the predictions.

Consider a particular locus placed at a distance p morgans from one end of the chromosome; it will be at a distance $l-p$ from the other end. Its contribution to the overall genetic variance is reduced by the covariances with the other $n-1$ loci in the same gamete generated by selection over generations. The population is at mutation–selection–drift equilibrium, and therefore the new covariances generated by selection between the locus and any other locus are considered to be constant over generations. The expected value in the current generation of the remaining fraction of the covariances originated y generations ago is

$$\mathrm{cov}\,(1-1/(2N_e))^y f'_y,$$

where $(1-1/(2N_e))^y$ is the reduction due to drift and f'_y represents the proportion of the remaining association of the locus in p with all the other loci in the same gamete. Again, the meaning of f'_y is equivalent to the proportion of the donor genome in a back-crossing process, but here the position of the locus of reference is fixed:

$$f'_y = \frac{v-1}{v}\frac{1}{2y} + \frac{1}{vl} \times \int_0^p \left(\frac{1+e^{-2x}}{2}\right)^y dx + \int_0^{l-p} \left(\frac{1+e^{-2x}}{2}\right)^y dx.$$

The asymptotic contribution of the locus at p to the gametic variation will be equal to the contribution if there were no linkage disequilibrium, that is the amount $N_e V_m/n$ which is identical for any locus, plus the sum of the disequilibrium covariances with all the

https://doi.org/10.1017/S0016672398003231 Published online by Cambridge University Press

other $n-1$ loci in the same gamete that were generated by selection during the previous infinite generations:

$$\frac{N_e V_m}{n}+(n-1)\sum_{i=0}^{\infty}\mathrm{cov}\left(1-\frac{1}{2N_e}\right)^i f'_i$$
$$=\frac{N_e V_m}{n}-(n-1)\frac{V_g}{n(n-1)}\frac{1-G}{2}\sum_{i=0}^{\infty}\left(1-\frac{1}{2N_e}\right)^i f'_i.$$

Dividing the equation by $N_e V_m/n$, the ratio between the marginal genetic variance (i.e. including linkage disequilibrium) and the standing genic variance (i.e. without linkage disequilibrium) contributed by the locus is obtained:

$$1-\frac{V_g}{N_e V_m}\frac{1-G}{2}\sum_{i=0}^{\infty}\left(1-\frac{1}{2N_e}\right)^i f'_i.$$

After some algebra, the equation can be simplified to the expression

$$1-\frac{V_g}{N_e V_m}\frac{1-G}{2}\left(\frac{v-1}{v}2+\frac{1}{vl}\left(\int_0^p F(x)\,dx+\int_0^{l-p}F(x)\,dx\right)\right), \quad (4)$$

where

$$F(x)=\frac{1}{\ln\left(\frac{1+e^{-2x}}{2}\right)+\ln\left(1-\frac{1}{2N_e}\right)}.$$

3. Application to particular selection methods

(i) *Truncation selection*

The main factor in the evolution of the genetic variance is the value of G. If this parameter is known for a given method of selection, predictive formulae for the genetic variance can be obtained under the infinitesimal model using the equations that were derived above. We shall now consider truncation selection in which individuals with the largest phenotypic values are selected. Every generation, a constant proportion of individuals is selected. Assuming normality in the distribution of genetic values at equilibrium, the proportion of the remaining genetic variance after selection is: $G=1-kh^2$ (Bulmer, 1971), where h^2 is the asymptotic heritability (i.e. $h^2=V_G/V_p$), $k=i(i-x)$, i is the selection intensity and x is the truncation point in standard deviation units. Although this assumption is not strictly valid, large deviations from normality are not expected under the infinitesimal model even with tight linkage (Turelli & Barton, 1994).

Substituting G into (3) and after some algebra, the quadratic equation to predict the asymptotic genetic variance at equilibrium is obtained:

$$V_G^2(1+kS)+V_G(V_e-2N_e V_m)-2N_e V_m V_e=0.$$

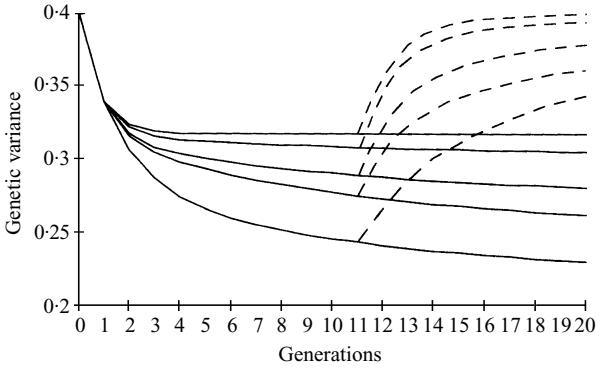

Fig. 1. Evolution of the genetic variance under truncation selection with initial heritability = 0·4 and selected proportion = 1/4 for different chromosome numbers. Population size is infinite and no mutation is assumed. From top to bottom: infinitely many, 10, 3, 2 and 1 chromosomes. The size of the chromosomes is always 1 morgan. The dashed lines represent the evolution of the variance after relaxation of selection in generation 11.

This equation for V_G is equivalent to the partial solutions given by Keightley & Hill (1987) for free recombination (their equation 10) and complete linkage (their equation 11). The equation also matches their numerical computations for a variety of values of linkage given in their figure 3.

The evolution of the genetic variance during the initial generations of a selection process in an infinite population can be predicted from (1) using the corresponding values of G_i for consecutive generations:

$$G_i=1-k_i\frac{V_{Gi}}{V_{Gi}+V_e},$$

where k_i allows for changes in the selection intensity over generations and V_{Gi} is the observed genetic variance in generation i. Fig. 1 shows an example of the evolution of the genetic variance for 20 generations in populations that were initially at linkage equilibrium. Selection was carried out at a constant intensity during the first half of the process and relaxed in the second half. The recovery of the original genetic variance, after relaxation, was slower than the reduction under selection, especially when the number of chromosomes was small.

Under truncation selection, the distribution of the genetic variance along the chromosome can be computed by numerical integration of (4). Fig. 2 gives the corresponding density of variance for some combinations of different parameters. Although the distribution is U-shaped with the maxima at the tips of the telomeric regions, most of the variability is evenly distributed along the chromosome. This validates the assumption of the uniform distribution of the new covariances between all the possible pairs

https://doi.org/10.1017/S0016672398003231 Published online by Cambridge University Press

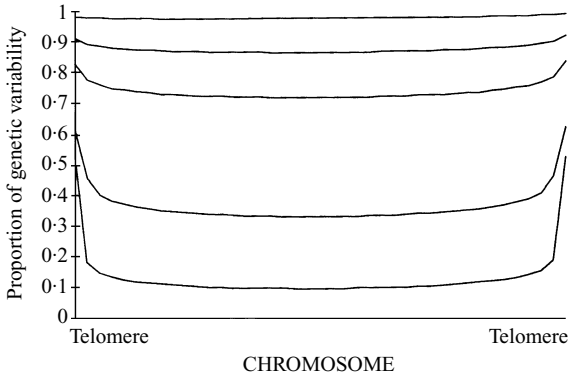

Fig. 2. Distribution of the genetic variation in a genome of one single chromosome at equilibrium under truncation selection. $l = 1$ morgan, $V_m = 10^{-3}$, selected proportion $= 1/2$ and population sizes of 10, 40, 100, 10^3 and 10^6 individuals (from top to bottom). The variation is given as a proportion of the variance with free recombination.

of loci and, therefore, the marginal genetic variance is approximately the same for all the loci.

(ii) *Stabilizing selection*

Most authors investigating the maintenance of genetic variation by stabilizing selection in natural populations have used the fitness function of Haldane (1954):

$$w(x) = \exp\ -\frac{1}{2}\frac{(x-\theta)^2}{\gamma}\ .$$

The formula gives the fitness of an individual with phenotypic value x when the optimum value is θ, where the parameter γ gives the strength of the stabilizing selection around the optimum. In an infinitesimal model with normal distribution of genetic values and environmental deviations, the remaining genetic variance after selection at any generation is

$$V_G* = V_G G = V_G\ \ 1-\frac{V_G}{V_G+V_S}$$

(see Bulmer, 1980), where $V_S = V_e+\gamma$. Applying the same derivation as before, the quadratic equation of the asymptotic genetic variance under mutation is obtained:

$$V_G^2(1+S)+V_G(V_S-2N_e V_m)-2N_e V_m V_S = 0. \qquad (5)$$

Our model assumes normality in the distribution of genetic values. The continuum of alleles model (Kimura, 1965) considers the more severe assumption of normality of the distribution of allele effects at individual loci. Although the reduction of variance under this model is entirely due to the erosion of variability within loci, the value of G would be the same in both models and (5) should be able to predict the genetic variance under Kimura's model: a number v of unlinked loci is equivalent to an infinitesimal model with v chromosomes each 0 morgans long. In this case, $S = (v-1+N_e)/v$ and linkage disequilibrium within chromosomes is not recovered by recombination. Substituting S in (5), for very large population size and $V_S \gg vV_m$, the equation of Kimura for the variance at equilibrium, when all the loci are equivalent, is obtained:

$$V_G \approx vV_m+\sqrt{(v^2V_m^2+2vV_m V_S)} \approx \sqrt{(2vV_m V_S)}.$$

4. Test of predictions

The accuracy of the predictions was checked by Monte Carlo simulations. Random mating populations with N diploid individuals were simulated. The quantitative trait was controlled by n additive loci evenly distributed among v linear chromosomes. All the loci were segregating for two alleles all the time. When one of the alleles was eventually lost at a locus, a new allele was introduced by mutation. The new copy of the allele either increased or decreased the score of the genotype by one unit, with probability 0·5 each. To calculate the effect of drift, an additional set of n neutral loci was allocated alternating with the selected loci. The neutral loci were also always segregating for two alleles as single copies of new neutral alleles were introduced at loci which become monomorphic. Selection was carried out on the population for thousands of generations so that the genetic system reached equilibrium. Then 10^4 additional generations were simulated and the asymptotic genetic variability was computed as the average value of the observed genetic variance during these generations. The standing genic variance was also computed as the average genetic variance if there were no linkage disequilibrium, i.e. as the sum of the individual genetic variance of the separate loci. The effective population size was approximated using the neutral loci in a similar way to that used by Charlesworth *et al.* (1993): the sum of the heterozygosity contributed by each neutral mutation during all the generations from mutation to fixation or loss was computed and then averaged (H) over all the mutational events. The effective population size was approximated as $N_e = (NH)/2$.

The results of the truncation selection are shown in Table 1. The simulations are in good agreement with predictions, especially if we consider the strong linkage disequilibrium generated by selection: generally, the amount of genic variance is twice or more the value of the observed genetic variance. The agreement is remarkably good with complete linkage. This observation is in conflict with that of Keightley & Hill (1987) who attributed a poor predictive accuracy to the infinitesimal model with complete linkage. Their observation could be a consequence of using the census size of the population instead of the effective

https://doi.org/10.1017/S0016672398003231 Published online by Cambridge University Press

Table 1. *Observed* (V_G) *and predicted values of the genetic variance under truncation selection for three proportions of selected individuals*

v	l	Proportion	N_e	V_m	V_G *le*	V_G	Predicted
1	0	0·25	8·3	13·25	228·1	94·8	98·1
		0·50	11·9	10·76	261·6	103·2	101·2
		0·75	20·9	8·28	353·1	117·1	111·2
1	1	0·25	16·7	10·04	367·0	179·3	195·0
		0·50	26·0	7·88	441·9	211·8	231·8
		0·75	43·5	6·05	547·7	275·6	299·5
10	1	0·25	35·1	6·56	584·8	322·4	322·8
		0·50	50·6	5·43	645·2	383·2	390·5
		0·75	70·5	4·66	709·7	487·9	489·9

V_G *le* is the observed genic variance, V_m is the mutational variance, v is the number of chromosomes and l the chromosome size. For all the combinations of parameters, the environmental variance was 400, the number of loci was 4800 and the number of reproducers was 100.

Table 2. *Observed* (V_G) *and predicted values of the genetic variance under stabilizing selection*

								Predicted	
v	l	n	N_e	V_e	V_m/V_e ($10^3\times$)	V_G *le*	V_G	Infin.	SHC
1	0	4800	126·2	400	5·28	538·9	160·2	159·2	524·7
		800	170·7	67·2	4·75	108·4	28·3	26·4	97·2
		80	191·3	6·76	4·82	11·3	2·78	2·72	5·31
		8	166·8	0·676	8·88	0·374	0·163	0·377	0·157
1	1	4800	193·9	400	4·60	702·4	526·3	540·5	718·1
		800	197·3	67·2	4·63	115·4	89·9	92·7	107·8
		80	209·8	6·76	5·12	9·85	8·05	10·59	5·86
		8	203·8	0·676	9·91	0·192	0·182	1·725	0·178
10	1	4800	194·8	400	4·55	712·6	651·7	643·0	693·4
		800	200·2	67·4	4·51	118·5	108·0	109·9	106·6
		80	183·8	6·76	5·18	9·70	9·08	11·62	5·61

Predictions were made using the infinitesimal model with linkage and the stochastic house of cards model (SHC). V_G *le* is the observed genic variance, V_m is the mutational variance, V_e is the environmental variance, n is the number of loci, v is the number of chromosomes and l the chromosome size. The number of reproducers was 200 and the intensity of stabilizing selection was $\gamma = 20V_e$.

population size in their predictions. With strong linkage, a large reduction in the effective population size is expected under selection (Nordborg *et al.*, 1996).

Table 2 shows the results of stabilizing selection. Predictions based on the stochastic house of cards model were also made using the equation

$$V_{G(SHC)} = \frac{2V_m N_e}{1+\alpha^2 N_e/V_S},$$

(Keightley & Hill, 1988; Burger *et al.*, 1989; Houle, 1989; Barton, 1989). In our table, $\alpha^2 = 1$. Although this model does not consider the effect of linkage, it has been argued repeatedly that linkage disequilibrium can be neglected (see Bulmer, 1989). To test the flexibility of the model to deviations from the infinitesimal effects of the genes, genetic systems with different numbers of selective loci were simulated. For the combinations of parameters considered in the simulations, predictions of the infinitesimal model overestimated the genetic variance when only eight selective loci were simulated, but the predictions were good for 80, 800 and 4800 loci. On the contrary, predictions of the stochastic house of cards model were good for eight loci per chromosome and overestimated the genetic variance when more than 80 loci were considered.

5. Discussion

In the infinitesimal model, the dynamics of the variation is determined entirely by linkage disequilibrium and drift. Under truncation or stabilizing selection, there is a reduction in the genetic variance

https://doi.org/10.1017/S0016672398003231 Published online by Cambridge University Press

due to negative linkage disequilibrium between pairs of loci (see Bulmer, 1980), but the gene frequencies at particular loci are unaffected by selection. This imaginary model predicts fairly well the genetic variance at equilibrium even under complete linkage if the gene effects are small enough, i.e. if the evolution of the gene frequencies is mainly dependent on drift. This makes the infinitesimal model inappropriate to predict the genetic variance at equilibrium in large populations, although changes in the variance for a few generations could probably be predicted even in very large populations as linkage disequilibrium is the main factor in the short term (Barton & Turelli, 1989).

The observed consistency of predictions and simulations is in conflict with the conclusions of Keightley & Hill (1987). They attributed the lack of accuracy of their predictions using the infinitesimal model to deviations in the distribution of the genetic values from normality under close linkage. But the disagreement is probably due to the fact that they did not consider the effective population size in making predictions. Directional selection reduces the effective size of populations (Santiago & Caballero, 1995) especially under close linkage (Nordborg *et al*., 1996). When the effective size is considered, the predictions are good, suggesting that the effect of the deviations from normality is small under truncation selection. This observation is in agreement with the prediction of quasi-normality of Turelli & Barton (1994) in the infinitesimal limit of a multilocus model.

The parameter S, which gives the effect of linkage on the genetic variability, is also a function of the effective population size. The effect of linkage becomes larger as the effective size increases and is small for very small population sizes. Keightley & Hill (1987) found the same effect in small populations under truncation selection. The reason is that drift makes the genes pass quickly through the population and, therefore, there is no time to build up strong linkage disequilibrium between them. This effect is amplified when the genes have a significant effect on the trait, as selection reduces the time to fixation or loss of the alleles.

Our model can be used to give some insights into the question of the evolution of the genetic variance under artificial selection. This is a matter of interest as response is dependent on the amount of genetic variability during the selection process. The evolution of the genetic variance depends on a complex of interactions of selection, linkage and drift as pointed out by Hospital & Chevalet (1996). Moreover, mutation can play an important role even in short-term selection processes. Our Fig. 1 suggests that, when the number of chromosomes is larger than 10, linkage disequilibrium between selected genes is not responsible for a large reduction in genetic variance during the first 20 generations of selection in comparison with free recombination. Although it is expected that all the original genetic variance of an infinite population will disappear with linkage, its decline is quite slow after a few generations. For the span of an artificial selection process in most farm species, the amount of genetic variance hidden behind linkage disequilibrium is probably small. The main effect of linkage on response is likely to be due to the reduction of the effective population size. A sign of this effect can be seen in Table 1 where large reductions of the asymptotic N_e values are produced even with 10 chromosomes.

Although the number of chromosomes and the chromosome lengths are parameters in the predictive equations of the genetic variability, changes in the distribution of chromosome lengths have a very small influence if the size of the whole genome is constant. Under truncation selection (25% selected), a mutational input of variation $0{\cdot}01V_e$ and an effective population size of 100 individuals, one single chromosome 3 morgans long yields the same genetic variability at equilibrium as 24 chromosomes each 0·125 morgans long ($h^2 = 0{\cdot}52$). The main parameter seems to be the total size of the genome. A similar observation was pointed out by Hill (1993) for the variance in the size of the donor genome in backcrossing programs. The similarity is not unexpected as the problem of the fraction of the original chromosome associated with a marker locus is a starting point of both derivations.

I thank W. G. Hill, P. D. Keightley and A. Caballero for helpful comments. This work was supported by grant PB95-0909-C02-02 from the Ministerio of Educacion y Cultura (Spain).

References

Barton, N. H. (1989). Divergence of a polygenic system subject to stabilizing selection, mutation and drift. *Genetical Research* **54**, 59–77.

Barton, N. H. & Turelli, M. (1989). Evolutionary quantitative genetics: How little do we know? *Annual Review of Genetics* **23**, 337–370.

Bulmer, M. G. (1971). The effect of selection on genetic variability. *American Naturalist* **105**, 201–211.

Bulmer, M. G. (1974). Linkage disequilibrium and genetic variability. *Genetical Research* **23**, 281–289.

Bulmer, M. G. (1976). The effect of selection on genetic variability: a simulation study. *Genetical Research* **28**, 101–117.

Bulmer, M. G. (1980). *The Mathematical Theory of Quantitative Genetics*. Oxford: Clarendon Press.

Bulmer, M. G. (1989). Maintenance of genetic variability by mutation–selection balance: a child's guide through the jungle. *Genome* **31**, 761–767.

Burger, R. (1989). Linkage and the maintenance of heritable variation by mutation–selection balance. *Genetics* **121**, 175–184.

Burger, R. (1993). Predictions of the dynamics of a polygenic character under directional selection. *Journal of Theoretical Biology* **162**, 487–513.

Burger, 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.

Charlesworth, B., Morgan, M. T. & Charlesworth, D. (1993). The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**, 1289–1303.

Fisher, R. A. (1918). The correlation between relatives on the supposition of mendelian inheritance. *Transactions of the Royal Society of Edinburgh* **52**, 399–433.

Haldane, J. B. S. (1919). The combination of linkage values and the calculation of the distances between the loci of linked factors. *Journal of Genetics* **8**, 299–309.

Haldane, J. B. S. (1954). The measurement of natural selection. In *Proceedings of the IX International Congress on Genetics Caryologia*, pp. 480–487.

Hanson, W. D. (1959). Early generation analysis of lengths of heterozygous chromosome segments around a locus held heterozygous with backcrossing or selfing. *Genetics* **44**, 833–837.

Hill, W. G. (1993). Variation in genetic composition in backcrossing programs. *Journal of Heredity* **84**, 212–213.

Hospital, F. & Chevalet, C. (1996). Interactions of selection, linkage and drift in the dynamics of polygenic characters. *Genetical Research* **67**, 77–87.

Houle, D. (1989). The maintenance of polygenic variation in finite populations. *Evolution* **43**, 1767–1780.

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 variation maintained by mutation-stabilizing selection balance in finite populations. *Genetical Research* **52**, 33–43.

Kimura, M. (1965). A stochastic model concerning the maintenance of genetic variability in quantitative characters. *Proceedings of the National Academy of Sciences of the USA* **54**, 731–736.

Lande, R. (1976). The maintenance of genetic variability by mutation in a polygenic character with linked loci. *Genetical Research* **26**, 221–235.

Nordborg, M., Charlesworth, B. & Charlesworth, D. (1996). The effect of recombination on background selection. *Genetical Research* **67**, 159–174.

Santiago, E. & Caballero, A. (1995). Effective size of populations under selection. *Genetics* **139**, 1013–1030.

Stam, P. & Zeven, A. C. (1981). The theoretical proportion of the donor genome in near-isogenic lines of self-fertilizers bred by backcrossing. *Euphytica* **30**, 227–238.

Turelli, M. (1984). Heritable genetic variation via mutation–selection balance: Lerch's zeta meets the abdominal bristle. *Theoretical Population Biology* **25**, 138–193.

Turelli, M. & Barton, N. H. (1994). Genetic and statistical analyses of strong selection on polygenic traits: what, me normal? *Genetics* **138**, 913–941.