Genetic variance and fixation probabilities at quantitative trait loci in mutation–selection balance

PETER D. KEIGHTLEY

Institute of Cell Animal and Population Biology, University of Edinburgh, West Mains Road, Edinburgh EH9 3JT Scotland

(*Received 26 October 1990 and in revised form 27 February 1991*)

Summary

Genetic variance and fixation probabilities are evaluated for a model of a quantitative trait at a balance between mutation, selection and drift in which many alleles can segregate at each locus. If the distribution of effects of new mutant alleles is such that mutations are unconditionally deleterious, as might be the case in natural populations, genetic variance maintained is proportional to the number of mutations occurring in the genome per generation, but is independent of the number of loci at which they appear. If selectively advantageous alleles can occur these tend to interfere to a greater extent with each others' fixation and increasing mutation rate leads to a decrease in the genetic variance as a fraction of the variance maintained in the absence of selection. Fixation probabilities of new mutant alleles approach that for neutral alleles with increasing mutation rate at a locus irrespective of their effects on fitness. The additive genetic variance contributed by the locus may appear to be 'decoupled' from the fixation rate of mutant alleles.

1. Introduction

For the most part, the loci which cause genetic variation in quantitative characters have not been amendable to direct analysis. There is little information on, for example, the number of alleles segregating, their distribution of effects and degrees of dominance, or their pleiotropic effects on fitness. On the other hand, some loci which have been identified and cloned usually because of the presence of mutations causing major phenotypic changes have been surveyed within populations at the DNA sequence or restriction enzyme levels. Such surveys indicate that very many variants typically segregate at loci in natural populations, and there is often linkage disequilibrium at the intra-genic level (Leigh Brown, 1989). Comparisons at the molecular level between populations and species show that there is much variation between loci in rates of fixation. This should relate in some way to within-population molecular variability at a locus and in turn to the amount of additive genetic variation in a quantitative trait which it affects. Many recent theoretical studies of the maintenance of genetic variation in quantitative traits by mutation–selection balance have involved comparisons between models in which different numbers of alleles can segregate at loci controlling the trait.

Kimura (1957) derived formulae for the fixation probability (*u*) of a new mutant allele at a locus under selection and drift for the two-allele case, and later (1969) for expected heterozygosity contributed by an allele until fixed or lost. Developments from Kimura's work have provided formulae for genetic variance maintained and rates of genetic progress in quantitative characters subject to recurrent mutation and directional selection (Hill, 1982), or stabilizing selection (Bulmer, 1972). A multiple allele model which was also initially analysed by Kimura (1965) and later by Lande (1976, 1980), assumed that an infinite number of alleles can segregate at a locus, and each allele can mutate to a new allele with a slightly different value from the previous state. Under these assumptions, the distribution of effects of alleles segregating at the locus is normal. If the trait is under stabilizing selection, however, this model is generally held not to be appropriate because normality is an unrealistic assumption. The arguments are complex, but depend on inferences of effects of new mutant alleles and mutation rates per locus based on experimental estimates of amount of new mutational variance per generation and the total number of loci likely to influence any one trait (Turelli, 1984). Later studies (Slatkin, 1987; Bulmer, 1989) confirmed that a two-allele model provides an accurate approximation

for genetic variance at equilibrium under stabilizing selection even if more than two alleles can segregate.

The presence of multiple alleles in models for maintenance of variation and response to selection in quantitative traits still presents analytical problems, however, and much of our own previous analysis in this area has been restricted to the case of two alleles per locus (Keightley & Hill, 1990). For a locus at mutation–selection balance in a finite population, genetic variance does not increase linearly with the number of alleles segregating (e.g. as a function of mutation rate or population size), but this phenomenon has not previously been quantified. Birky & Walsh (1988) showed, however, that the fixation probability of a mutant allele approaches that for neutral alleles, as the product of mutation rate at the locus and population size increases, and Barton (personal communication) has derived formulae for the amount by which the fixation probability of an advantageous allele is reduced by alleles generating variance of fitness in the region linked to the locus. This paper investigates the relationship between the amount of additive genetic variance maintained in a quantitative trait and fixation probabilities of alleles contributing to the variance in finite populations in simulation models in which the number of alleles segregating at a locus can be very large.

2. Model

(i) *Mutation*

Consider a locus at which there is no intragenic recombination affecting a quantitative trait in a random mating diploid population of effective size N with discrete generations. The expected number of mutations per generation is μ. An additive model is assumed in which each mutant deviation, a, is the difference between the homozygous states, and these are samples from a stationary distribution, $f(a)$, and are added to the current value at the locus. This differs from a model of the mutation process proposed by Cockerham & Tachida (1987) in which the value of the new allele is independent of the current state, and is sampled from some distribution. In the present model, absolute values of effects of mutations on the trait are sampled from a gamma distribution with shape parameter one-half:

$$f(|a|) = \alpha^{\frac{1}{2}} e^{-\alpha a} a^{-\frac{1}{2}} / \Gamma(\tfrac{1}{2}),$$

where $\Gamma(\ )$ is the gamma function, and α defines the spread of the distribution according to

$$E(a^2) = 3/(4\alpha^2), \quad E(|a|) = 1/(2\alpha) \text{ and } V(|a|) = 1/(2\alpha^2).$$

A gamma distribution was chosen as this provides a highly leptokurtic distribution of effects of new mutant alleles, i.e. the model is of many silent or nearly silent mutations (e.g. at the third base pair of a codon), and

of relatively few causing drastic changes to the gene product.

(ii) *Selection*

Two variations of the basic gamma distribution are considered which relate either to natural or artificial selection at the locus.

(1) *Deleterious alleles only.* If the trait has previously been subject to the rigours of natural selection and is at an adaptive peak, almost all new alleles are likely to be deleterious. To model this case, mutant deviations are sampled from the gamma distribution and unconditionally given a negative sign, i.e. the gamma distribution is reflected about $a = 0$ such that there is zero density for $a > 0$, so $s = -|a|$.

(2) *Advantageous and deleterious mutants.* Mutant effects have equal probability of being positive or negative. This is a model of artificial selection of a trait for which there is no *a priori* reason to expect that there would be more deleterious than advantageous mutations. The gamma distribution is reflected about $a = 0$ such that there is equal density for $a > 0$ and $a < 0$. For truncation selection the selection coefficient of a mutant allele would be $s = ia/\sigma$, but for these purposes, i/σ is assumed to be unity so $s = a$.

(iii) *Rare mutation, up to two alleles* $(N\mu \to 0)$

When the product of mutation rate and effective population size is small (i.e. $N\mu < 1$) it can be assumed that no more than two alleles segregate at the locus, so fixation probability and genetic variance maintained can be obtained from Kimura's formulae. The fixation probability is expressed as 'relative fixation probability', i.e. as a fraction of the fixation probability if the mutants were selectively neutral, $1/(2N)$. The fixation probability of a mutation of selective value s is given by Kimura's (1957) formula:

$$u(s) = (1 - e^{-s})/(1 - e^{-2Ns}),$$

for mutants with initial frequency $1/(2N)$. With a continuous distribution of s, the relative fixation probability is therefore

$$u_R = 2NE[u(s)] = 2N \int [(1 - e^{-s})/(1 - e^{-2Ns})] f(s) \, ds.$$

The 'relative variance' is the equilibrium additive genetic variance among individuals contributed by the locus as a proportion of the neutral variance which is the variance maintained in the absence of selection,

$$V_N = 2NV_M = N\mu E(a^2),$$

where V_M is the expected increment in variance at the locus among individuals per generation from mutation. In this case the asymptotic variance in the trait is proportional to the product of mutation rate and expected heterozygosity contributed by a mutant

during its lifetime, $H = 4[u(s) - 1/2N]/s$, (Kimura, 1969). The relative variance is therefore

$$V_R = N\mu \int [(s^2/4) H(s) f(s) ds]/V_N.$$

Relative variance and fixation probability were evaluated by numerical integration using Simpson's rule over the range $-\infty$, 0 for natural selection, and $-\infty$, ∞ for artificial selection.

(iv) *Monte Carlo simulation of the general multi-allele case*

For the general case of segregation of many alleles at the locus, relative fixation probability and asymptotic variance were evaluated by Monte Carlo simulation. The population consisted of N diploid individuals with a single locus affecting the trait. Fertility selection was carried out, and the relative fertility, W_i, of an individual was assigned according to $W_i = 1 + X_i - \bar{X}$, where X_i is the value of the individual and $\bar{X}$ is the population mean. The number of mutations per haploid per generation was sampled from a Poisson distribution with parameter μ. Mutant deviations were sampled from a gamma distribution, given sign appropriate to the selection regime and added to the current value at the locus. The population was started from a homozygous state and allowed to equilibrate for $6N$ generations. Thereafter, additive genetic variance among individuals and numbers of alleles fixed per generation were averaged over independent replicates. Relative variance was computed by dividing by the asymptotic neutral variance which was measured by computing the steady state variance among individuals at a locus with no effect on fitness, the variance at which was increased by $V_M/2$ units each generation. Because of constraints on computing resources the largest value of $N\mu$ simulated was 10 with $N = 80$ (see below).

3. Results

(i) *Validity of simulation using a small population*

To determine if a small population with high mutation rate is a good approximation of a large population with correspondingly small mutation rate, simulation runs for a range of population size were carried out. These showed that relative variance and fixation probability reach asymptotes with increasing N as long as $N\mu$ and $E(N|s|)$ are constant (i.e. if N is changed and corresponding changes are made to μ and $E(|s|)$), and that using a population size of $N = 80$ gives values close to the asymptote for $E(N|s|)$ < 10 (data not shown).

(ii) *Relative variance*

Figure 1*a* shows relative variance for natural selection (all mutants deleterious with respect to fitness). With
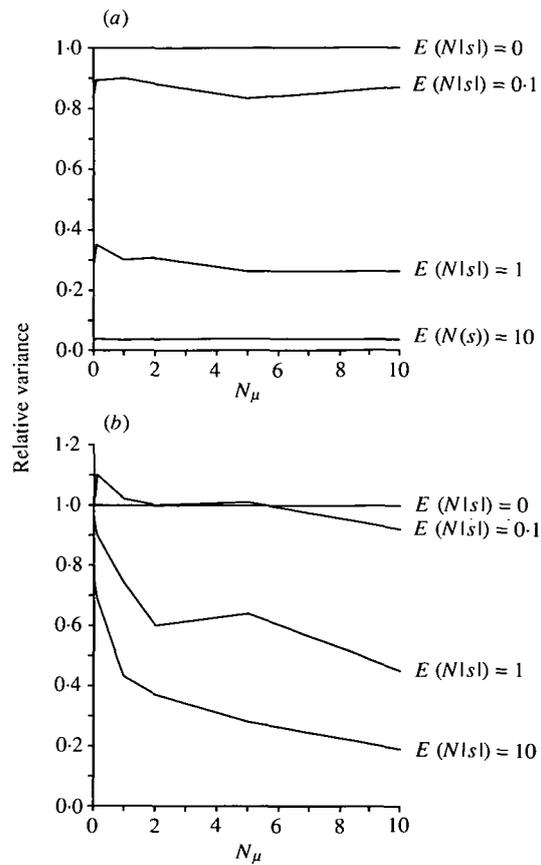

Fig. 1. Relative genetic variance for three values of $E(N|s|)$ with variation in the mutation rate $N\mu$. Kimura's two-allele approximation was used for $N\mu \to 0$; other results were computed using Monte Carlo simulation and a population of size $N = 80$. (*a*) Natural selection (all mutants deleterious). (*b*) Artificial selection (half mutants advantageous).

weak selection $(E(N|s|) \to 0)$, relative variance is one because mutant alleles behave as if they are neutral, but increasing the strength of selection leads to a reduction in relative variance because selection tends to eliminate deleterious alleles. Relative variance does not depend to a great extent on $N\mu$, implying that the two-allele approximation is appropriate for this type of selection regime. For intermediate strengths of selection, however, a very slight reduction in the relative variance occurs as $N\mu$ increases. This slight reduction was found to be greater if mutations have equal effects rather than a gamma distribution (data not shown). The reduction can be attributed to two related causes. (i) Selection generates association of alleles removed from their equilibrium frequencies. If we considered the locus as a chromosome of map length zero, these associations would be analogous to a linkage disequilibrium covariance component (see e.g. Bulmer, 1980, ch. 9) which tends to be negative under selection. In the simulations, the 'genic variance' $(V_g = \Sigma a^2 q[1-q]/2$, obtained by considering each new mutant as occurring at a separate independent locus and segregating in Hardy–Weinberg proportions) was found not to decrease with increasing

10

$N\mu$. (ii) Skew in the distribution of genotypic values generated by selection which leads to a higher rate of elimination of deleterious alleles (Hill & Robertson, 1966).

Relative variance for artificial selection (half mutant alleles advantageous and half deleterious) is shown in Fig. 1*b*. With low mutation rates and as long as selection is not too strong, the relative variance approaches one. This result agrees with Hill (1982) who showed that asymptotic variance is $4NV_{\mathrm{M}}\gamma$, where γ is the fraction of the mutational variance contributed by mutants of positive effect (one-half in this case). There is, however, a severe reduction in variance with increasing mutation rate and, in contrast to natural selection (Fig. 1*a*), the reduction increases with increasing strength of selection. The reasons for this reduction of variance are, however, similar to those for natural selection above.

It can be inferred from Fig. 1 that the *actual* genetic variance at the locus is proportional to μ, but becomes independent of N for both selection regimes if s is constant.

(iii) *Relative fixation probability*

Figure 2*a* shows relative fixation probability at the locus with natural selection. The points for $N\mu \to 0$ were obtained using Kimura's formula, and these agree with the Monte Carlo simulation using a low mutation rate. If $E(N|s|)$ is much greater than 1 and the mutation rate is low, relative fixation

probability becomes very small. Increasing the mutation rate leads to a gradual increase in relative fixation probability, and eventually the fixation probability becomes the same as for neutral alleles (see Birky & Walsh, 1988). Interestingly, even with strong natural selection the neutral fixation probability is approached with increasing $N\mu$, but relative variance is independent of $N\mu$ (Fig. 1*a*). In essence, if there is little variation already at the locus, any new mutant allele is strongly deleterious and tends to be eliminated very quickly. If the mutation rate is sufficiently high such that there are many alleles segregating at the locus, however, a new mutation may appear at the individual chromosome with the best genotype with respect to the locus (with probability $1/2N$) and become fixed by hitch-hiking. This does not, however, generate additive variance in the trait. With advantageous alleles (Fig. 2*b*) increasing $N\mu$ leads to a much faster approach to the neutral fixation probability, because lifetimes of advantageous alleles in the population are longer than for deleterious alleles and there is therefore greater chance of mutations interfering with one another. An alternative way of explaining the above phenomenon is that the effective population size in which a given mutant allele is selected becomes very small due to the input of large numbers of mutations and its fate depends to a large extent on chance (Hill & Robertson, 1966; Birky & Walsh, 1988).

4. Discussion

(i) *Models*

Several models of the mutation process might have been chosen for this study, but there are inevitably unrealistic assumptions. The present model is 'stepwise' because the new allelic state is conditional on the previous state, with the new allelic deviation added to the current value. This contrasts with Cockerham & Tachida's (1987) model where the new allelic state does not depend on the previous state, but is sampled from a stationary distribution. A better model would include a mixture of the above: some substitutions may have step-wise effects on the gene product (e.g. a change in activity followed by a change in concentration), but mutations which lead to null activity could occur from any allelic state. The assumption of a continuous distribution of allelic deviations is also unrealistic because these are likely to fall into discrete classes. However, with equal absolute values of effects of mutations rather than the gamma distribution in the model the pattern of the results changed little (data not shown). Foley (1987) derived approximations for the fixation probability of mutant alleles at loci under stabilizing selection and these are also appropriate approximations to the left hand end of the curves in Fig. 2*a* ($N\mu$ small). Although the model here involves alleles with fixed selection coefficients it
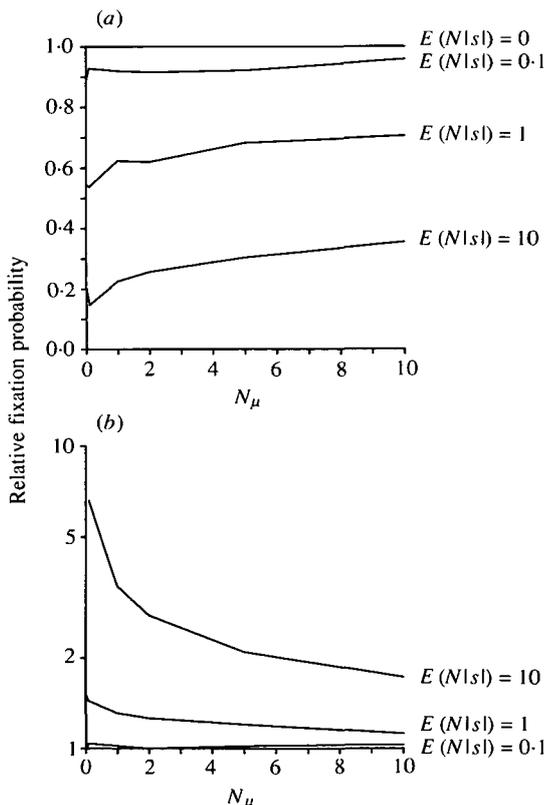

Fig. 2. Relative fixation probability computed as explained in the legend to Fig. 1.

is very similar to one of pure stabilizing selection, because although selection coefficients of new mutant alleles under stabilizing selection are gene frequency-dependent, they tend to be negative and most disadvantageous when rare (Robertson, 1956). Although the present model of natural selection assumes that all mutant alleles are deleterious, presumably some advantageous alleles occur otherwise population mean fitness would continuously decline, but the fraction of advantageous alleles need not be very large (Kimura, 1983, ch. 8).

(ii) *Genetic variance and fixation probability*

Providing mutant alleles are deleterious, the present analysis shows that the two-allele model gives an accurate approximation for asymptotic genetic variance providing mutant alleles are deleterious. The principle reason for this is that deleterious alleles become eliminated at a high rate and have little chance of interfering with the fixation of other alleles at the locus. A critical parameter needed to predict the asymptotic genetic variance is the total number of mutations occurring per generation at loci affecting the trait, $\lambda = n\mu$ where n is the number of loci. The fixation probability of deleterious alleles can, however, approach that for neutral alleles so the fixation rate would bear little relation to the variance maintained. If natural selection favoured heterozygotes, however, the presence of multiple alleles would have a strong effect on genetic variance maintained, because alleles would have longer life-expectancies in the population and therefore greater opportunity to interfere with one another. The actual genetic variance as a function of N would be concave upwards (because lifetime heterozygosity per mutant becomes very large with increasing N), in contrast to the case of deleterious alleles for which it is concave downwards (with neutral or advantageous alleles, variance is approximately linear with N).

With artificial selection, mutations of positive effect on the trait are expected to occur. Because mutant alleles become fixed by selection (in contrast to the model of natural selection where they do not), alleles tend to interfere with each others' fixation, so increasing mutation rate leads to a reduction in relative genetic variance. This is analogous to the reduction in V_G observed because of linkage.

(iii) *Implications*

For populations subject to directional artificial selection, the presence of multiple alleles can be ignored and the two-allele approximation applies because the mutation rate per locus is unlikely to be greater than 10^{-5} per generation and effective population sizes are generally much less than 10^3. Natural populations may be very much larger and it is conceivable that the product of effective population size and mutation rate

can exceed the largest value (10) shown in the results. The relative genetic variance at a quantitative trait locus would be independent of $N\mu$, but the fixation probability of mutant alleles could approach the neutral prediction. Thus, quantitative genetic variance and fixation rates at the quantitative trait loci can be 'decoupled' from one another, but populations would have to remain very large for long periods in order to approach the asymptotic state.

References

Birky, C. W. & Walsh, J. B. (1988). Effect of linkage on molecular evolution. *Proceedings of the National Academy of Sciences, USA* **85**, 6414–6418.

Bulmer, M. G. (1972). Genetic variability of polygenic characters under optimizing selection, mutation and drift. *Genetic Research* **19**, 17–25.

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.

Cockerham, C. C. & Tachida, H. (1987). Evolution and maintenance of quantitative genetic variation by mutations. *Proceedings of the National Academy of Sciences, USA* **84**, 6205–6209.

Foley, P. (1987). Molecular clock rates at loci under stabilizing selection. *Proceedings of the National Academy of Sciences, USA* **84**, 7996–8000.

Hill, W. G. (1982). Rates of change in quantitative traits from fixation of new mutations. *Proceedings of the National Academy of Sciences, USA* **79**, 142–145.

Hill, W. G. & Robertson, A. (1966). The effects of linkage on limits to artificial selection. *Genetical Research* **8**, 269–294.

Keightley, P. D. & Hill, W. G. (1990). Variation maintained in quantitative traits with mutation-selection balance: pleiotropic side-effects on fitness traits. *Proceedings of the Royal Society of London Series* B **242**, 95–100.

Kimura, M. (1957). Some problems of stochastic processes in genetics. *Annals of Mathematical Statistics* **28**, 882–901.

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

Kimura, M. (1969). The number of heterozygous nucleotide sites maintained in a finite population due to a steady flux of mutations. *Genetics* **61**, 893–903.

Kimura, M. (1983). *The Neutral Theory of Molecular Evolution*. Cambridge University Press.

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

Lande, R. (1980). The genetic covariance of characters maintained by pleiotropic mutations. *Genetics* **94**, 203–215.

Leigh Brown, A. J. (1989). Population Genetics at the DNA level: a review of the contribution of restriction enzyme studies. *Oxford Surveys in Evolutionary Biology* **6**, 207–242.

Robertson, A. (1956). The effect of selection against extreme

deviants based on deviation or on homozygosis. *Journal of Genetics* **54**, 236–248.

Slatkin, M. (1987). Heritable variation and heterozygosity under a balance between mutations and stabilizing selection. *Genetical Research* **50**, 53–62.

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