Variability in genetic parameters among small populations

By P. J. AVERY and W. G. HILL

Institute of Animal Genetics, Edinburgh EH9 3JN

SUMMARY

For a model in which quantitative traits are assumed to be determined solely by additive genes at many loci, formulae are developed for the variance among replicated small populations of size N, maintained without selection, of the additive genetic variance, heritability, genetic correlations and similar parameters. The base population is assumed to be in linkage equilibrium, but it is argued that most of the variation in the within-line additive variance (V_{At} at generation t) is due to linkage disequilibrium caused by sampling. If $\overline{r_t^2}$ is the squared correlation of gene frequencies averaged over all pairs of loci at time t, the coefficient of variation (CV) of V_{At} equals $\sqrt{(2\overline{r_t^2})}$, with similar formulae for other parameters.

The formulae are evaluated for models of loci distributed uniformly along the chromosome but much of the disequilibrium is due to loci on different chromosomes. For unlinked loci $CV(\overline{V_{At}})$ reaches $\sqrt{4/(3(N))}$, and for mammalian models, this value is not greatly exceeded. The variance in successive generations has a correlation of at least one-half due to the maintenance of linkage disequilibrium. The magnitude of this variance in parameters and their autocorrelation with time shows that accurate predictions cannot be made about genetic parameters in the base population from single replicate results.

1. INTRODUCTION

There is an extensive theory in quantitative genetics for predicting the variance in mean performance between replicated small populations and the expected genetic variance within such populations (Falconer, 1960; Crow & Kimura, 1970). Results are only obtained readily, however, when no selection is practised. The formulae are simple in the case of additive genes, these variances being proportional to the variance in the foundation population and the extent of inbreeding or drift. With dominance, the proportional changes in variance depend also on gene frequencies (Robertson, 1952), and are more complicated with epistasis. Even with additive gene action, however, there has not until recently been any theory for predicting the variation among replicates in the genetic *variance* within replicates, perhaps largely because it was thought that this variance of variance would be highly dependent on numbers of loci and the distribution of their effects. Recently, however, Bulmer (1976) has shown that if many loci affect the trait, the variation of within-line variance between generations of the same replicate and by implication,

between replicates at the same generation, is contributed largely by linkage disequilibrium between pairs of loci, and if the number of loci is large, the distribution of individual gene effects and frequencies may be unimportant. Bulmer considered only two particular models of distribution of loci along chromosomes. In this paper some of his results are extended and generalized, and the formulae are developed to include correlated traits.

The model is restricted to additive genes and formally to the assumption of no selection, either natural or artificial. Without this assumption the algebra becomes formidable and some generality is lost. Providing selection pressures at individual loci are not strong, it is likely that parts of the results carry across to some selection models, and examples are discussed by Bulmer (1976). Some preliminary results of the following analysis are included in another paper (Hill, 1976), which also includes a discussion on the variation among replicates in response to directional selection. Apart from considerations of response to selection, the results to be described have relevance to the reliability of estimates of, for example, genetic variance, heritabilities and genetic correlations from populations, and the inferences that can be made about successive generations in the same subpopulations, the base population from which it was drawn and other similar populations.

2. ANALYSIS

Consider a set of replicated random mating lines, each of effective size N, sampled from a large base population in linkage equilibrium. There is no mutation, migration or selection and generations are discrete. In the following analysis there are assumed to be two alleles at each locus, but the results can be extended to multiple alleles. At locus i let the allele having high value for some quantitative trait have frequency q_i and let the difference in value between heterozygote and homozygote be a_i, with all effects being additive both within and between loci. The additive variance in the base population is V_A, and in a replicate line at generation t, it is V_{At}. Addition of a subscript t to other variables also denotes generation.

The additive variance in the base population is

$$V_A = 2\sum_i a_i^2 q_i(1-q_i), \tag{1}$$

since there is assumed to be linkage equilibrium and Hardy–Weinberg proportions initially; and from well known theory (e.g. Falconer, 1960)

$$E(V_{At}) = (1-1/(2N))^t V_A = (1-F_t) V_A, \tag{2}$$

where F_t is the inbreeding coefficient. In a particular replicate line in which the gene frequency at locus i is q_{it} and the disequilibrium or covariance of frequencies between loci i and j is D_{ijt},

$$V_{At} = 2\sum_i a_i^2 q_{it}(1-q_{it}) + 4\sum_{i<j}\sum a_i a_j D_{ijt}, \tag{3}$$

where V_{At} is taken as twice the variance between chromosomes, thus ignoring any departure from Hardy–Weinberg equilibrium. The validity of ignoring departures from Hardy–Weinberg will be mentioned later.

Table 1. *Values of $E(D^2)/\pi_0$ (calculated by moment generating matrix*
(Hill & Robertson, 1968), approximate $E(D^2)/\pi_0$ (from (8)) and
$E(q_{it}(1-q_{it})\,q_{jt}(1-q_{jt})/(\pi_0(1-1/(2N))^{2t})\ (=\alpha)$,   *where* $\pi_0 = q_i(1-q_i)\,q_j(1-q_j)$
$N = 20$

$t =$	2	5	10	20
		$c = 0.5$		
$E(D^2)/\pi_0$, exact	0·0288	0·0263	0·0205	0·0123
$E(D^2)/\pi_0$, approx.	0·0281	0·0256	0·0199	0·0120
α	1·00001	1·00008	1·00023	1·00053
		$c = 0.1$		
$E(D^2)/\pi_0$, exact	0·0415	0·0659	0·0679	0·0462
$E(D^2)/\pi_0$, approx.	0·0404	0·0640	0·0649	0·0428
α	1·00003	1·00059	1·00360	1·01500
		$c = 0.01$		
$E(D^2)/\pi_0$, exact	0·0453	0·0916	0·1309	0·1404
$E(D^2)/\pi_0$, approx.	0·0442	0·0888	0·1239	0·1221
α	1·00003	1·00088	1·00717	1·04755
		$c = 0$		
$E(D^2)/\pi_0$, exact	0·0458	0·0953	0·1425	0·1672
$E(D^2)/\pi_0$, approx.	0·0466	0·0923	0·1348	0·1443
α	1·00003	1·00092	1·00775	1·05506

(i) *Variance of additive variance*

From (3)

$$E(V^2_{At}) = 4E\{\sum_i a_i^4 q_{it}^2(1-q_{it})^2 + \sum\sum_{i>j} a_i^2 a_j^2[2q_{it}(1-q_{it})\,q_{jt}(1-q_{jt}) + 4D_{ijt}^2]\} \qquad (4)$$

because terms such as $E[q_{it}(1-q_{it})\,D_{ijt}]$, $E[D_{ijt}D_{ikt}]$ are all zero for populations initially in linkage equilibrium (Hill, 1974*a*). From Crow & Kimura (1970, p. 337)

$$E[q_{it}^2(1-q_{it})^2] = \tfrac{1}{5}q_i(1-q_i)\,[(1-F_t)-(1-F_t)^6] + q_i^2(1-q_i)^2(1-F_t)^6. \qquad (5)$$

The other moments in (4), $E[q_{it}(1-q_{it})\,q_{jt}(1-q_{jt})]$ and $E(D_{ijt}^2)$, can be evaluated using the moment generating matrix of Hill & Robertson (1968). No simple explicit formula for each moment is available, but the results can be simplified providing t/N is not too large. It is shown in Table 1 that, unless linkage is very tight and t/N is large,

$$E[q_{it}(1-q_{it})\,q_{jt}(1-q_{jt})] = q_i(1-q_i)\,q_j(1-q_j)\,(1-F_t)^2 \qquad (6)$$

to a good approximation, which we shall subsequently assume to be adequate. (Mathematically this is equivalent to ignoring the other terms in the moment generating matrix.) Also we shall use Sved & Feldman's (1973) formula for r_{ijt}^2, strictly the expectation of the ratio of D_{ijt}^2 to $q_{it}(1-q_{it})\,q_{jt}(1-q_{jt})$, to approximate the ratio of expectations of these quantities. Their formula is

$$E(r_{ijt}^2) = \frac{1-[(1-1/(2N))\,(1-c_{ij})^2]^t}{(1+(2N-1)2c_{ij}-c_{ij}^2)}, \qquad (7)$$

where c_{ij} is the recombination fraction between loci i and j. Therefore using (6)

$$E(D_{ijt}^2) = E(r_{ijt}^2)\,(1-F_t)^2\,q_i(1-q_i)\,q_j(1-q_j) \qquad (8)$$

where $E(r_{ijt}^2)$ is substituted as in (7). The fit of the approximation (8) is also shown

in Table 1. Bulmer (1976) excluded the term $(1-F_t)^2$ in (8) and the term involving t in (7) and thus his results are less satisfactory unless both t/N is very small and the recombination fraction relatively large.

Using (1), (2), (4), (5), (6) and (8), we obtain the between replicate (line) variance of the within-line variance

$$V(V_{At}) = E(V_{At}^2) - E^2(V_{At})$$
$$= 4\sum_i a_i^4\{\tfrac{1}{5}q_i(1-q_i)\,[(1-F_t)-(1-F_t)^6]+q_i^2(1-q_i)^2\,[(1-F_t)^6-(1-F_t)^2]\}$$
$$+ 16(1-F_t)^2\sum_{i<j}\sum a_i^2 a_j^2 q_i(1-q_i)\,q_j(1-q_j)\,E(r_{ijt}^2). \tag{9}$$

Rewriting (9), and defining $r_{iit}^2 = 1$,

$$V(V_{At}) = 8(1-F_t)^2\sum_i\sum_j a_i^2 a_j^2 q_i(1-q_i)\,q_j(1-q_j)\,E(r_{ijt}^2)$$
$$+4\sum_i a_i^4\{\tfrac{1}{5}q_i(1-q_i)\,[(1-F_t)-(1-F_t)^6]+q_i^2(1-q_i^2)\,[(1-F_t)^6-3(1-F_t)^2]\}. \tag{10}$$

Now let us consider the magnitude of the two sets of terms in (10). If there are n loci affecting the trait there are $\tfrac{1}{2}n(n-1)$ terms involving pairs of loci and n involving single loci. From (7), or more simple direct arguments (Hill & Robertson, 1968), $E(r_{ijt}^2) = 1/(2N)$ at generation one and increases subsequently. In early generations, where terms of $(t/N)^2$ can be ignored, the second term in (10) is $4\sum_i a_i^4 q_i(1-q_i)t/2N$.

Thus, in early generations unless initial gene frequencies are very extreme such that $q_i(1-q_i) \gg q_i^2(1-q_i)^2$, and unless there are very few loci, the term involving pairs of loci in (10) is much larger than the term involving only single loci. This statement will hold also in later generations providing there are many loci, and agrees with Bulmer (1976) that most variation in variance is due to linkage disequilibrium.

Taking, therefore, only the first term in (10), and rewriting,

$$V(V_{At}) = 8(1-F_t)^2\,[\sum_i a_i^2 q_i(1-q_i)]^2\,E(\overline{r_i^2}) + \sum_i\sum_j a_i^2 a_j^2 q_i(1-q_i)\,q_j(1-q_j)\,E(r_{ijt}^2-\overline{r_i^2}), \tag{11}$$

where $\overline{r_i^2}$ is the mean, over all pairs of loci, of r_{ijt}^2. The first term in (11) is simply $2E^2(V_{At})\,E(\overline{r_i^2})$ from (1) and (2). The second term is less simple, but again we neglect it by making further assumptions, perhaps the least tenable of all.

If there is no association between the distance apart of a pair of loci and the product of their contributions to variance of the trait, then on average the second term in (11) will be zero. This assumes that individual traits are not affected by 'blocks' of genes. Thus equation (11) simplifies to

$$\mathrm{var}\,(V_{At}) = 2E(\overline{r_i^2})\,E^2(V_{At}); \tag{12}$$

and the coefficient of variation (CV) of the additive variance is

$$\mathrm{CV}(V_{At}) = \sqrt{(2\overline{r_i^2})}, \tag{13}$$

where in (13), and subsequently, expectation of $\overline{r_i^2}$ is implied. In the first generation $\overline{r_i^2} = 1/(2N)$ for all loci and $\mathrm{CV}(V_{At}) = \sqrt{(2/2N)}$. This can also be obtained directly by considering the estimate of variance in a normal sample of size $2N$.

Multiple alleles. With more than two alleles at each locus, the derivation is much more complicated and we merely outline the proof. Let a_{hit} and q_{hit} be the effect and frequency, respectively, at locus i at time t of allele h and let D_{hifjt} be the disequilibrium between alleles h at locus i and f at locus j at time t. After some manipulation, the variance in a replicate at generation t (cf. equation (3)) can be shown to be

$$V_{At} = \sum_i \sum_{h \neq k} (a_{hi} - a_{ki})^2 q_{hit} q_{kit} + 4 \sum_{i<j} \sum_h \sum_f a_{hi} a_{fj} D_{hifjt}.$$

Expectations of terms in V_{At}^2 can be computed using formulae given by Hill (1975) and ones analogous to those given in Hill (1974a). The same approximations are then used as in the two allele proof (i.e. neglecting single loci terms and assuming no association between distance between loci and the product of their contributions to the variance of the trait). Then equation (12) is obtained as before, i.e.

$$\mathrm{var}\,(V_{At}) = 2E(\overline{r_i^2})\,E^2(V_{At}),$$

where $\overline{r_i^2}$ is now defined as the average over all pairs of loci of r_{ijt}^2 and $E(r_{ijt}^2)$ is taken as

$$E(r_{ijt}^2) = \frac{E[\sum_h \sum_k D_{hifjt}^2]}{E[\sum_{h \neq k} q_{hit} q_{kit} \sum_{f \neq g} q_{fjt} q_{gjt}]}.$$

As with the two allele formulation, this is an approximation, for from Hill (1975),

$$E(r_{ijt}^2) = E\left[\frac{\sum_h \sum_f D_{hifjt}^2}{\sum_{h \neq k} q_{hit} q_{kit} \sum_{f \neq g} q_{fjt} q_{gjt}}\right].$$

Similarly, formulae given subsequently for autocorrelations of variance and other quantities also hold whether there are two or many alleles at each locus.

(ii) *Autocorrelation of variance*

Since linkage disequilibrium generated in a replicate line is not expected to immediately return to zero, there will be a correlation of the additive variance over generations within a replicate. From standard statistical theory, for generations t and $t+k$, where $k > 0$,

$$E(V_{At} V_{At+k}) = E(V_{At} E(V_{At+k}|V_{At})). \tag{14}$$

Using equation (3),

$$E(V_{At+k}|V_{At}) = E[4\sum_{i<j} a_i a_j D_{ijt+k}|V_{At}] + \left(1 - \frac{1}{2N}\right)^k 2\sum_i a_i^2 q_{it}(1 - q_{it}). \tag{15}$$

From Hill & Robertson (1968),

$$E(D_{ijt+k}|D_{ijt}) = (1-c)^k \left(1 - \frac{1}{2N}\right)^k D_{ijt},$$

which, when used in (15) and then (15) substituted in (14), gives

$$E(V_{At} V_{At+k}) = E\left[V_{At}\left(1 - \frac{1}{2N}\right)^k \left(2\sum_i a_i^2 q_{it}(1 - q_{it}) + 4\sum_{i<j} a_i a_j D_{ijt}(1 - c_{ij})^k\right)\right]. \tag{16}$$

By using exactly the same approach and approximations as were used to derive $V(V_{At})$, (16) gives

$$\text{cov}\,(V_{At},\,V_{At+k}) = \left(1 - \frac{1}{2N}\right)^k 2E^2(V_{At})\,E(\overline{r_i^2(1-c)^k}), \qquad (17)$$

where $\overline{r_i^2(1-c)^k}$ is the average over all pairs of loci of $r_{ijt}^2(1-c_{ij})^k$ for specified k.

From equation (2), $E(V_{At+k}) = (1 - 1/(2N))^k\,E(V_{At})$ and thus the correlation of variances between generations t and $t+k$ is

$$
\begin{aligned}
\text{corr}\,(V_{At}V_{At+k}) &= \frac{\text{cov}\,(V_{At}V_{At+k})}{[\text{var}\,(V_{At})\,\text{var}\,(V_{At+k})]^{\frac12}} \\[2mm]
&= \frac{(1 - 1/(2N))^k\,2E^2(V_{At})E(\overline{r_i^2(1-c)^k})}{[2E^2(V_{At})\,E(\overline{r_i^2}) \times 2E^2(V_{At})\,(1 - 1/(2N))^{2k}\,E(\overline{r_{i+k}^2})]^{\frac12}} \\[2mm]
&= \frac{E(\overline{r_i^2(1-c)^k})}{[E(\overline{r_i^2})\,E(\overline{r_{i+k}^2})]^{\frac12}}.
\end{aligned} \qquad (18)
$$

Before proceeding further, we should review the important assumptions used to obtain (12) and (18): there is initial linkage equilibrium and Hardy–Weinberg proportions are assumed to be maintained; there is no selection; the total inbreeding is not very high, say $t < N$ generations (equivalent to $F = 40\,\%$); the number of loci affecting the trait is large; and there is no association between map distance and the product of gene frequencies and effects for pairs of loci. Fewer assumptions are required for earlier equations. Bulmer (1976) proved that the departure from Hardy–Weinberg proportions of the within line variance was small and dependent on number of loci. Thus as the number of loci increases it will become negligible as with other single loci effects, compared to the effect of linkage disequilibrium. Also it is independent from generation to generation and thus Hardy–Weinberg departure will not introduce a strong autocorrelation as does disequilibrium (cf. equation 18).

3. EVALUATION OF FORMULAE

An expression for r_{ijt}^2 is given in (7), from which it can be shown that for unlinked loci

$$r_{ijt}^2 = \frac{2}{3N + \frac12}\left(1 - \left[\frac14\left(1 - \frac{1}{2N}\right)\right]^t\right) \sim \frac{2}{3N}\left(1 - \frac{1}{4^t}\right) \qquad (19)$$

unless N is very small. Regardless of the tightness of linkage, $r_1^2 = 1/(2N)$ and subsequently $\overline{r_i^2}$ is not less than that which applies if all loci are unlinked; thus $r_2^2 \geqslant 5/(8N)$ and $\overline{r_i^2}$ reaches an asymptote of at least $2/(3N)$ very quickly. So after 3 or 4 generations and if any loci are linked, $\text{CV}(V_{At}) = \sqrt{(2r_i^2)} > \sqrt{(4/(3N))}$ or 0·52, 0·37, 0·26, 0·16 and 0·12 for $N = 5$, 10, 20, 50 and 100.

Let us now consider loci on the same chromosome and assume as previously that that there are very many (formally infinitely many) loci and that position has no relation to effect. Let us also assume that these loci are distributed uniformly along the chromosome with no interference. If the chromosome has map length l morgans, the density function of distance x between pairs of loci is the triangular distribution

$f(x) = 2(l - x)/l^2$. The relation between recombination fraction and map distance is $c = (1 - e^{-2x})/2$. Using these two relations and (7), the mean value of r^2 can be computed for single chromosomes.

Extending further to consider m chromosomes, let the kth chromosome have map length l_k, and the total map length be $L = \Sigma l_k$. Thus a proportion

$$\Sigma l_k^2/L^2 = [1 + \text{CV}^2(l_k)]/m$$

of pairs of loci are on the same chromosome, and the remainder on different chromosomes with a recombination fraction between them of $\frac{1}{2}$. Taking all the relevant formulae together, we obtain

$$\overline{r_t^2} = \frac{2}{L^2} \sum_{k=1}^{m} \int_0^{l_k} \frac{1 - [(1 - 1/(2N))(1 - (1 - e^{-2x})/2)^2]^t}{1 + (2N - 1)[(1 - e^{-2x}) - \frac{1}{4}(1 - e^{-2x})^2]} (l_k - x)\, dx$$

$$+ \frac{1}{L^2}\left(1 - \sum_{k=1}^{m} l_k^2\right) \frac{2}{3N + \frac{1}{2}} \left[1 - \frac{1}{4^t}\left(1 - \frac{1}{2N}\right)^t\right]. \tag{20}$$

This formula has been evaluated by numerical integration. A little insight can be obtained, however, by considering t large (when strictly some other assumptions are violated) and individual chromosomes sufficiently short that Nl_k is small, i.e. $Nl_k \ll 1$ and terms in l_k^2 can be ignored relative to l_k. Then

$$\lim_{t \to \infty} \overline{r_t^2} = \frac{1}{L^2} \sum_k \frac{(4Nl_k + 1)\log(4Nl_k + 1)}{8N^2} - \frac{1}{2NL} + \frac{2}{3N}\left(1 - \frac{\Sigma l_k^2}{L^2}\right). \tag{21}$$

Equation (21) can be improved by removing some restrictions, but becomes more complicated. In general it shows that the asymptotic value of $\overline{r_t^2}$, and hence of $\text{CV}(V_{At})$, is a function of Nl_k, but this holds only if Nl_k and l_k are small enough.

Two models of distribution of chromosome lengths have been used. In the first all chromosomes are assumed to be of equal length, i.e. $l_k = L/m$. In the second the chromosome lengths are given relative lengths $1, 2, \ldots$, up to m. This leads to

$$l_k = 2kL/[m(m + 1)] \tag{22}$$

so that with $m = 4$, for example, and $\bar{l} = L/4 = 1$, the lengths are 0·4, 0·8, 1·2 and 1·6. The proportion of pairs of loci on the same chromosome is $1/m$ in the first model and $\frac{2}{3}[(2m + 1)/(m(m + 1))]$ (which tends to $4/(3m)$ with large m) in the second. Values of $\overline{r_t^2}$ for both models are given in Table 2. They show that there is little difference in results between them, so in the following calculations we shall assume that the chromosome lengths vary as in model 2.

4. RESULTS

(i) *Coefficient of variation of additive variance.* Results for $\text{CV}(V_{At})$ for a range of values of chromosome number and total map length with fixed population size ($N = 20$) are given in Fig. 1, and for different population sizes but fixed total map length in Fig. 2. The case of all loci unlinked is shown in Fig. 1 as $L \to \infty$, $m \to \infty$, in which case $\text{CV}(V_{At})$ soon reaches $\sqrt{(4/(3N))} = 0·26$. The figures also show that if there

P. J. AVERY AND W. G. HILL

are many chromosomes, as in mammals for example, the value of $\mathrm{CV}(V_{At})$ given for unlinked loci does not greatly underestimate the correct value. The total map length and the number of chromosomes both influence the result, but these have a smaller effect than does population size, at least when there is more than one chromosome. Only if there is one chromosome is an asymptotic value for $\mathrm{CV}(V_{At})$ not approached rapidly, and as illustrated by (19), the time at which this asymptote is reached depends little on population size.

Table 2. *Values of* $100\overline{r_t^2}$ *for* $N = 20$ *and two models of distribution of chromosome length,* (i) *equal,* (ii) *variable (equation 22). There are* m *chromosomes of total map length* L

		$t^* =$	2		5		10		20	
L	m		(i)	(ii)	(i)	(ii)	(i)	(ii)	(i)	(ii)
1	2		3·73	3·76	5·55	5·62	6·95	7·02	8·16	8·24
	5		3·42	3·47	4·56	4·72	5·60	5·80	6·65	6·87
	20		3·20	3·22	3·70	3·80	4·12	4·30	4·70	4·95
2·5	2		3·52	3·52	4·60	4·61	5·25	5·26	5·76	5·77
	5		3·36	3·38	4·20	4·26	4·76	4·83	5·25	5·31
	20		3·19	3·21	3·65	3·72	3·98	4·09	4·34	4·47
10	2		3·25	3·25	3·71	3·71	3·88	3·88	4·02	4·02
	5		3·23	3·23	3·67	3·67	3·83	3·84	3·96	3·97
	20		3·17	3·18	3·53	3·55	3·67	3·69	3·79	3·82

* At $t = 1$, $100\overline{r_t^2} = 2\cdot5$ for both models.

For a much wider range of parameters than can be shown in the figures, the asymptotic values of $\mathrm{CV}(V_{At})$ are given in Table 3. These will be rather poorer approximations, if there is tight linkage, for a large number of generations are required before they are reached and some of the assumptions made previously are invalidated.

(ii) *Checks by simulation.* Some checks of the approximations made in deriving $\mathrm{CV}(V_{At})$ have been made by Monte Carlo simulation. The model comprised 20 loci, assumed to be on one chromosome with various combinations of map distances, effects and initial frequencies, or 20 loci each on different chromosomes. One hundred replicates of each simulation were run for five generations, computing costs limiting further runs or generations. In Table 4 the results from simulation are compared with predictions given by (13) and (20). The fit is seen to be excellent.

(iii) *Autocorrelation of variance.* In a replicate line, in which in one generation the variance is higher than expectation, due to an excess of positive linkage dis-equilibrium, it is likely to remain above expectation in the following generations, since, even with unlinked loci, only one-half of the disequilibrium is lost each generation. A formula for the autocorrelation of genetic variances k generations apart is given by (18). If all loci are unlinked, (18) reduces to

$$\mathrm{corr}\,(V_{At}, V_{At+k}) = \left(\frac{1}{2}\right)^k \overline{(r_t^2/r_{t+k}^2)^{\frac{1}{2}}}, \tag{23}$$

which very quickly asymptotes to $(\frac{1}{2})^k$ as t increases. In general, we expect the auto-correlation to exceed this value if there are many closely linked loci except when t is small and var (V_{At}) is still increasing rapidly (cf. Fig. 1).

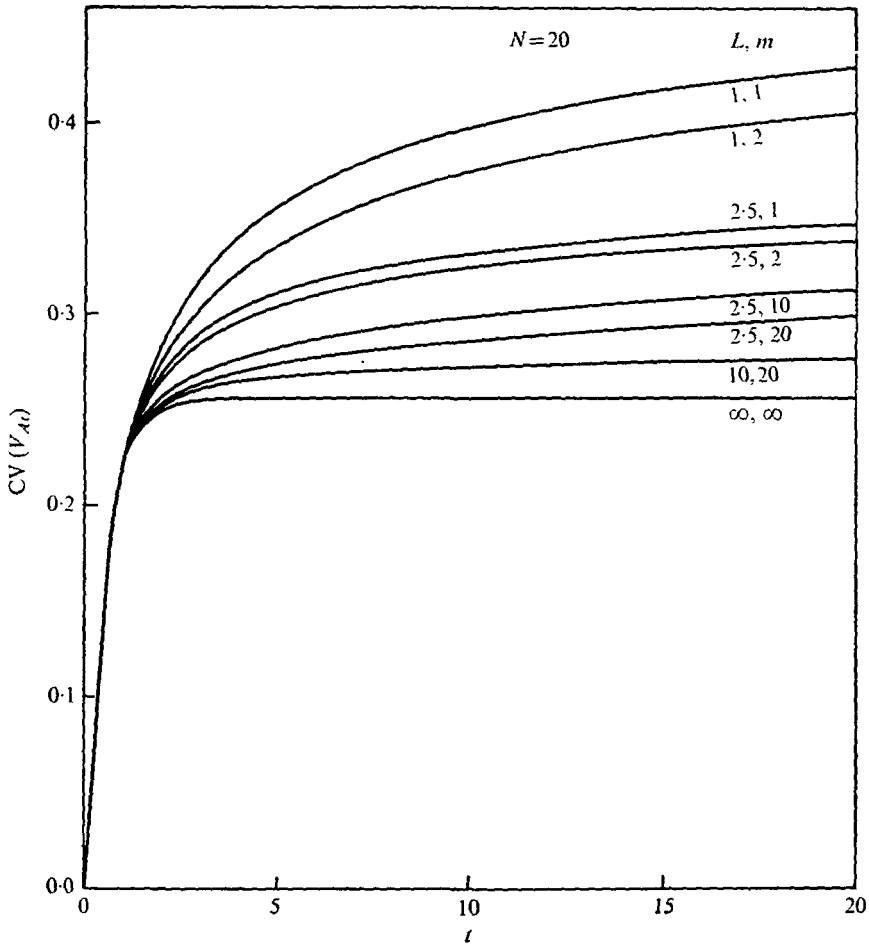Fig. 1. The coefficient of variation of the genetic variance (i.e. CV(V_{At})) is plotted against time for various values of the total map length, L, and the number of chromosomes, m. The population size, N, of each replicate equals 20.

Examples of the autocorrelation are given in Table 5 for a range of parameter values. As loci become more tightly linked the autocorrelation increases proportionately more as the number of generations apart, k, increases. Thus with a few chromosomes and a total map length of only 1 or 2, the autocorrelation may approach 40 %. This is because corr (V_{At}, V_{At+k}) depends on a weighted average of r_i^2; the pairs of loci more closely linked have both a higher value of r_i^2 and $(1-c)^k$.

The magnitude of the autocorrelation of additive variance is changed remarkably little as population size changes if other parameters remain constant (see Table 5), even though the variance and the coefficients of variation of the additive variance

14

are highly dependent on population size (cf. Fig. 2). This invariance with N begins to break down as t becomes large as is seen by the asymptotic results. The invariance arises because, for small t, $r_t^2 \simeq f(c)/2N$ where $f(c)$ is a function of the recombination fraction alone. Thus $\overline{r_t^2} \simeq A/2N$ and $\overline{r_t^2}(1-c)^k \simeq B/2N$ where A and B are independent of N, and when corr (V_{At}, V_{At+k}) is computed from (18), the terms in N
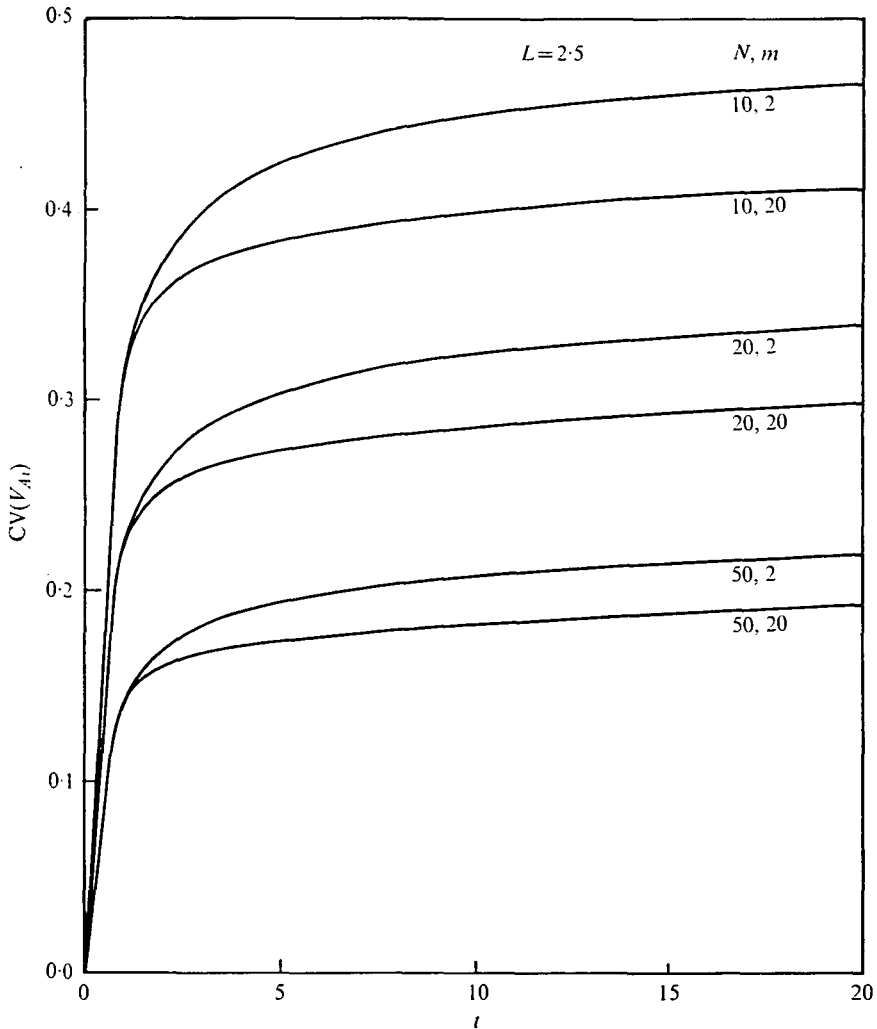

Fig. 2. The coefficient of variation of the genetic variance (i.e. CV(V_{At})) is plotted against time for various values of the population size, N, and the number of chromosomes, m, keeping the total map length, L, constant.

cancel out, i.e. corr $(V_{At}, V_{At+k}) = B(t)/[A(t)\,A(t+k)]^{\frac{1}{2}}$. Table 5 also clearly shows that, as k increases, the speed of convergence to the limit decreases considerably. Correlations can still be very high even for large k, e.g. for $k = 10$, $N = 20$, $L = m = 1$ (one chromosome of length one morgan) the correlations are 11·0 %, 24·0 % and 45·6 % for $t = 1$, 5 and $\to \infty$.

Table 3. *Asymptotic values of* $\mathrm{CV}(V_{At}) = \sqrt{(2\overline{r^2})}$ *as* $t \to \infty$, *using the model of different chromosome lengths*

N	L	NL	$m = 1$	2	5	10	20	$\to \infty$
20	0·1	2	0·873	0·755	0·604	0·501	0·416	0·257
20	1·0	20	0·460	0·439	0·405	0·376	0·347	0·257
20	2·5	50	0·363	0·356	0·343	0·330	0·316	0·257
20	10	200	0·289	0·288	0·287	0·285	0·281	0·257
10	5	50	0·431	0·428	0·420	0·413	0·403	0·362
50	1	50	0·321	0·308	0·286	0·268	0·248	0·163

Table 4. *Simulation (O) and Predicted (P) results for* $\mathrm{CV}(V_{At})$ *% for different models of effect, initial gene frequencies and spacing of loci. Twenty loci are considered, distributed on a single chromosome.* $N = 20$

	L	Effects	Spacing	Initial frequencies	1	2	3	4	5
P	0·2	Any	Any	Any	22	31	36	40	44
O	0·2	Equal	Equal	0·5	24	32	37	43	45
P	$\to\infty$	Any	Any	Any	22	25	26	26	26
O	$\to\infty$	Equal	Equal	0·5	22	25	26	28	25
P	2·0	Any	Any	Any	22	27	30	31	32
O	2·0	Equal	Equal	0·5	24	26	29	32	31
O	2·0	Equal	Equal	Uniform* (0·1–0·9)	24	27	27	32	35
O	2·0	Uniform* (0–9)	Equal	0·5	24	30	35	35	35
O	2·0	Equal	†Two blocks	0·5	23	28	33	32	34
O	2·0	Equal	‡Four blocks	0·5	22	26	28	33	33
O	2·0	Exponential*	Equal	0·5	21	26	29	30	27

The table headers: **Model** spans Effects, Spacing, Initial frequencies; **Time (t)** spans 1, 2, 3, 4, 5.

* Sampled.

† Two blocks of 10 genes each with $c = 0·05$ between adjacent loci and $c = 0·439$ between blocks.

‡ Four blocks of five genes each with $c = 0·05$ between adjacent loci and $c = 0·269$ between blocks.

5. EXTENSION OF FORMULAE TO EXAMINE THE VARIANCE OF OTHER DERIVED PARAMETERS

(i) *Heritability.* Variances between replicates of within replicate heritability values and responses can be obtained, at least approximately, from the results on additive variance. If it is assumed that the environmental variance (V_E) remains constant

$$V(h_t^2) = V[V_{At}/(V_{At} + V_E)].$$

Using Taylor's approximation,

$$V(h_t^2) \simeq \left[\frac{\mathrm{d}h^2}{\mathrm{d}V_{At}}\right]^2 V(V_{At}) = \frac{V_E^2}{[E(V_{At}) + V_E]^4} \, V(V_{At})$$

$$= \mathrm{CV}^2(V_{At}) \, E^2(h_t^2) \, E^2[1 - h_t^2]]$$

approximately. Removing the expectation signs to simplify,

$$V(h_t^2) = 2\overline{r_t^2}[h_t^2(1-h_t^2)]^2.$$

If the total inbreeding is low, $h_t^2 \sim h^2$, and

$$SD(h_t^2) \sim \sqrt{(2\overline{r^2})}\, h^2(1-h^2). \tag{24}$$

Table 5. *Autocorrelation (%) of additive genetic variance between generations t and $t+k$ as a function of population size, N, total map length, L, and number of chromosomes, m*

			$k =$	1		2			4		
N	L	m	$t = 1$	5	$\to\infty$	1	5	$\to\infty$	1	5	$\to\infty$
20	0·2	1	68·8	88·8	96·3	54·7	80·0	92·9	40·2	66·9	86·8
	1	1	61·7	79·6	87·7	44·2	65·2	78·3	27·0	46·7	65·1
		2	56·6	75·3	85·5	38·3	60·2	75·9	22·6	43·4	64·1
	2·5	1	54·9	70·0	77·8	35·1	51·6	63·6	17·8	32·0	47·5
		2	53·0	67·9	76·4	32·9	49·3	62·1	16·3	30·6	46·7
		20	46·7	57·0	67·7	24·8	35·4	51·0	9·1	18·5	37·3
	10	20	46·2	54·4	58·9	23·9	31·2	37·7	7·7	12·9	20·6
	> 0	$\to\infty$	44·8	50·0	50·0	21·9	25·0	25·0	5·4	6·2	6·2
10	2·5	2	53·2	67·8	74·1	33·2	49·2	58·6	16·5	30·4	42·1
		20	46·8	56·9	64·3	24·9	35·1	45·9	9·0	18·2	31·0
50	2·5	2	52·9	68·0	79·0	32·8	49·4	66·2	16·2	30·7	52·3
		20	46·6	57·1	71·8	24·7	35·5	57·2	9·0	18·7	45·0

For $h^2 = 0·5$, for example $SD(h_t^2) \sim \frac{1}{4}CV(V_{At})$ and results can be obtained from Figs. 1 and 2; the values of $SD(h_t^2)$ will be a smaller proportion of $CV(V_{At})$ at more extreme heritability values. Alternatively $CV(h_t^2) \sim \sqrt{(2\overline{r^2})}\,(1-h^2)$. The auto-correlation of heritability values in successive generations will be given by formulae similar to (18).

(ii) *Response.* The response at generation t from artificial selection of intensity $\bar{\imath}$ standard deviations is, with the same assumptions as before,

$$R_t = \bar{\imath}V_{At}/\sqrt{(V_{At}+V_E)}.$$

The same procedure can be used to compute the variation in response due to variation in variance within lines. This is not the variation in response due to sampling of mean gene frequency in the usual drift sense.

Thus,

$$V(R_t) \simeq \bar{\imath}^2(1-\tfrac{1}{2}h_t^2)^2\, V(V_{At})/(V_{At}+V_E)$$

and

$$CV(R_t) \simeq (1-\tfrac{1}{2}h_t^2)\,CV(V_{At}) \tag{25}$$

$$\sim (1-\tfrac{1}{2}h^2)\sqrt{(2\overline{r_t^2})}. \tag{26}$$

If heritability is low, the coefficient of variation of response roughly equals that of the additive variance, in excess of 26 % with $N = 20$ for example after a few generations.

(iii) *Cumulative variance and response.* Motivated by a need to compute the variance of cumulative response to several generations of selection, which is a function of the sum of the additive variances over generations, let us consider the variation between replicates in these quantities. Thus, let $V_{AT}^* = \sum_{T=0}^{t} V_{AT}$ denote the cumulative variance up to generation t, which includes $t+1$ terms. Now

$$V(V_{At}^*) = \sum_T V(V_{AT}) + 2\sum\sum_{T'<T} \text{cov}\,(V_{AT}, V_{AT'})$$

the parts of which we have already evaluated. For presentation, the results have been standardized as $\text{CV}(V_{At}^*)$, where

$$E(V_{At}^*) = 2N[1 - (1-1/(2N))^{t+1}]\,V_A.$$

Since the convention in this analysis has been to take the additive variance at generation 0 as V_A, but with $V(V_{A0}) = 0$, $\text{CV}(V_{A1}^*) = [V(V_{A1})]^{\frac{1}{2}}/[V_A + E(V_{A1})]$, i.e. there are more terms in the denominator than numerator. An alternative would have been to take the sampled first generation as the base point, giving

$$\text{CV}(V_{A1}^*) = \text{CV}(V_{A1}),$$

since no conceptual infinite population at generation 0 could be sampled. Assuming N to be reasonably large and ignoring the changes in $E(V_{At})$ due to inbreeding, this alternative method of presenting values of $\text{CV}(V_{At}^*)$ would have given results roughly $(t+1)/t$ times as great.

Examples of $\text{CV}(V_{At}^*)$ are given in Fig. 3. With unlinked loci $\text{CV}(V_{At}^*)$ reaches a maximum in very few generations and then declines; this decline is not due to a reduction in $V(V_{At}^*)$ but to the fact that it tends to asymptote quickly because covariances among generations a long way apart become very small, while $E(V_{At}^*)$ continues to rise. With $N = 20$ the maximum in $\text{CV}(V_{At}^*)$ is near 14 %, or $\text{CV}^2(V_{At}^*)$ is one-half of the maximum of $\text{CV}^2(V_{At})$ for single generations. This holds also approximately for other population sizes.

The variation in cumulative response due to variation in within line variance up to generation $t+1$ depends on $V(V_{At}^*)$, and cf. (25)

$$\text{CV}(R_t^*) \sim (1 - \tfrac{1}{2}h^2)\,\text{CV}(V_{At}^*). \tag{27}$$

These arguments are pursued further by Hill (1976), and the contribution of this source of variation in response relative to that from drift in mean is discussed, although no exact answers can be given.

(iv) *Genetic covariance.* The analysis presented above can readily be extended to the variances of genetic covariances and correlations between two traits if both of them can be assumed to be genetically determined solely by additive genes. Assume locus i has effect a_i on trait X (as before) and b_i on trait Y. The genetic covariance between traits X and Y is, at generation t,

$$\text{cov}_{At} = 2\sum_i a_i b_i q_{it}(1 - q_{it}) + 2\sum\sum_{i<j} (a_i b_j + a_j b_i)\,D_{ijt} \tag{28}$$

in which the first terms in (28) are due to pleiotropy and the second to disequi-

librium. We assume, as before, that $D_{ij0} = 0$, i.e. there is initial equilibrium, and denote the initial covariance as cov_A, in which case

$$E(\mathrm{cov}_{At}) = (1 - F_t)\,\mathrm{cov}_A \tag{29}$$
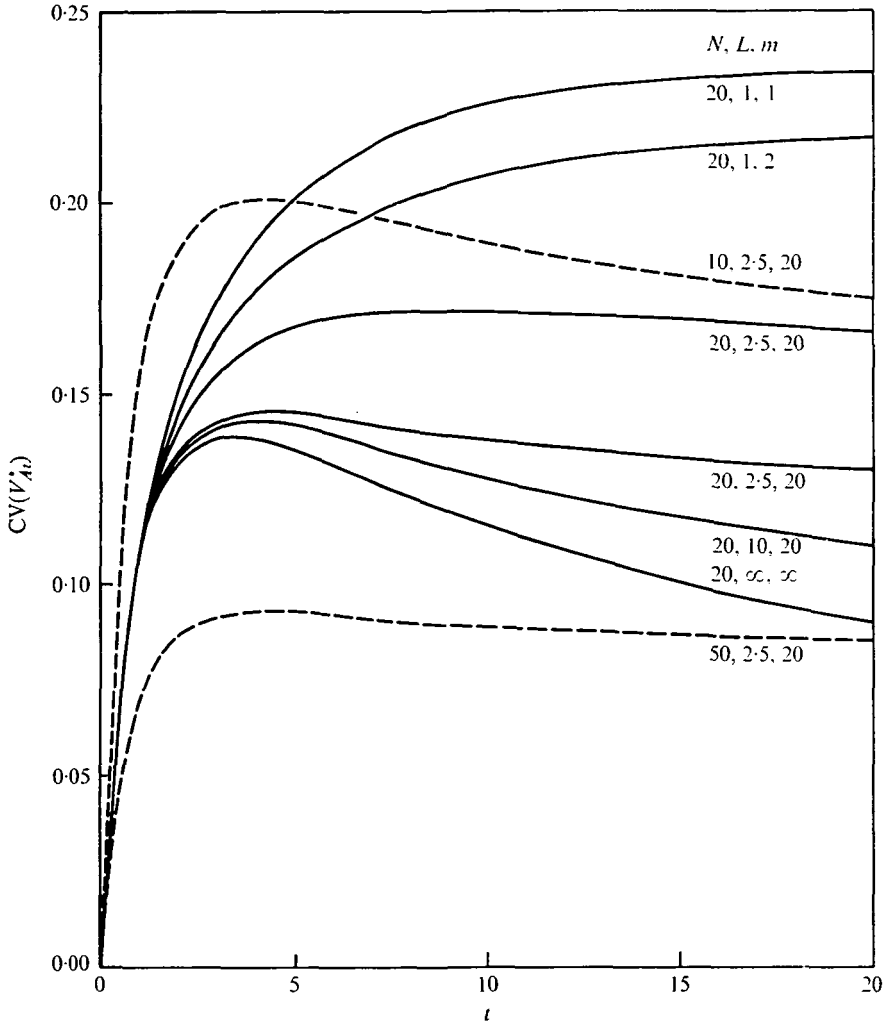
Fig. 3. The coefficient of variation of the genetic variance (i.e. $\mathrm{CV}(V_{At}^*)$) is plotted against time for various values of the population size, N, the number of chromosomes, m, and the total map length, L.

as in (2). Ignoring terms involving single loci, as in the derivation of $V(V_{At})$,

$$V(\mathrm{cov}_{At}) = 2(1 - F_t)^2 \sum_i \sum_j (a_i b_j + a_j b_i)^2 q_i(1 - q_i)\, q_j(1 - q_j)\, r_{ijt}^2.$$

Assuming, as before, no correlation between crossover probability and effects,

$$V(\mathrm{cov}_{At}) = 2(1 - F_t)^2 \overline{r_t^2}[\sum_i \sum_j 2a_i^2 b_j^2 q_i(1 - q_i)\, q_j(1 - q_j)$$
$$+ \sum_i \sum_j 2a_i a_j b_i b_j q_i(1 - q_i)\, q_j(1 - q_j)]$$
$$= (1 - F_t)^2 \overline{r_t^2}(V_{AX} V_{AY} + \mathrm{cov}_A^2), \tag{30}$$

where V_{AX} and V_{AY} are the initial additive genetic variances in traits X and Y. The terms in (30) in generation 1, when $\overline{r^2} = 1/2N$, are the extension to a covariance of formulae for sampling variances of variance estimates (Tallis, 1959).

For traits which are initially uncorrelated, the variance in cov_{At} depends solely on the products of variances at the two loci, in which case $\mathrm{CV}(\mathrm{cov}_{At})$ is infinite. For initially correlated traits, using (29) and (30),

$$\mathrm{CV}(\mathrm{cov}_{At}) = (1/\rho^2 + 1)\overline{r_t^2}, \tag{31}$$

where $\rho = \mathrm{cov}_A/\sqrt{(V_{AX}V_{AY})}$ is the initial genetic correlation between the traits. Equation (31) shows that $\mathrm{CV}(\mathrm{cov}_{At})$ is at least as big as $\mathrm{CV}(V_{At})$, becoming much larger as the original genetic correlation becomes smaller.

The phenotypic covariance (cov_{Pt}) will, if the environmental covariance remains constant, have the same variance as does cov_{At}. Its coefficient of variation will be smaller if the environmental and genetic covariances are of the same sign.

(v) *Correlations.* Providing that the coefficients of variation of V_{AXt} and V_{AYt} are not too large, the following approximate formula for $V(\rho_t)$ can be obtained using Taylor's series.

$$V(\rho_t) = \left(\frac{\partial \rho_t}{\partial \mathrm{cov}_{At}}\right)^2 V(\mathrm{cov}_{At}) + 2\frac{\partial \rho_t}{\partial \mathrm{cov}_{At}}\frac{\partial \rho_t}{\partial V_{AXt}}\mathrm{cov}(\mathrm{cov}_{At}, V_{AXt}) + \dots + \left(\frac{\partial \rho_t}{\partial V_{AYt}}\right)^2 V(V_{AYt}). \tag{32}$$

Using normal theory, and extrapolating to our situation as in (30),

$$\mathrm{cov}\,(V_{AXt}, V_{AYt}) = 2(1 - F_t)^2\overline{r_t^2}\,\mathrm{cov}_A^2,$$

$$\mathrm{cov}\,(V_{AXt}, \mathrm{cov}_{At}) = 2(1 - F_t)^2\overline{r_t^2}\,V_{AX}\,\mathrm{cov}_A,$$

and similarly for other terms. Substituting in (32), we have

$$V(\rho_t) = \overline{r_t^2}(1 - \rho^2)^2. \tag{33}$$

The formula for the variance of the phenotypic correlation is less appealingly simple, except when $r_P = r_G = 0$ initially, when its variance is $h_X^2 h_Y^2 \overline{r_t^2}$, as could be guessed from (33).

For traits in which the genetic correlation is small, $V(\rho_t) \sim \overline{r_t^2}$ and thus is in excess of $2/(3N)$ (the asymptotic value of $\overline{r_t^2}$ for unlinked loci). For $N = 20$, for example, the standard deviation of ρ_t must exceed $\sqrt{\frac{1}{30}} = 0 \cdot 18$. Thus replicate populations could have very different genetic correlations.

(vi) *Correlated responses.* When selection is practised on trait X, the correlated response on trait Y is expected to be, at generation t,

$$Q_{Yt} = \tilde{\imath}\,\mathrm{cov}_{At}/\sqrt{(V_{AXt} + V_{EX})}. \tag{34}$$

Similar calculations to those above give

$$V(Q_{Yt}) = \frac{\tilde{\imath}^2(1 - F_t)^2\overline{r_t^2}}{E(V_{AXt}) + V_{EX}}\{V_{AX}V_{AY} + \mathrm{cov}_A^2[2(1 - \tfrac{1}{2}E(h_{Xt}^2))^2 - 1]\} \tag{35}$$

and, for $\rho > 0$, the coefficient of variation of the correlated response

$$\mathrm{CV}(Q_{Yt}) = [\overline{r_t^2}(1/\rho^2 + 1 - 2h_X^2 + \tfrac{1}{2}h_X^4)]^{\frac{1}{2}} \tag{36}$$

assuming $E(h^2_{Xt})$ approximately equals the initial heritability h^2_X. Thus if the initial correlation is low, the correlated response in Y has a much higher coefficient of variation than the direct response in X (cf. 26).

6. DISCUSSION

In this paper some first attempts have been made to derive a theory to describe the variation among small populations of quantities such as the genetic variance, heritability and genetic correlation of quantitative traits. This extends previous theory which dealt essentially either with variation in mean performance between populations, or with mean variance within populations. Following Bulmer (1976) the variation in variance is argued to be due almost entirely to linkage disequilibrium, assuming the trait is influenced by several loci, so there are many more pairs of loci than single loci. Fortunately there is available a large theory for pairs of linked loci in population genetics which could be utilised in the extension to quantitative traits.

The populations have been assumed to be initially in linkage equilibrium. With no selection, the mean disequilibrium over replicate populations remains zero in each successive generation, and the disequilibrium itself tends to zero in all replicates as fixation occurs. In the segregating populations, however, there is a considerable variation among replicates in the amount of disequilibrium which occurs, the variation being greater for tightly linked loci but not negligible even for pairs of loci on different chromosomes. This variance in disequilibrium, summed over all pairs of loci affecting the trait, is thus shown to induce a considerable variation among replicate populations in the variances and other second moments measured within populations. Although the disequilibrium is transient, it does not on average decline to zero in a single generation, the reduction being not more than one-half (the value for unlinked loci). Therefore there is a correlation of at least one-half between the genetic variances in successive generations, a consequence of the disequilibrium perhaps equally important to the variation in variance. Estimates of variances, heritabilities and correlations made in several generations from a single population are therefore likely to be unreliable predictors of these quantities in other (conceptual) replicate populations, or in the base population from which the replicate was drawn.

Initial disequilibrium

If the base population is not initially in linkage equilibrium, the variance among replicates in additive variance will differ from that given previously, especially in early generations. The expected variance will not change in proportion to the inbreeding coefficient (cf. (2)) and could increase if there were much negative disequilibrium; and the formulae given for variance (e.g. (4)) in variance are also incorrect. With unlinked loci, however, the expected disequilibrium is halved each generation in a random mating population (and becomes zero in the F2 of the cross of two inbred lines) and the equilibrium value of $r^2 = 2/(3N)$ is reached very quickly,

essentially within four generations. Therefore for mammals where most pairs of genes are on different chromosomes, the variance in additive variance will very soon approach the same value, $2\overline{r^2}V_{At}^2$, whether or not the population is initially in linkage equilibrium. With a high proportion of tightly linked loci, this value will be approached more slowly, especially for populations such as two-way crosses of inbred lines in which there is extreme initial disequilibrium.

Table 6. *General table for two particular examples* ($N = 10$)

Variable	Equation source	Mouse			Drosophila		
		$t = 1$	5	10	1	5	10
$E(V_{At})$	(2)	$0.95V_A$	$0.774V_A$	$0.599V_A$	$0.95V_A$	$0.774V_A$	$0.599V_A$
Var (V_{At})	(12)	$0.090V_A^2$	$0.083V_A^2$	$0.051V_A^2$	$0.090V_A^2$	$0.146V_A^2$	$0.103V_A^2$
$CV(V_{At})$	(13)	0.316	0.372	0.377	0.316	0.495	0.537
corr (V_{At}, V_{At+1})	(18)	0.462	0.536	0.547	0.510	0.674	0.715
$CV(V_{At}^*)$	(30), (31)	0.154	0.194	0.174	0.154	0.278	0.302
$SD(h_t^2), h^2 = \frac{1}{2}$	(26)	0.079	0.093	0.094	0.079	0.124	0.134
$SD(h_t^2), h^2 = \frac{1}{4}$	(26)	0.059	0.070	0.071	0.059	0.093	0.101
$SD(\rho_t)\ \rho = 0$	(39)	0.224	0.263	0.267	0.224	0.350	0.380
$SD(\rho_t)\ \rho = \frac{1}{2}$	(39)	0.167	0.197	0.200	0.167	0.263	0.285

Examples

The effects of population size, map length and number of chromosomes, separately or together, have been shown in the tables and graphs. In order to gain some feeling for the magnitude of these effects in a practical context, we consider two examples, mouse and *Drosophila*. The mouse has ($m =$) 20 chromosome pairs and the total map length (L) has been estimated as 14 morgans, obtained by summing map lengths between the most distant mapped genes on each chromosome and rounding upwards to allow for unmarked chromosome ends (from map of J. Womack (1976)). Using chiasma frequency data the total map length of the mouse is around 20 morgans (Green, 1966), but the former figure was used. No correction for lack of crossing over in the X chromosome in males was made. The mouse example is probably typical of most domestic mammals which have large chromosome numbers but of unknown map length, although for man a figure of 23 morgans has been given from chiasma frequency (Strickberger, 1968, p. 345). In the other example, that of *Drosophila melanogaster*, the total map length is taken as 2·84 over 3 chromosomes, using the linkage maps of Lindsley & Grell (1967). As there is no crossing over in males, the relation between mean recombination fraction and distance apart, x, of pairs of loci is $\frac{1}{4}(1-e^{-2x})$ rather than $\frac{1}{2}(1-e^{-2x})$ as used in previous formulae and the results have been computed accordingly. Although rather an unusual example, *Drosophila* is included as it is so widely used experimentally. Results are given in Table 6 using an effective population size of 10. For example, with an initial heritability of 25 %, the standard deviation of heritability between replicates is predicted to reach about 7 % with the mouse and 10 % with *Drosophila* after 10 generations, most or all of these values being reached in 5 generations. The values given in both cases assume the unequal distribution (equation 22) of

chromosome lengths, but there would be little difference if equal or the actual lengths were used (cf. Table 2). Nor would the mouse results be substantially affected if a total map length of 20 morgans was used, or even if all loci were assumed to be on different chromosomes. If a population size of 40 rather than 10 were used, the coefficients of variation and standard deviations shown in Table 6 would be roughly halved and the correlation, corr $(V_{\Delta t}, V_{\Delta t+1})$, little affected.

Assumptions

A large number of assumptions have been made in obtaining the results which have been presented, essentially so that the answers could be reduced to functions of measurables such as effective population size, chromosome numbers and map length. Some of the assumptions simply reduce the range of validity of the answers, for example the number of generations (t) expressed relative to the population size as t/N should not exceed about unity and for a better fit, one-half. Other assumptions are more fundamental. There are assumed to be many loci affecting the quantitative trait, formally that there are an infinite number of loci such that variance of frequency at single loci and departures from Hardy–Weinberg proportions are irrelevant. There is assumed to be no association between map distance and effects of genes on a trait and in evaluating $\overline{r_i^2}$ that genes are uniformly distributed along the chromosome. In *Drosophila* it is widely accepted that this assumption of independence does not hold well, for blocks of genes have been identified affecting bristle number, for example, but less convincing information applies for mammals (see Bodmer & Parsons (1962) for general review).

Within our theoretical framework such clumping of effects could be incorporated by assuming that the total relevant map length were smaller than the total length, but the extent of reduction required is uncertain. Our simulation results show that even with only 20 genes of equal effect the formulae work satisfactorily. When the genes have a distribution of effects, rather more loci are required for the approximations to be satisfactory, but even so, they do not seem biologically unreasonable, given the complexity of organisms.

The assumptions of additivity at all loci and no changes from natural or artificial selection are obviously very restrictive. Removal of these assumptions while obtaining answers in terms of variances and covariances among traits which do not require specifying individual gene effects and frequency are bound to be difficult. With unselected additive genes, the variance of the genetic variance involves only fourth moments among gene or gene pair frequencies, whilst with dominance, higher moments are required.

Selection

With selection, terms in one order of moments always involve moments of higher order, so simple recurrence relations cannot be obtained. Bulmer (1976) has, however, considered some selection problems and shown by simulation that results could be predicted at least approximately. Also Hill (1976) has used some of the results given here (notably of $V(V_{\Delta t}^*)$) for neutral loci in discussing variation in

response to selection, but only by assuming that gene effects were small and selection sufficiently weak and short term so that the variance among selected populations was likely to be the same as among unselected populations maintained at the same effective size.

Formulae have been presented for variation in response to selection contributed by variation in variance, and it may be useful to consider the significance of this term relative to the variation in response contributed by binomial sampling of genes, discussed in many previous papers and termed genetic drift. Because selection changes parameters, we assume for illustration the effects of a single generation of selection in lines which may already have diverged by drift from the base population. A further consideration of the problem is given elsewhere (Hill, 1976).

From (25) and (26), the variance in response due to variation in variance is

$$V_v(R_t) \simeq \bar{\imath}^2 h_t^4 V_{Pt} (1 - \tfrac{1}{2} h_t^2)^2 . 2\overline{r_t^2},$$

where V_{Pt} is the phenotypic variance at generation t. The variance in response due to genetic sampling of the mean is, from Hill (1974b),

$$V_m(R_t) \simeq h_t^2 V_{Pt} [1 - (1 - 0 \cdot 2 - p) h_t^2] / N \qquad (37)$$

approximately, where p is the proportion of the population selected, and terms due to error of estimation of the genetic mean from the individuals actually measured is ignored. As an example, let us assume that all loci are unlinked, so after three or so generations we can take $2\overline{r_t^2} = 4/(3N)$; then the ratio of contributions to variance in response to one generation of selection from variance and mean, assuming these are independent, is

$$V_v(R_t)/V_m(R_t) \simeq \frac{4\bar{\imath}^2 h_t^2 (1 - \tfrac{1}{2} h_t^2)^2}{3[1 - (1 - 0 \cdot 2 - p) h_t^2]}. \qquad (38)$$

With, for example, typical values of $p = 0 \cdot 2$, $\bar{\imath} = 1 \cdot 4$ giving $\bar{\imath}^2 = 2$ and taking $h^2 = 0 \cdot 5$ and ignoring subsequent small expected changes in h_t^2 due to inbreeding, (38) gives

$$V_v(R_t)/V_m(R_t) \simeq 1 \cdot 07,$$

i.e. the terms are of approximately equal magnitude. For $h^2 = 0 \cdot 2$, the ratio (38) is a little under one-half. Thus the variation in variance can make a significant contribution to variance in response.

Modifying formulae of Hill (1971) to correspond with (37), from sampling the mean, the variance of correlated response in trait Y is

$$V_m(Q_{Yt}) \simeq h_{PYt}^2 V_{PYt} [1 - (1 - 0 \cdot 2 - p) h_{Xt}^2 \rho_t^2] / N. \qquad (39)$$

Combining (39) with (34) and (36) gives the formula corresponding to (38) for correlated traits,

$$V_v(Q_{Yt})/V_m(Q_{Yt}) \simeq \frac{2\bar{\imath}^2 h_{Xt}^2 [1 + \rho_t^2 (1 - 2h_{Xt}^2 + \tfrac{1}{2} h_{Xt}^i)]}{3[1 - (1 - 0 \cdot 2 - p) h_{Xt}^2 \rho_t^2]}.$$

With $p = 0 \cdot 2$, $\bar{\imath}^2 = 2$ and $h_{Xt}^2 = 0 \cdot 5$ as before, and with $\rho = 0 \cdot 5$, for example, the above ratio equals about 3/4, somewhat less than for the direct response.

Error of estimation

Unfortunately, it seems unlikely that experimental tests with animals can be done with sufficient accuracy to validate the formulae which have been obtained, for it is not possible with limited facilities to estimate genetic variances, heritabilities or corrections within a population with much accuracy. For example, the standard error of the estimate of heritability from a half-sib analysis with s families of size n is

$$SE(\hat{h}^2) = 4(1 - \tfrac{1}{4}h^2)\,[1 + (n-1)\tfrac{1}{4}h^2]\left[\frac{2}{n(n-1)(s-1)}\right]^{\frac{1}{2}} \tag{40}$$

(Falconer, 1960), although rather more efficient alternative methods are available. For a fixed total number, ns, recorded, the standard error is minimized by taking $n = 4/h^2$, so with $h^2 = 0.4$ and $n = 10$, $SE(h^2) = 1.02/\sqrt{s}$, approximately, or about 0.2 with $s = 25$. For comparison, the standard deviation of the heritability between replicate lines maintained with random mating and 25 males and 25 females, giving $N = 50$, is from (24), $SD(h^2) = h^2(1-h^2)\sqrt{(2\bar{r^2})} > 0.039$, the latter value applying if all loci are unlinked. The sampling error of estimation is clearly much larger than the true variance. The nearest analogy between the variance of estimation and variance amongst replicates is obtained by considering all families to be infinitely large. Then (40) gives $SE(h^2) = (1 - \tfrac{1}{4}h^2)\sqrt{(2/s)}$, which with s families is roughly double the true value in the first generation, since s genotypes rather than $2N = 4s$ chromosomes from s pairs of individuals are sampled. Also due to sampling one cannot estimate the environmental variance exactly and thus small further errors are introduced. An equivalent analysis of sampling errors of genetic correlation estimates would give the same conclusion.

Despite the many assumptions and restrictions which have been made, our results have highlighted the large degree of variability of parameters expected among small populations. Without replication, inferences about heritabilities and correlations in a base population cannot be made accurately from derived subpopulations.

REFERENCES

BODMER, W. F. & PARSONS, P. A. (1962). Linkage and Recombination in Evolution. *Advances in Genetics* 11, 1–87.

BULMER, M. G. (1976). The effect of selection on genetic variability: a simulation study. *Genetical Research* 28, 101–118.

CROW, J. F. & KIMURA, M. (1970). *An Introduction to Population Genetics Theory.* New York, Evanston and London: Harper and Row.

FALCONER, D. S. (1960). *Introduction to Quantitative Genetics.* Edinburgh and London: Oliver & Boyd.

GREEN, E. L. and staff of Jackson Laboratory (1966). *Biology of the Laboratory Mouse*, 2nd edition. New York, Toronto, London, Sydney: McGraw-Hill.

HILL, W. G. (1971). Design and efficiency of selection experiments for estimating genetic parameters. *Biometrics* 27, 293–311.

HILL, W. G. (1974*a*). Disequilibrium among several linked neutral genes in finite populations. II. Variances and covariances of disequilibria. *Theoretical Population Biology* **6**, 184–198.

HILL, W. G. (1974*b*). Variability of response to selection in genetic experiments. *Biometrics* **30**, 363–366.

HILL, W. G. (1975). Linkage disequilibrium among multiple neutral alleles produced by mutation in finite populations. *Theoretical Population Biology* **8**, 117–126.

HILL, W. G. (1976). Variation in response to selection. *Proceedings of International Conference on Quantitative Genetics*; Iowa State University Press, Ames (in the Press).

HILL, W. G. & ROBERTSON, A. (1968). Linkage disequilibrium in finite populations. *Theoretical and Applied Genetics* **38**, 226–231.

LINDSLEY, D. L. & GRELL, E. H. (1967). *Genetic Variations of Drosophila melanogaster.* Washington: Carnegie Institution Press.

OHTA, T. & KIMURA, M. (1969). Linkage disequilibrium due to random genetic drift. *Genetical Research* **13**, 47–55.

ROBERTSON, A. (1952). The effect of inbreeding on the variation due to recessive genes. *Genetics*, **37**, 189–207.

ROBERTSON, A. (1976). Effects of linkage and population size on selection response and limits. *Proceedings of International Conference on Quantitative Genetics*; Iowa State University Press, Ames (in the Press).

STRICKBERGER, M. W. (1968). *Genetics.* London: Collier-Macmillan and New York: Macmillan.

SVED, J. A. & FELDMAN, M. W. (1973). Correlation and probability methods for 1 and 2 loci. *Theoretical Population Biology* **4**, 129–132.

TALLIS, G. M. (1959). Sampling errors of genetic correlation coefficients calculated from analyses of variance and covariance. *Australian Journal of Statistics* **1**, 35–43.

WOMACK, J. E. (1976). Linkage map of the laboratory mouse. From 'Mouse Newsletter' July 1976 (No. 55) (privately circulated).