Theor Appl Genet (1983) 65:25–30

Effect on linkage disequilibrium of selection for a quantitative character with epistasis

J. P. Mueller and J. W. James
School of Wool and Pastoral Sciences, University of New South Wales, Kensington, 2033, Australia

Received November 30, 1982
Communicated by J. S. F. Barker

Summary. Selection for a character controlled by additive genes induces linkage disequilibrium which reduces the additive genetic variance usable for further selective gains. Additive × additive epistasis contributes to selection response through development of linkage disequilibrium between interacting loci. To investigate the relative importance of the two effects of linkage disequilibrium, formulae are presented and results are reported of simulations using models involving additive, additive × additive and dominance components. The results suggest that so long as epistatic effects are not large relative to additive effects, and the proportion of pairs of loci which show epistasis is not very high, the predominant effect of linkage disequilibrium will be to reduce the rate of selection response.

Key words: Linkage disequilibrium – Selection – Genetic variance – Quantitative traits – Epistasis

Introduction

Selection for a quantitative character will in general lead to response through two mechanisms; change in gene frequency, and generation of linkage disequilibrium. Griffing (1960) showed that gene frequency responses were dependent on additive genetic variance, while linkage disequilibrium responses were dependent on epistatic variance of the additive × additive type. If V_A and V_{AA} denote the additive and additive by additive variances, σ_P the phenotypic standard deviation and i the standardised selection differential, the response to one generation of selection is $(i/\sigma_P)(V_A + V_{AA}/2)$ provided three-locus and higher order epistasis is negligible. Gene frequency changes persist when selection ceases, but linkage disequilibrium decays and the contribution of epistatic variance to response gradually disappears. These changes in the genetic constitution of the population will produce alterations in the genetic variance which Griffing (1960) assumed negligible in the short term for traits controlled by a large number of loci each of small effect. Nei (1963) discussed changes in genetic variance produced by selection, and showed that the linkage disequilibrium between any pair of loci would be small when gene effects were small. He therefore ignored it and dealt further only with changes in variance produced by gene frequency changes.

Bulmer (1971), on the other hand, using a model of a normally distributed quantitative character with no epistasis, showed that selection could lead to an immediate appreciable decrease in genetic variance due to establishment of linkage disequilibrium. If the phenotypic variance of selected parents is a fraction $1 - K$ of that in the whole population, the additive genetic variance in the progeny is reduced by a fraction $\frac{1}{2} K h^2$, where h^2 is the heritability of the trait. Further selection adds additional linkage disequilibrium, while recombination reduces that previously accumulated, so that a steady state is reached fairly rapidly. Thus linkage disequilibrium, by reducing the usable additive genetic variance, will decrease the rate of response to selection. This contrasts with the effect discussed by Griffing, where linkage disequilibrium contributes positively to selection response. For traits showing both additive and epistatic genetic variance it is necessary to know the relative importance of these effects in order to predict responses to continuing selection.

The complexity of the general case has prevented theoretical solutions more advanced than those dis-

cussed above, although Felsenstein (1965) has discussed the general nature of the effect of directional selection on linkage disequilibrium. In such circumstances simulation may be very helpful in gaining an understanding of the problem. For small populations Gill (1965) considered the contribution of epistatic variance to selection response to be grossly overestimated by Griffing's formula. Young (1967) simulated selection in a large population and concluded that the predictive ability of h^2 (recalculated each generation) is poor when genes show epistasis. In neither set of simulations was any attempt made to specify which component of overall genetic variation led to discrepancies. To do this, it is necessary to partition the total variance into a series of components and to discover what changes are produced by selection in all components. In this paper we present results of a series of simulations designed to investigate such changes in detail.

Partition of genetic variance

A completely general model appears too complex to be readily analysed, so attention is restricted to a narrower class of models. Suppose we are dealing with a trait in a diploid species which has a finite number of loci affecting the trait. With no epistasis, the genetic value G of the trait for an individual is given by the sum over loci of the genotypic values at each locus. We suppose that the n loci are divided into $\frac{1}{2}$ n pairs, with epistasis within but not between these pairs of loci. Thus, for the j-th pair of loci the genetic effect g_j is determined by the particular combinations of alleles present at the two loci, while overall genetic value is given by $G = \sum_{j}^{\text{pairs}} g_j$. The total genetic variance V_G can be calculated directly and in addition

$$V_G = \sum_{j}^{\text{pairs}} \text{var}(g_j) + 2 \sum_{j \neq j'} \text{cov}(g_j, g_{j'}).$$

The second term is the sum of covariances between pairs and will be denoted C_{LB}. It will be zero if there is linkage equilibrium. For any pair of loci, the variance can be partitioned into additive and dominance components at each locus, plus the epistatic components, and a covariance term due to linkage disequilibrium. If α_i and δ_i are the additive effect and the dominance effect at the i-th locus which has gene frequency p_i the additive and dominance variances are

$$V_{Ai} = 2\, p_i (1 - p_i)\, \alpha_i^2$$

and

$$V_{Di} = 4\, p_i^2 (1 - p_i)^2\, \delta_i^2.$$

If only additive × additive epistasis occurs, the epistatic variance at the j-th pair of loci will be

$$V_{AAj} = 4\, p_{j_1} (1 - p_{j_1})\, p_{j_2} (1 - p_{j_2})\, (\alpha\alpha)_j^2,$$

where j_1 and j_2 denote the loci of the j-th pair, and $(\alpha\alpha)_j$ denotes the additive × additive effect at these loci. Then

$$\text{var}(g_j) = V_{Aj_1} + V_{Aj_2} + V_{Dj_1} + V_{Dj_2} + V_{AAj} + C_{LWj}.$$

Here, C_{LWj} represents the covariance contribution to the variance of genetic values within the j-th pair of loci, and is zero in the absence of linkage disequilibrium. Then

$$\sum_{j}^{\text{pairs}} \text{var}(g_j) = \sum_{i}^{\text{loci}} (V_{Ai} + V_{Di}) + \sum_{j}^{\text{pairs}} V_{AAj} + \sum_{j}^{\text{pairs}} C_{LWj}.$$

This last sum we refer to as C_{LW}. In the absence of epistasis there would be no need to separate C_{LB} and C_{LW} and the sum of these terms is then C_L of Bulmer, but possible differences in their behaviour is the object of this investigation. One further component needs to be considered, namely C_{HW} (Bulmer 1976) which is the contribution due to deviations from Hardy-Weinberg distribution at the various loci. This is calculated as the difference between the actual value of the variance at each locus and the value as given above, assuming Hardy-Weinberg distribution. Since it is possible to calculate actual values of var (g_j) and also the theoretical variance components (V_A, V_D, V_{AA}), we can calculate C_{LW}. Since we can also find V_G, C_{LB} can be found by subtraction. Thus, all components can be calculated.

Simulation procedures

A monoecious population with selfing permitted, in which selection was for a character controlled by 40 unlinked loci, each with gene frequency 0.5, was simulated. In each generation 100 individuals were produced, ranked on phenotype, and the best 50 were selected as parents of the next generation. Phenotypes were calculated by adding to G (genotypic value) a random normally distributed variable with mean zero and variance equal to the initial total genetic variance. Thus, the initial broad sense heritability was 0.5. Offspring were produced by combining gametes at random from the parents. Genotypic value was determined from the genotype and the particular genetic model simulated, there being four models. The loci were divided into 20 pairs and, for each pair, values were specified for all 9 genotypes. The total genotypic value was the sum of these 20 values (g_j). For convenience, the models are named according to the types of genetic variance present in the base population. The genotypic values for the 4 models were as follows:

Model	AABB	AaBB	aaBB	AABb	AaBb	aaBb	AAbb	Aabb	aabb
A	4	3	2	3	2	1	2	1	0
AA	4	2	0	2	2	2	0	2	4
A+AA	8	4	0	4	2	0	0	0	0
D+AA	7	10	11	10	11	10	11	10	7

Model A+AA was used by Young (1967) and model D+AA by Gill (1965). Although for models AA and D+AA there is no additive variance in the base population, changes in gene frequency will generate such variance. For these models

$$V_A = 8 \sum_{j=1}^{20} [p_{j_1}(1-p_{j_1})(2p_{j_2}-1)^2 + p_{j_2}(1-p_{j_2})(2p_{j_1}-1)^2] \quad \text{(AA)}$$

and

$$V_A = 2 \sum_{j=1}^{20} [p_{j_1}(1-p_{j_1})(3-4p_{j_2}-2p_{j_1})^2 + p_{j_2}(1-p_{j_2})(3-4p_{j_1}-2p_{j_2})^2]. \quad \text{(D+AA)}$$

Ten replicate populations were run over 20 generations for each model, means and standard errors of various quantities being estimated from these.

Results

Figure 1 shows results from Model A. As in the simulations of Bulmer (1976), the changes in the early generations are consistent with predictions from the infinite loci model. For instance, since $K = 0.637$, we expect from Bulmer (1971) that at generation 3 the genetic variance will be 15.7 and obtained $V_G = 15.1 \pm 0.7$. The average change in gene frequency over this time period was 0.08. As selection proceeds, changes in gene frequency produce larger effects on V_A, but the total genetic variance always remains smaller than V_A, showing that C_L is negative. It has not been thought worthwhile to show C_{HW} as its magnitude is negligible, averaging -0.07 over the 20 generations. In an unselected finite population

$$\xi(C_{HW}) = -\sum_{i}^{loci} p_i(1-p_i)/N$$

where N is the effective population size. In our case this gives -0.1 as the expected value in the base population, and as gene frequencies shift from 0.5 it declines. It will therefore be ignored in what follows.

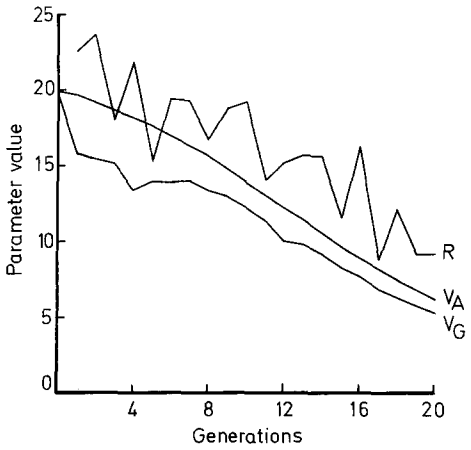

Fig. 1. Changes in parameters with generations of selection for Model A. For definition of parameters see text. The scale for R is 10 times that for other parameters

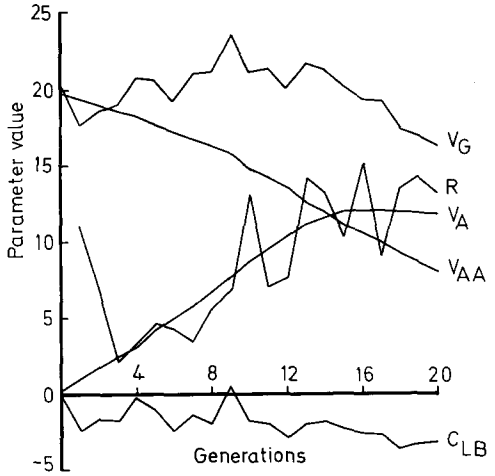

Fig. 2. Changes in parameters with generations of selection for Model AA. For definition of parameters see text. The scale for R is 10 times that for other parameters. (C_{LW} is not shown because it is negligible)

Results for Model AA are shown in Fig. 2. Since for this model there is no additive variance in the base population, any initial response should be due to linkage disequilibrium. However, in a finite population, genetic drift will shift frequencies of genes from 0.5, and then selection will move gene frequencies either both up or both down at each pair, but the average gene frequency over all pairs will still be 0.5. V_A will have a maximum when gene frequencies are $\frac{1}{2} \pm \frac{1}{4}\sqrt{2}$ at both loci, subject to the restriction of equal gene frequencies. Since selection acts equally on both loci, this restriction is reasonable. There is a clear rise in V_A over this period, the changes in gene frequency being such that maximum V_A was reached at about the end of the simulation. The epistatic variance declined as gene frequencies changed with the result that the total variance remained relatively constant. C_{LB} was consistently negative, especially in the later generations when h^2 was higher. The dominant factor affecting the response to selection was clearly the change in V_A. The first generation response predicted from Griffing (1960) is 1.26, while the observed value was 1.10 ± 0.29. The response declines as predicted for two generations before the development of additive genetic variance causes it to rise again. With selection beyond 20 generations the approximately constant rate of response observed over generations 13 to 20 would decrease as V_A approached zero.

Figure 3 shows the results for model A+AA. In this model there are favoured (plus) alleles whose frequencies will all increase. As frequencies increase, V_A will rise until it reaches a maximum at $p = \frac{3}{4}$ (Young 1967). Maximum V_A is about 1.5 times its initial value and fairly constant over generations 6–11, where gene frequencies rise from 0.66 to 0.78 (SE $\pm$ 0.01) Despite the

increase in V_A, response does not increase very much, because C_{LB} is negative, and about 10 times the absolute value of C_{LW}, which is positive. Griffing's (1960) formulae for response to repeated selection ignore changes in gene frequency and C_{LB}. But, because $V_A - C_{LB}$ remains approximately constant, the predictions are fairly accurate. Thus, for example, the following values show reasonable agreement.

Generation	Cumulative response	
	Predicted	Observed
4	19.1	19.2 ± 1.3
7	32.7	35.6 ± 1.6
20	91.4	97.6 ± 1.4

This agreement is, however, fortuitous, in the sense that with a different set of initial gene frequencies, for example, such an outcome could not be expected. Equally, with a larger number of loci the change in gene frequency at each would be smaller and C_{LB} would be relatively more important. With tight linkage, the contribution of V_{AA} to response would be greater, but C_{LB} would also be larger in magnitude.

Figure 4 shows the results obtained with model D+AA. In contrast to the previous 3 models, there is in this case very little additive genetic variance at any time, and total response to selection is very small, being only 0.6 initial phenotypic standard deviations over 20 generations (compared with 5.1, 2.8 and 6.9 for models A, AA and A+AA respectively). V_G declines fairly slowly as V_{AA} and V_D decline and V_A rises, though the changes are not very great. Since for this model $(\alpha\alpha)_j$ is negative for interacting loci it produces negative linkage disequilibrium. As seen in Fig. 4, C_{LW} is consistently negative. On the other hand C_{LB} fluctuates above and below zero over the 20 generations, and has relatively large standard errors every generation. The observed values of C_{LB} appear to be mainly influenced by sampling fluctuations. Thus, in this case, the epistatic effect on linkage disequilibrium is relatively more important than that due to the Bulmer effect. This is to be expected since in this case V_A is always small and the magnitude of C_{LB} is proportional to heritability.

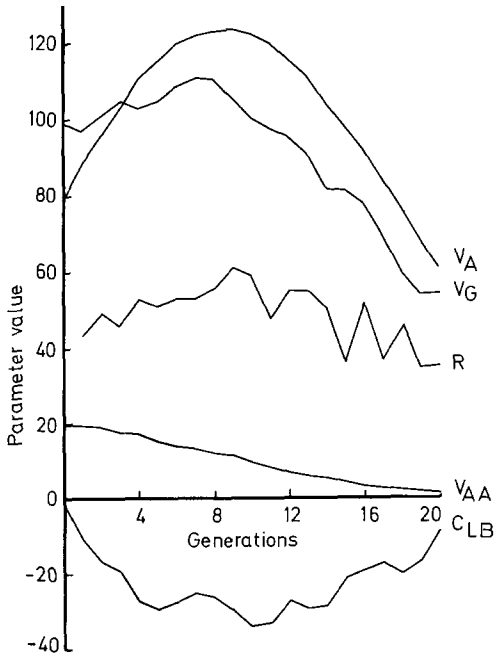

Fig. 3. Changes in parameters with generations of selection for Model A+AA. For definition of parameters see text. The scale for R is 10 times that for other parameters. (C_{LW} is not shown because it is negligible)

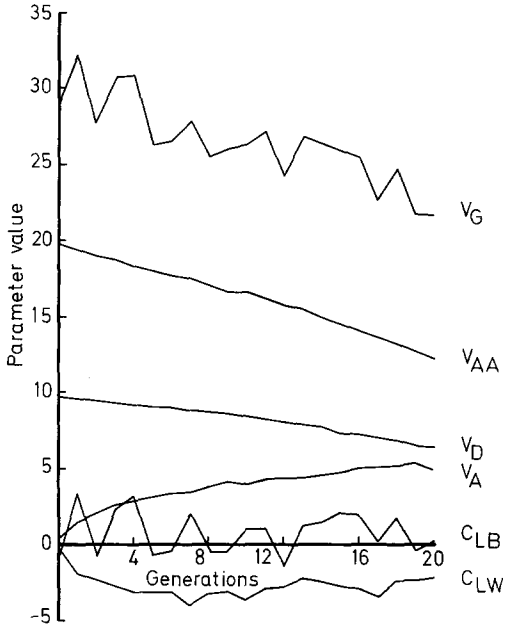

Fig. 4. Changes in parameters with generations of selection for Model D+AA. For definition of parameters see text. (R is not shown because it is negligible)

Discussion

It will be useful to discuss the behaviour of genetic variance components in relation to well known theory. A finite population not undergoing selection will have its genetic variance in subsequent generations affected by sampling in three ways.

1. V_A will be reduced at a rate 1/2 N per generation due to drift in gene frequency at individual loci.

2. There will be random deviations from Hardy-Weinberg proportions, again proportional to 1/N, but these are not cumulative.

3. There will be random fluctuations in linkage disequilibrium. Linkage disequilibrium is, on average, zero but will vary between lines. In our simulations,

therefore, this will not affect mean linkage disequilibrium effects but will affect the standard errors. This effect also is proportional to 1/N. With selection for an additive trait, there are added two further effects.

4. There is a reduction in usable additive variance due to induced negative correlations between loci.

5. When appreciable changes in gene frequency occur, V_A will be affected. In the long run, V_A must decline as useful alleles approach fixation but, in the shorter term, V_A may increase as gene frequencies approach intermediate values (i.e., favoured alleles are initially less common). If the trait also shows epistasis,

6. there is a contribution of epistatic variance to response due to establishment of favourable correlations between loci. (Differential selection in males and females would give a contribution to deviation from Hardy-Weinberg proportions, but the effect is inversely proportional to the number of loci and, for polygenic traits, is of little importance.)

Our concern in this paper has been mainly with the relative importance of (4) and (6). We have not been much interested in random effects, so have used a moderately large population size to keep them small. Nevertheless, random drift is very important in model AA since, initially, V_A is zero and, in an infinite population, gene frequency would remain unchanged. However, random drift causes deviations to arise and give rise to a positive V_A. At this point, selection begins systematically to affect gene frequencies, as is clear from the results of the AA model.

The formulae for C_{LB} and C_{LW} assumed (1), (2) and (3) negligible. For the time span considered and the effective size used, (1) is not of much consequence in comparison with (4) and (5). Since for Model A $N\,i\,\alpha/\sigma_P = 6.3$ (considerably greater than 1) where i is the standardised selection differential and σ_P is the phenotypic standard deviation, selection effects will be much more important than random drift (Robertson 1960). Except for Model AA in the early generations, the same will be true to varying extents for the other models, but as α is not constant, a specific $N\,i\,\alpha/\sigma_P$ value cannot be given for them. Deviations from Hardy-Weinberg (2) were calculated and found to have very little effect on the variance. The magnitude of (3) is important for contributing to the variance between populations. However, the major forces will be those of (4) and (6), with (5) also being important in some circumstances. For instance, Sorensen and Hill (1982) postulated an additive model with one or more genes of large effect as compatible with results from selection on abdominal bristle number in *Drosophila*, suggesting that the Bulmer effect, though important, can be less marked than the modification of V_A caused by gene frequency changes.

In order to compare (4) and (6) we need to see their relative effects on linkage disequilibrium. Using the results of Griffing (1960) and Nei (1963) the linkage disequlibrium generated at the j-th pair of loci is given by

$$D_j \approx \frac{i}{\sigma_P} p_{j_1}(1-p_{j_1})\, p_{j_2}(1-p_{j_2})\,(\alpha\alpha)_j - \Delta p_{j_1}\Delta p_{j_2} \qquad (1)$$

where Δp_{j_1} and Δp_{j_2} are the changes in gene frequency at these loci produced by selection. This equation may be written

$$D_j \approx \frac{i}{\sigma_P} p_{j_1}(1-p_{j_1})\, p_{j_2}(1-p_{j_2}) \left[(\alpha\alpha)_j - \frac{i}{\sigma_P}\alpha_{j_1}\alpha_{j_2}\right]. \qquad (2)$$

When α/σ_P is small, $(\alpha\alpha)_j$ will be the dominant term in the expression in [ ], provided the loci show non-negligible epistasis. Thus, for pairs of loci of this type, linkage disequilibrium will be mostly determined by $(\alpha\alpha)$. On the other hand, with n loci there are $\frac{1}{2}n(n-1)$ pairs of loci, and though the $\alpha_{j_1}\alpha_{j_2}/\sigma_P$ terms may be small, the sum of such a large number may be appreciable.

Assume there is no epistasis, so that C_{LW} and C_{LB} can be combined to give C_L. Then

$$C_L = -4\sum_{j \neq j'} \Delta p_{j_1}\Delta p_{j_2}\alpha_{j_1}\alpha_{j_2}.$$

For simplicity, assume there are n identical loci with completely additive gene action. Then, summing over pairs of loci

$$C_L = -2n(n-1)\,\Delta p^2\,\alpha^2.$$

The selection response is $R = 2n\alpha\Delta p$ so

$$C_L = -\frac{1}{2}\left(1-\frac{1}{n}\right)R^2$$

and since $R = i\,h^2\,\sigma_P$ we have

$$C_L = -\frac{1}{2}\left(1-\frac{1}{n}\right) i^2 h^4 \sigma_P^2$$

$$\approx -\frac{1}{2} i^2 h^2 V_A.$$

This differs from the result of Bulmer (1971) only in the use of i^2 instead of $i(i-x)$, the discrepancy coming from the approximations involved in our approach.

The point of importance is that the negative term in the expression for D_j has order of magnitude $1/n^2$. This is likely to be unimportant for a pair of loci which interact, and may be ignored for such a pair, as was done by Griffing (1960). However, the number of pairs of loci is of order n^2 and the total effect of these terms is not always negligible, as shown by Bulmer (1971). It is rather unlikely that all pairs of loci interact and, therefore, the epistatic effect will probably be contributed by only a small fraction of the pairs of loci

and, therefore, be of less importance than the additive effect on linkage disequilibrium. Indeed, even with all pairs of loci showing interaction of equal but small size, the overall contribution of linkage disequilibrium to selection response may not be positive.

In our simulations, only $\frac{1}{2}$ n pairs of loci interacted, but the $(\alpha\alpha)$ effects were large enough to yield large amounts of V_{AA}. The clear separation into epistatic and additive pairs of loci may be somewhat artificial, but allowed a simple partition of linkage disequilibrium effects into C_{LB} and C_{LW}. Other models are of course possible. However, if more pairs of loci interact, the $(\alpha\alpha)$ effects at such loci would have to be smaller, since the magnitude of V_{AA} cannot increase indefinitely in relation to the phenotypic variance. Thus, it seems reasonable to conclude that our results would hold for a wider class of models. Thus, for those pairs of loci with appreciable epistasis, the Griffing effect will predominate but, overall, the Bulmer effect will remain important and will reduce the additive genetic variance usable by selection. The rate of genetic gain will be reduced and it will take more generations for fixation of favourable alleles to occur. When population size is small enough for random drift to be important, sampling fluctuations will occur over a longer period, increasing the probability of fixation of unfavourable alleles.

Our simulations have all assumed no linkage. Tighter linkage will increase the amount of linkage disequilibrium, but will not alter any of the principles.

Acknowledgements. This work was carried out by J.P.M. while holding a scholarship from the Instituto Nacional de Tecnologia Agropecuaria, INTA (Argentina).

References

Bulmer MG (1971) The effect of selection on genetic variability. Am Nat 195:201–211

Bulmer MG (1976) The effect of selection on genetic variability: a simulation study. Genet Res, Camb 28:101–117

Felsenstein J (1965) The effect of linkage on directional selection. Genetics 52:349–363

Gill JL (1965) A Monte Carlo evaluation of predicted selection response. Aust J Biol Sci 18:999–1007

Griffing B (1960) Theoretical consequences of truncation selection based on the individual phenotype. Aust J Biol Sci 13:307–343

Nei M (1963) Effect of selection on the components of genetic variance. In: Hanson WD, Robinson HF (eds) Statistical genetics and plant breeding NAS-NRC Publ 982:501, 515

Robertson A (1960) A theory of limits in artificial selection. Proc R Soc Edinburgh Sect B 153:234–249

Sorensen DA, Hill WG (1982) Effect of short-term directional selection on genetic variability: experiments with *Drosophila melanogaster*. Heredity 48:27–33

Young SSY (1967) Computer simulation of directional selection in large populations. II. The additive × additive and mixed models. Genetics 56:73–87