THE EFFECTS OF LINKAGE AND LINKAGE DISEQUILIBRIUM ON THE COVARIANCES OF NONINBRED RELATIVES*

B. S. WEIR, C. CLARK COCKERHAM, and JOHN REYNOLDS

Department of Statistics, North Carolina State University, P.O. Box 5457, Raleigh, North Carolina 27650, U.S.A.

SUMMARY

The effects of linkage and linkage disequilibrium on the genetic variances and covariances of noninbred relatives are formulated for quantitative traits with additive and dominance effects but without epistasis. Assortative mating is excluded. Linkage disequilibrium between two loci introduces a covariance between their additive effects and between their dominance effects. The usual coefficients of additive and dominance variances found by counting paths through common ancestors suffice to express the covariances of relatives, which now include the additive and dominance covariances. The linkage parameter, or recombination fraction, comes into play only when relating the additive or dominance covariances from one generation to another.

1. INTRODUCTION

MANY authors have considered the covariance between the genotypic values of relatives in a variety of circumstances, with the most general treatment being given by Gallais (1974). Gallais has reviewed the relevant literature and allowed for linkage, linkage disequilibrium, epistasis and inbreeding for two loci affecting a trait. We consider here only the effects of linkage and linkage disequilibrium, without epistasis or inbreeding, on the covariances of relatives. These effects are not readily displayed in the complicated formulations of Gallais, which are necessitated by his very general treatment. We also provide some examples.

2. MODEL AND FREQUENCIES

For a trait governed by genes at loci A and B, we express the genotypic value of an individual with genotype A_iB_j/A_kB_l as

$$G_{kl}^{ij} = \mu_0 + a^i + a_k + b^i + b_l + d_k^i + c_l^j$$

where a and d denote additive and dominance effects at the A locus, and b and c denote corresponding effects at the B locus. We let the genotypic frequencies $\mathcal{P}$ of the initial ancestors, as a base of reference, be

$$\mathcal{P}_{kl}^{ij} = \mathcal{P}_{ij}^{kl} = P_{ij}P_{kl} = (p_iq_j + \mathcal{D}_{ij})(p_kq_l + \mathcal{D}_{kl})$$

where the P's are gametic frequencies, p's and q's are gene frequencies at loci A and B, respectively, and $\mathcal{D}$'s are the usual measures of linkage

disequilibrium. Thus, the initial ancestors are in Hardy–Weinberg equi-
librium and exhibit linkage disequilibrium.

Gene frequencies are considered to remain constant through matings,
but subsequent measures of linkage disequilibrium, denoted by D's, will
change as

$$D' = \frac{1+\lambda}{4} (D_m + D_f)$$

for each meiosis, where D' is for gametes produced by an individual formed
from (male and female) gametes having disequilibria D_m and D_f. The
linkage parameter λ is one minus twice the recombination fraction. For
convenience we have dropped the allelic subscripts i, j on the D's. We keep
D_m and D_f separate to allow for there being different numbers of individuals
in the paths from paternal and maternal parents back to the initial ancestors.

The disequilibrium for subsequent gametes requires this kind of expan-
sion through each individual back to the gametes in the initial ancestors. If
there are n_r individuals in the path from a present gamete to the two gametes
forming the rth initial ancestor, the probability that the gamete has descen-
ded through this pathway without recombination is $2[(1+\lambda)/4]^{n_r}$, where the
factor 2 accounts for the two gametes in the initial ancestor. All initial
gametes are assumed to have the same disequilibrium. Without inbreeding,
all initial ancestors for any gamete are distinct so that the total probability, α,
of gametic integrity for a present gamete follows from summing over paths

$$\alpha = 2 \sum_r \left(\frac{1+\lambda}{4}\right)^{n_r}.$$

Linkage disequilibrium for our subject gamete depends on initial dis-
equilibrium and requires the gamete to have descended without recom-
bination from an initial gamete, so that

$$D = \alpha \mathcal{D}.$$

For equal values of n_r, $n_r = n$, as would be the case for distinct genera-
tions, summation for α is over 2^{n-1} distinct paths to 2^{n-1} ancestors, and then
we recover the familiar result for gametes uniting to form the nth genera-
tion, $\alpha = [(1+\lambda)/2]^n$, and

$$D = \left(\frac{1+\lambda}{2}\right)^n \mathcal{D}.$$

3. VARIANCES

Since additive or dominance gene effects weighted by frequencies all sum
to zero over any index, and since the D's sum to zero over any index, the
genotypic mean values do not change for any member of any pedigree. The
variance changes however, and for the initial population it is

$$\sigma_G^2 = \sum_i \sum_j \sum_k \sum_l P_{ij} P_{kl} (a^i + a_k + b^j + b_l + d_k^i + c_l^j)^2$$

which decomposes into

additive variance
$$\sigma_A^2 = 2\sigma_a^2 + 2\sigma_b^2 = \sum_i p_i(a^i)^2 + \sum_k p_k(a_k)^2$$
$$+ \sum_j q_j(b^j)^2 + \sum_l q_l(b_l)^2$$

dominance variance
$$\sigma_D^2 = \sigma_d^2 + \sigma_c^2 = \sum_i \sum_k p_i p_k (d_k^i)^2 + \sum_j \sum_l q_j q_l (c_l^j)^2$$

additive covariance
$$4\mathscr{C}_{ab} = 2\sum_i \sum_j \mathscr{D}_{ij}(a^i)(b^j) + 2\sum_k \sum_l \mathscr{D}_{kl}(a_k)(b_l)$$

dominance covariance
$$2\mathscr{C}_{cd} = 2\sum_i \sum_j \sum_k \sum_l \mathscr{D}_{ij}\mathscr{D}_{kl}(d_k^i)(c_l^j).$$

The two covariances of additive effects, one for each uniting gamete, are considered to be the same initially. Linkage disequilibrium is also seen to cause a covariance between dominance effects, but there is no covariance between additive and dominance effects.

For an individual in any pedigree without inbreeding formed by the union of male and female gametes whose probabilities of integrity since the initial gametes are α_m and α_f, the variance is

$$\sigma_G^2 = \sigma_A^2 + \sigma_D^2 + 2(\alpha_m + \alpha_f)\mathscr{C}_{ab} + 2(\alpha_m \alpha_f)\mathscr{C}_{cd}.$$

4. COVARIANCE OF RELATIVES

The same kind of pedigree path finding arguments are used to express the covariance of the genotypic values of noninbred relatives. The arguments are based on the probabilities that genes received by relatives X and Y are derived from a single ancestral gamete. The probabilities that the alleles in one gamete, m or f (male or female) in each of X and Y have descended from one gamete in a common ancestor U in the pedigree are written as φ_{mm_U}, φ_{ff_U}, φ_{mf_U} and φ_{fm_U}. These probabilities are a refinement of those given by Malécot (1948) in part and Kempthorne (1957) in general. The total probability that X and Y receive genes in male gametes, for example, from a single ancestral gamete follows from summing over common ancestors

$$\varphi_{mm} = \sum_U \varphi_{mm_U}.$$

When there are n_U individuals in the loop through common ancestor U between relatives X and Y for their male gametes

$$\varphi_{mm_U} = (\tfrac{1}{2})^{n_U}.$$

The covariance between X and Y without linkage disequilibrium is (Kempthorne, 1957)

$$\mathscr{C}_{G_X G_Y} = \tfrac{1}{2}(\varphi_{mm} + \varphi_{ff} + \varphi_{mf} + \varphi_{fm})\sigma_A^2 + (\varphi_{mm}\varphi_{ff} + \varphi_{mf}\varphi_{fm})\sigma_D^2$$

and our purpose is to elaborate the effects of linkage disequilibrium.

For the effects on the additive covariance $\mathscr{C}_{ab}$ we note that the probability of nonallelic male genes A in X and B in Y (or B in X and A in Y) having descended from the same gamete in a particular common ancestor of X and Y is the same as that for allelic genes (e.g., A's in X and Y).

Consequently, for the covariance of X and Y, the contribution to the covariance of additive effects contributed by ancestor U is

$$\varphi_{mm_U}(\alpha_{m_U} + \alpha_{f_U})\mathscr{C}_{ab}$$

where the terms α_{m_U} and α_{f_U} relate the male and female gametes of U back to the initial founders. We let

$$\gamma_U = \alpha_{m_U} + \alpha_{f_U}$$

and the covariance of a and b effects between X and Y, taking account of the male and female gametes received by each, is

$$\sum_U (\varphi_{mm_U} + \varphi_{ff_U} + \varphi_{mf_U} + \varphi_{fm_U})\gamma_U\mathscr{C}_{ab}.$$

When all common ancestors have the same value of γ, the covariance of a and b effects is

$$(\varphi_{mm} + \varphi_{ff} + \varphi_{mf} + \varphi_{fm})\gamma\mathscr{C}_{ab}.$$

If, further, $\alpha_m = \alpha_f = \alpha$, as would be the case for all ancestors in the same distinct generation, $\gamma = 2\alpha$. For initial common ancestors, $\gamma = 2$, and this situation of not considering individuals earlier than the common ancestors of the relatives in question may be the usual one.

For there to be a covariance between dominance effects between X and Y, each of two AB pairs between A genes in X and B genes in Y (or vice versa) must stem from an ancestral gamete. Because of no inbreeding this happens only when they can be traced to two distinct common ancestors. If male genes stem from common ancestor U and female genes from common ancestor V, the contribution to the covariance is

$$2\varphi_{mm_U}\varphi_{ff_V}(\tfrac{1}{4}\gamma_U\gamma_V)\mathscr{C}_{cd}$$

where the factor of 2 is for the A genes in X and the B genes in Y or vice versa, and the factor of $\frac{1}{4}$ is for the averaging of the probabilities over the four pairings of gametes between U and V. Now adding the case of a male and female gene stemming from each of U and V gives the total covariance of c and d effects between X and Y

$$\tfrac{1}{2}\sum_U\sum_V (\varphi_{mm_U}\varphi_{ff_V} + \varphi_{mf_U}\varphi_{fm_V})\gamma_U\gamma_V\mathscr{C}_{cd}.$$

The total covariance formula for X and Y is now

$$\mathscr{C}_{G_XG_Y} = \tfrac{1}{2}\sum_U (\varphi_{mm_U} + \varphi_{ff_U} + \varphi_{mf_U} + \varphi_{fm_U})[\sigma_A^2 + 2\gamma_U\mathscr{C}_{ab}]$$

$$+ \sum_U\sum_V (\varphi_{mm_U}\varphi_{ff_V} + \varphi_{mf_U}\varphi_{fm_V})[\sigma_D^2 + \tfrac{1}{2}\gamma_U\gamma_V\mathscr{C}_{cd}]. \tag{1}$$

When all common ancestors are in the same, nth, discrete generation, $\gamma = 2\alpha$, where

$$\alpha = \alpha_m = \alpha_f = \left(\frac{1+\lambda}{2}\right)^n$$

and

$$\mathscr{C}_{G_XG_Y} = \tfrac{1}{2}(\varphi_{mm} + \varphi_{ff} + \varphi_{mf} + \varphi_{fm})(\sigma_A^2 + 4\alpha\mathscr{C}_{ab})$$

$$+ (\varphi_{mm}\varphi_{ff} + \varphi_{mf}\varphi_{fm})(\sigma_D^2 + 2\alpha^2\mathscr{C}_{cd}).$$

Consequently, for relatives stemming from common ancestors in a single generation of a random mating population in linkage disequilibrium, the coefficients of additive and of dominance quadratic terms are the same as for linkage equilibrium. We just have to broaden the definitions of the additive and dominance terms to include the respective covariances. Ewens (1979) calls these broader terms the "true" additive and dominance variances, and he also gives a numerical demonstration of the effects of linkage disequilibrium on the covariance of relatives.

When Y is a descendant of ancestor X,

$$\mathcal{C}_{G_X G_Y} = \tfrac{1}{2}\varphi[\sigma_A^2 + 2(\alpha_{m_X} + \alpha_{f_X})\mathcal{C}_{ab}]$$

where the single path of n individuals between X and Y provides the term $\varphi = (\tfrac{1}{2})^n$.

5. EXAMPLES

Evaluation of the covariance formulae is straightforward and requires the identification of common ancestors, the counting of individuals in paths to these ancestors and further evaluation from common ancestors to initial ancestors. It is a simple matter to recover covariances for cases considered by Gallais (1974). If X and Y are parent and offspring,

$$\mathcal{C}_{G_X G_Y} = \tfrac{1}{2}(\sigma_A^2 + 4\mathcal{C}_{ab})$$

while if X and Y are half sibs

$$\mathcal{C}_{G_X G_Y} = \tfrac{1}{4}(\sigma_A^2 + 4\mathcal{C}_{ab})$$

or full sibs

$$\mathcal{C}_{G_X G_Y} = \tfrac{1}{2}(\sigma_A^2 + 4\mathcal{C}_{ab}) + \tfrac{1}{4}(\sigma_D^2 + 2\mathcal{C}_{cd}).$$

The two sib cases require the parents of X and Y to both be in the initial population, while the parent in the parent-offspring case is an initial ancestor. The ease of using equation (1) however is better illustrated with the following three more complex pedigrees.

(i) Three-quarter sibs

When X and Y are three-quarter sibs with common parent U and common grandparents V, W (see fig. 1), the terms arising from paths to common ancestors are

$$\varphi_{mm_U} = \tfrac{1}{2}, \qquad \varphi_{ff_V} = \varphi_{ff_W} = \tfrac{1}{8}$$

so that

$$\mathcal{C}_{G_X G_Y} = \tfrac{1}{8}[3\sigma_A^2 + (4\gamma_U + \gamma_V + \gamma_W)\mathcal{C}_{ab}] + \tfrac{1}{8}[\sigma_D^2 + \tfrac{1}{4}\gamma_U(\gamma_V + \gamma_W)\mathcal{C}_{cd}].$$

If V, W are in a discrete generation n, and U in the following generation $n+1$

$$\mathcal{C}_{G_X G_Y} = \frac{1}{8}\left\{3\sigma_A^2 + \left[8\left(\frac{1+\lambda}{2}\right)^{n+1} + 4\left(\frac{1+\lambda}{2}\right)^n\right]\mathcal{C}_{ab}\right\}$$

$$+ \frac{1}{8}\left[\sigma_D^2 + 2\left(\frac{1+\lambda}{2}\right)^{2n+1}\mathcal{C}_{cd}\right].$$
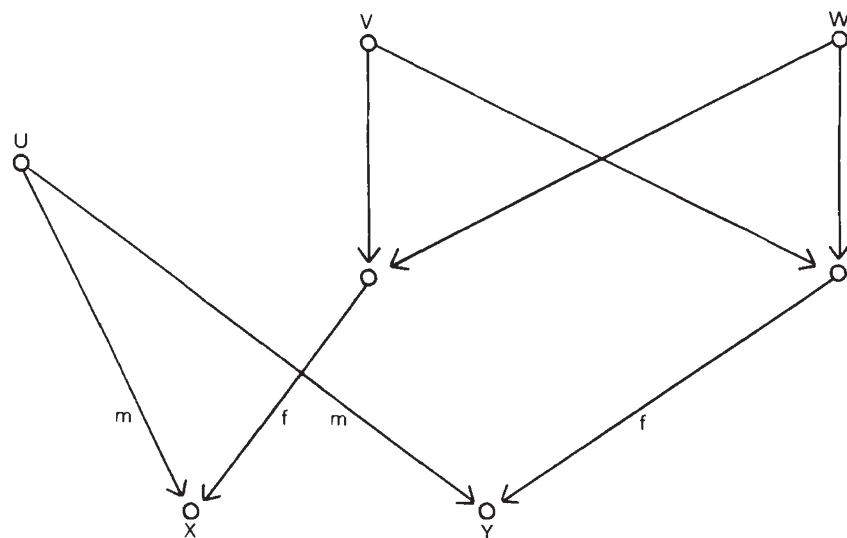
Fig. 1.—Pedigree for three-quarter sibs.

Alternatively, if V, W and U are in the same generation n,

$$\mathscr{C}_{G_X G_Y} = \frac{3}{8}\left[\sigma_A^2 + 4\left(\frac{1+\lambda}{2}\right)^n \mathscr{C}_{ab} \right] + \frac{1}{8}\left[\sigma_D^2 + 2\left(\frac{1+\lambda}{2}\right)^{2n} \mathscr{C}_{cd} \right].$$

(ii) *Three-way cross*

When X and Y have three common grandparents, U, V, W as in fig. 2, the paths to common ancestors give

$$\varphi_{mf_U} = \varphi_{fm_W} = \tfrac{1}{4}, \qquad \varphi_{mm_V} = \tfrac{1}{8}$$

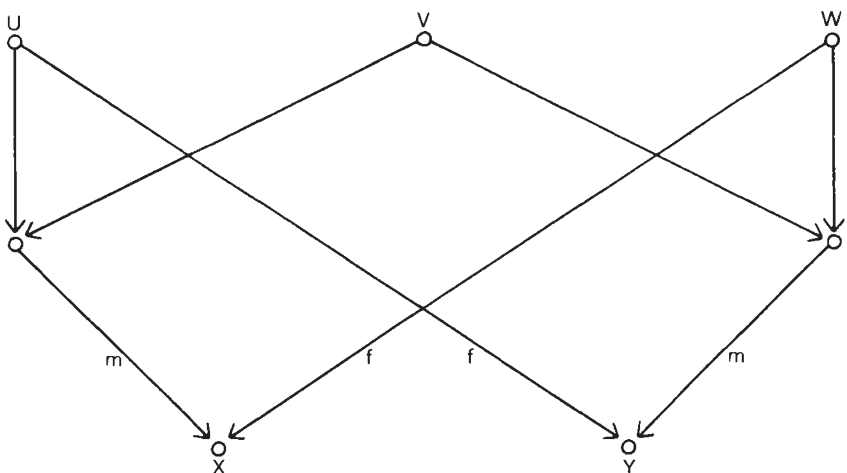
Fig. 2.—Pedigree for three-way cross.

so that

$$\mathscr{C}_{G_X G_Y} = \tfrac{1}{16}[5\sigma_A^2 + (4\gamma_U + 4\gamma_W + 2\gamma_V)\mathscr{C}_{ab}] + \tfrac{1}{16}[\sigma_D^2 + \tfrac{1}{2}\gamma_U\gamma_W\mathscr{C}_{cd}].$$

If U, V, W are all in generation n,

$$\mathscr{C}_{G_X G_Y} = \frac{5}{16}\left[\sigma_A^2 + 4\left(\frac{1+\lambda}{2}\right)^n \mathscr{C}_{ab}\right] + \frac{1}{16}\left[\sigma_D^2 + 2\left(\frac{1+\lambda}{2}\right)^{2n} \mathscr{C}_{cd}\right].$$

(iii) *Four-way cross*

When X and Y have four common grandparents U, V, W, Z as shown in fig. 3,

$$\varphi_{mm_U} = \varphi_{ff_Z} = \varphi_{mf_W} = \varphi_{fm_V} = \tfrac{1}{8}$$
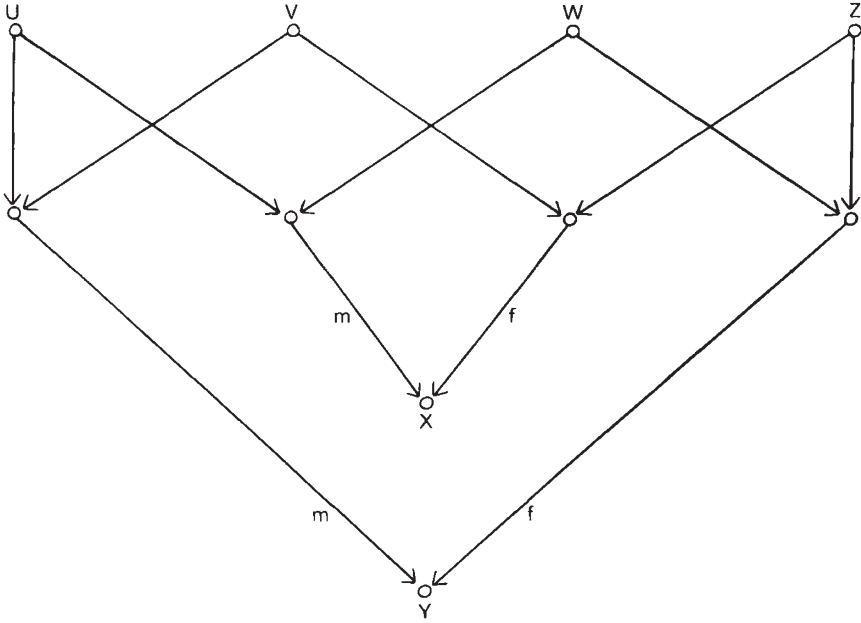

FIG. 3.—Pedigree for four-way cross.

so that

$$\mathscr{C}_{G_X G_Y} = \tfrac{1}{4}[\sigma_A^2 + \tfrac{1}{2}(\gamma_U + \gamma_V + \gamma_W + \gamma_Z)\mathscr{C}_{ab}] + \tfrac{1}{32}[\sigma_D^2 + \tfrac{1}{4}(\gamma_U\gamma_Z + \gamma_V\gamma_W)\mathscr{C}_{cd}]$$

and, when U, V, W, Z are in generation n,

$$\mathscr{C}_{G_X G_Y} = \frac{1}{4}\left[\sigma_A^2 + 4\left(\frac{1+\lambda}{2}\right)^n \mathscr{C}_{ab}\right] + \frac{1}{32}\left[\sigma_D^2 + 2\left(\frac{1+\lambda}{2}\right)^{2n} \mathscr{C}_{cd}\right].$$

6. Discussion

While our formulations have been phrased in terms of two loci, the extension to any number of loci is straightforward. We extend the

definitions of σ_A^2 and σ_D^2 to include all loci, and the covariances, $\mathscr{C}_{ab}$'s and $\mathscr{C}_{cd}$'s, as well as γ's, for each pair of loci, must be summed together. The additive and dominance coefficients in terms of φ's remain the same. There may be occasions, such as in natural populations undergoing a particular selection regimen, where linkage disequilibrium remains constant over time. In this case all γ's $= 2$, and all common ancestors have the same linkage disequilibrium.

The covariance formulae among relatives indicate in a general way the extent to which estimates of additive and dominance variances based on differences of these covariances can be biased by linkage disequilibrium. Unfortunately, the bias, and even its sign, requires an exact knowledge of the genetic effects, gene frequencies, and linkage disequilibria. The possible magnitude of the effects of linkage disequilibrium is very large since there are, for n loci, $n(n-1)/2$ $\mathscr{C}_{ab}$'s and $\mathscr{C}_{cd}$'s as compared to n $2\sigma_a^2$'s and σ_d^2's in the sums. However, there can be considerable cancellation in the sums of $\mathscr{C}_{ab}$'s or $\mathscr{C}_{cd}$'s.

There is a probable basic difference between the $\mathscr{C}_{ab}$'s and $\mathscr{C}_{cd}$'s as to sign, pointed out by Comstock and Robinson (1952). This can be demonstrated for two alleles at each locus. Let the genotypic values be

A_1A_1	A_1A_2	A_2A_2		B_1B_1	B_1B_2	B_2B_2
u	gu	$-u$	,	v	hv	$-v$

where g and h reflect dominance of the favorable genes A_1, B_1. We have to account for only one distinct coefficient of linkage disequilibrium $\mathscr{D}_{A_1B_1} = \mathscr{D}$. The covariances are

$$\mathscr{C}_{ab} = \mathscr{D}[1 - g(p_1 - p_2)][1 - h(q_1 - q_2)]uv$$

$$\mathscr{C}_{cd} = 4\mathscr{D}^2 ghuv.$$

For partial dominance of genes, $-1 < g, h < 1$, the sign of $\mathscr{C}_{ab}$ is determined by the sign of $\mathscr{D}$ and is plus or minus as the A_1B_1 genes are in coupling or repulsion, respectively. On the other hand, the sign of $\mathscr{C}_{cd}$ is determined by the product of the signs of g and h. With dominance of the favorable genes, $g, h > 0$, which appears to be the case in some species, $\mathscr{C}_{cd}$ is positive. Comstock and Robinson (1952) made these discoveries in formulating estimates of dominance for particular mating designs. The conclusions apply to all types of relatives, but are not as clear-cut for multiple alleles.

The inclusion of epistasis in conjunction with correlations due to linkage disequilibrium leads to the usual statistical problems with interactions and correlations in that most effects become correlated, and variances or covariances of relatives become unduly complicated. Even with linkage equilibrium, linkages, $\lambda > 0$, affect the covariances of relatives with epistasis (Cockerham, 1956), since we then require probabilities, analogous to the φ's, for A and B genes received by each relative being jointly descended from a common ancestor. Only when there is complete correlation for a block of genes, i.e., all λ's $= 1$, is there simplification when including epistasis and linkage disequilibrium. In this case there is a single multiple allelic series with allelic frequencies $P_{ijk\ldots}$, and the variance can be operationally broken into additive, $\sigma_{A^*}^2$, and dominance, $\sigma_{D^*}^2$, parts, which in conjunction with the usual additive and dominance coefficients give the correct covariances of relatives. However, we do not specify details of the compositions of

$\sigma^2_{A^*}$ and $\sigma^2_{D^*}$ in terms of the various types of genetic variances and covariances.

In finite populations, it is sometimes plausible to consider linkage disequilibrium to be zero on the average, *i.e.*, $\mathscr{E}D = 0$ where $\mathscr{E}$ denotes expectation, but $\mathscr{E}D^2 \neq 0$ (Hill and Robertson, 1968). This, of necessity, introduces a covariance of dominance effects as studied by Avery and Hill (1978). Finite populations introduce inbreeding and drift which have other effects on the genetic variances and on covariances of relatives, and the general consequences are most easily clarified with our descent measures.

7. REFERENCES

AVERY, P. J., AND HILL, W. G. 1978. The effect of linkage disequilibrium on the genetic variance of a quantitative trait. *Advances in Applied Probability, 10,* 4-6.

COCKERHAM, C. CLARK. 1956. Effect of linkage on the covariances between relatives. *Genetics, 41,* 138-141.

COMSTOCK, R. E., AND ROBINSON, H. F. 1952. Estimation of average dominance of genes, in *Heterosis,* (ed. J. Gowen) pp. 494-516. Iowa State College Press, Ames, Iowa.

EWENS, W. J. 1979. *Mathematical Population Genetics.* Springer-Verlag, Berlin.

GALLAIS, A. 1974. Covariances between arbitrary relatives with linkage and epistasis in the case of linkage disequilibrium. *Biometrics, 30,* 429-446.

HILL, W. G., AND ROBERTSON, A. 1968. Linkage disequilibrium in finite populations. *Theor. Appl. Genet., 38,* 226-231.

KEMPTHORNE, O. 1957. *An Introduction to Genetic Statistics.* Wiley, New York.

MALECOT, G. 1948. *Les Mathematiques de l'Hérédité,* Masson et Cie, Paris.