Variance of Neutral Genetic Variances Within and Between Populations for a Quantitative Character

Zhao-Bang Zeng and C. Clark Cockerham

Department of Statistics, North Carolina State University, Raleigh, North Carolina 27695-8203

Manuscript received February 28, 1991
Accepted for publication June 6, 1991

ABSTRACT

The variances of genetic variances within and between finite populations were systematically studied using a general multiple allele model with mutation in terms of identity by descent measures. We partitioned the genetic variances into components corresponding to genetic variances and covariances within and between loci. We also analyzed the sampling variance. Both transient and equilibrium results were derived exactly and the results can be used in diverse applications. For the genetic variance within populations, $\hat{\sigma}_w^2$, the coefficient of variation can be very well approximated as

$$CV(\hat{\sigma}_w^2)_t \simeq \sqrt{\frac{6}{5m(1-\theta_{1_0})}\left\{\left(1-\frac{1}{2N}\right)^{-t}-1\right\}+\frac{2}{3N}\left(1-\left(\frac{1}{4}\right)^t\right)+\frac{2}{n}}$$

for a normal distribution of allelic effects, ignoring recurrent mutation in the absence of linkage, where m is the number of loci, N is the effective population size, n is the sample size, θ_{1_0} is the initial identity by descent measure of two genes within populations and t is the generation number. The first term is due to genic variance, the second due to linkage disequilibrium, and third due to sampling. In the short term, the variation is predominantly due to linkage disequilibrium and sampling; but in the long term it can be largely due to genic variance. At equilibrium with mutation

$$CV(\hat{\sigma}_w^2)_\infty \simeq \sqrt{\frac{1}{4Num}+\frac{2}{3N}+\frac{2}{n}}$$

where u is the mutation rate. The genetic variance between populations is a parameter. Variance arises only among sample estimates due to finite sampling of populations and individuals. The coefficient of variation for sample genetic variance between populations, $\hat{\sigma}_b^2$, can be generally approximated as

$$CV(\hat{\sigma}_b^2)_t \simeq \sqrt{\frac{2}{S-1}}$$

when the number of loci is large where S is the number of sampling populations.

THERE have been considerable analyses of the genetic variation within and between finite populations. The original work of WRIGHT (1951, 1952) partitioned the total genetic variance of populations $\sigma_T^2 = (1+\theta)\sigma_a^2$ into components of variance within and between populations $\sigma_w^2 = (1-\theta)\sigma_a^2$, $\sigma_b^2 = 2\theta\sigma_a^2$, where θ is the coancestry coefficient between individuals in the same population and is the same as the inbreeding coefficient, and σ_a^2 is the genetic variance in an infinite random mating population. By incorporating mutation into the model COCKERHAM and TACHIDA (1987) formulated the genetic variation within and between populations with mutation in the same framework as WRIGHT's partition and expressed the results in terms of identity by descent measures. In contrast, CHAKRABORTY and NEI (1982) and LYNCH and HILL (1986), working on a different mutation model, reached somewhat different results. For a discussion of the differences of the results between the two models, see COCKERHAM and TACHIDA (1987).

These are, however, analyses of expectations of estimates of genetic variances. Estimates of genetic variances have variation. Variation arises among genetic variances within populations because of drift and mutation ("genetic sampling") and finite sampling of individuals ("statistical sampling"). For the genetic variance between populations, however, variance arises only among sample estimates due to finite sampling of populations and individuals.

Compared with the analysis of genetic variances, analysis of the variance of genetic variances within and between populations is limited. BULMER (1976, 1980) and AVERY and HILL (1977) analyzed the variance of genetic variance within populations without

mutation for two alleles in a locus for a relatively short time. They derived approximate results and observed that if many loci affect the character the variance of genetic variance within populations is contributed to largely by linkage disequilibrium between pairs of loci, and the distribution of individual allelic effects and frequencies may be unimportant in the short term. LYNCH and HILL (1986) analyzed the equilibrium variance of the genetic variance within populations with mutation for the constant variance mutation model. LYNCH (1988b) also approximated the sampling variance of the genetic variance between populations without mutation. Only approximate results were attempted in these analyses however.

In this paper we attempt to derive complete and exact solutions of variance of genetic variances within and between populations for a quantitative character with additive effects of genes undergoing mutation. The analysis is greatly facilitated by utilizing various identity by descent measures. Both transient and equilibrium results are derived, and the results are evaluated for different founder populations. Where possible, we try to extract approximate formulae from the complex solutions. These are then compared with the results of some previous studies and the analysis is presented in the discussion. There are similarities in the approach of using descent measures to the papers of WEIR, AVERY and HILL (1980), WEIR and HILL (1980), COCKERHAM and WEIR (1983) and WEIR, REYNOLDS and DODDS (1990).

VARIANCE OF GENETIC VARIANCE WITHIN POPULATIONS

Definitions and assumptions: Let us consider a set of independent replicate random mating (including selfing) monoecious diploid populations, each consisting of N individuals in each distinct generation, all stemming from the same founder population, which is assumed to be at linkage equilibrium. We consider a quantitative character contributed to by m loci, which are otherwise neutral, undergoing mutation. Mutation is of the form that a random gene in a locus mutates to the ith allele with probability v_i each generation. The total mutation rate for the locus is $u = \sum_i v_i$, $i = 1, 2, \ldots, K$ for K alleles and the equilibrium frequency for the ith allele is $p_{i*} = v_i/u$. We consider only additive effects of genes within and between loci.

Let the genotypic value for a genotype with alleles A_i and A_j at a locus be $G_{ij} = x_i + x_j$, where the x's are considered to be identically independently distributed with some distribution with mean zero, variance σ_x^2, and the fourth central moment μ_4. Genotypes formed by the union of gametes A_iB_k and A_jB_l have frequency $P_{jl}^{ik} = P_{ik}^{jl}$. Sums of genotypic frequencies indicated by dots for the indices summed, provide various marginal totals. Gametic frequencies for A_iB_k, for example, are denoted by

$$P_{\cdot\cdot}^{ik} = \sum_j \sum_l P_{jl}^{ik}.$$

For convenience, allelic frequencies have an alternative notation

$$p_i = P_{\cdot\cdot}^{i\cdot} \text{ for } A_i \text{ and } q_k = P_{\cdot\cdot}^{\cdot k} \text{ for } B_k.$$

All of these frequencies are expected population values so that they are expected values over all possible replicate populations maintained with identical histories.

Let the expected gene and gametic frequencies in a population be distinguished by a degree hat $\mathring{}$ (or other notations indicated below), as in $\mathring{p}_i$. Then, with random union of gametes as in our mating system, the expected genetic variance in a population is defined as

$$\mathring{\sigma}_w^2 = \sum_r \sum_i \sum_j \mathring{p}_i \mathring{p}_j (x_i + x_j)^2 + 2 \sum_{r \neq s} \sum_i \sum_k \sum (\mathring{P}_{\cdot\cdot}^{ik} + \mathring{P}_{\cdot\cdot}^{i\cdot}{}_{\cdot k}) x_i x_k - (2 \sum_r \sum_i \mathring{p}_i x_i)^2 \quad (1)$$

where the summations of r and s are over loci and the summations of i, j, k and l are over alleles with respect to the implied locus. This can be decomposed into components representing gene effects (genic variance) within loci and the effect of linkage disequilibrium which is the covariance between loci

$$\mathring{\sigma}_w^2 = \mathring{\sigma}_{wg}^2 + \mathring{\sigma}_{wL}^2 \quad (2)$$

with

$$\mathring{\sigma}_{wg}^2 = 2 \sum_r \sum_i \mathring{p}_i (1 - \mathring{p}_i) x_i^2 - 2 \sum_r \sum_{i \neq j} \sum \mathring{p}_i \mathring{p}_j x_i x_j \quad (3)$$

$$\mathring{\sigma}_{wL}^2 = 2 \sum_{r \neq s} \sum_i \sum_k \sum (\mathring{P}_{\cdot\cdot}^{ik} + \mathring{P}_{\cdot\cdot}^{i\cdot}{}_{\cdot k} - 2 \mathring{p}_i \mathring{q}_k) x_i x_k. \quad (4)$$

The variance of $\mathring{\sigma}_w^2$ among replicate populations is

$$\mathrm{Var}_b(\mathring{\sigma}_w^2) = \mathscr{E}(\mathring{\sigma}_w^2)^2 - \mathscr{E}(\mathring{\sigma}_w^2 \check{\sigma}_w^2) \quad (5)$$

where $\mathring{\sigma}_w^2$ and $\check{\sigma}_w^2$ are used to denote the variances for two distinct replicate populations, which is the difference between the variance for unrelated populations and the covariance for replicate populations. In analyzing the variance of variance (5), we will dissect it into parts of the variance corresponding to the different genetic variance components (2).

Identity measures: Analysis is greatly facilitated by utilizing identity by descent measures. With mutation they are the probabilities that the genes are identical by descent and none of them has mutated. The required descent measures are defined and listed in Table 1. They involve genes from four populations, both for within and between loci. Explicit solutions of these descent measures for random mating monoecious populations are given in APPENDICES A and B. Many of these descent measures for one and

TABLE 1

Definitions of identity measures

Description	Identity measures	
One locus		
Within population		
Two genes	$\theta_1 = \text{Prob}(a \equiv b)$	
Three genes	$\gamma_1 = \text{Prob}(a \equiv b \equiv c)$	
Four genes	$\delta_1 = \text{Prob}(a \equiv b \equiv c \equiv d)$	$\delta_2 = \text{Prob}(a \equiv b, c \equiv d)$
Between two populations		
Two genes	$\theta_1' = \text{Prob}(a \equiv a')$	
Three genes	$\gamma_1' = \text{Prob}(a \equiv b \equiv a')$	
Four genes	$\delta_1' = \text{Prob}(a \equiv b, a' \equiv b')$	$\delta_2' = \text{Prob}(a \equiv b \equiv a' \equiv b')$
	$\delta_3' = \text{Prob}(a \equiv a', b \equiv b')$	$\delta_4' = \text{Prob}(a \equiv b \equiv c \equiv a')$
	$\delta_5' = \text{Prob}(a \equiv b, c \equiv a')$	
Among three populations		
Three genes	$\gamma_1'' = \text{Prob}(a \equiv a' \equiv a'')$	
Four genes	$\delta_1'' = \text{Prob}(a \equiv b, a' \equiv a'')$	$\delta_2'' = \text{Prob}(a \equiv b \equiv a' \equiv a'')$
	$\delta_3'' = \text{Prob}(a \equiv a', b \equiv a'')$	
Among four populations		
Four genes	$\delta_1''' = \text{Prob}(a \equiv a', a'' \equiv a''')$	$\delta_2''' = \text{Prob}(a \equiv a' \equiv a'' \equiv a''')$
Two loci		
Within population		
On two gametes	$\tilde{\theta}_1 = \text{Prob}(a \equiv b, A \equiv B)$	
On three gametes	$\tilde{\gamma}_1 = \text{Prob}(a \equiv b, A \equiv C)$	
On four gametes	$\tilde{\delta}_1 = \text{Prob}(a \equiv b, C \equiv D)$	
Between two populations		
On two gametes	$\tilde{\theta}_1' = \text{Prob}(a \equiv a', A \equiv A')$	
On three gametes	$\tilde{\gamma}_1' = \text{Prob}(a \equiv a', A \equiv B')$	$\tilde{\gamma}_2' = \text{Prob}(a \equiv b, A \equiv A')$
On four gametes	$\tilde{\delta}_1' = \text{Prob}(a \equiv b, A' \equiv B')$	$\tilde{\delta}_2' = \text{Prob}(a \equiv a', B \equiv B')$
	$\tilde{\delta}_3' = \text{Prob}(a \equiv b, C \equiv A')$	
Among three populations		
On three gametes	$\tilde{\gamma}_1'' = \text{Prob}(a \equiv a', A \equiv A'')$	
On four gametes	$\tilde{\delta}_1'' = \text{Prob}(a \equiv a', B \equiv A'')$	$\tilde{\delta}_2'' = \text{Prob}(a \equiv b, A' \equiv A'')$
Among four populations		
On four gametes	$\tilde{\delta}_1''' = \text{Prob}(a \equiv a', A'' \equiv A''')$	

Identity by descent measures are defined as the probability of genes being identical by descent without mutations (denoted by ≡). The representation of random distinct genes from populations is

	Population 1	Population 2	Population 3	Population 4
Locus 1	a b c d	a′ b′	a″	a‴
	| | | |	| |	|	|
Locus 2	A B	A′ B′	A″	A‴

Genes connected by a line are on the same gamete. Only a representative gene arrangement is presented above.

two populations have been given by H. TACHIDA and C. C. COCKERHAM (unpublished results) and are reproduced here. We use somewhat different notations here due to the complexity of the problem considered. As before, θ is used for two genes, γ for three genes, and δ for four genes. For genes involving two populations a prime is used, two primes for three populations, and three primes for four populations. Tilde (~) is used for descent measures involving two loci.

These descent measures are functions of effective population size N, mutation rate u, recombination rate r for two loci, generation time t and also initial descent measures which depend on the founder population. The initial values of primed descent measures are equal to the initial values of their unprimed counterparts within populations. At equilibrium with mutation, except that $\delta_{1_\infty}' = \theta_{1_\infty}^2$ and $\tilde{\delta}_{1_\infty}' = \theta_{1_\infty}\theta_{1_\infty}$, all other primed descent measures are zero.

One-locus analysis: We first analyze the variance of genetic variance within populations for a locus by assuming that the genetic variance is averaged over many samples from the same population. In so doing we analyze only the *population component of variance*, $\text{Var}_b(\hat{\sigma}_w^2)$, *of the genetic variance* which takes into account genetic drift and mutation that occurred during the period of separation of replicate populations, ignoring the sampling variance of the genetic variance due to finite sampling of individuals within replicate populations which we will discuss later.

From Equation 3 the genetic variance among indi-

TABLE 2

Gene frequency functions associated with $\mathscr{E}_x\mathring{Z}\check{Z}$

	μ_4	σ_x^4
$\mathscr{E}_x\mathring{Z}_1\check{Z}_1$	$\sum_i \mathring{p}_i\check{p}_i$	$1 - \sum_i \mathring{p}_i\check{p}_i$
$\mathscr{E}_x\mathring{Z}_2\check{Z}_2$	$\sum_i \mathring{p}_i^2\check{p}_i^2$	$\sum_{i\neq j} \mathring{p}_i^2\check{p}_j^2$
$\mathscr{E}_x\mathring{Z}_3\check{Z}_3$		$2\sum_{i\neq j} \mathring{p}_i\check{p}_i\mathring{p}_j\check{p}_j$
$\mathscr{E}_x\mathring{Z}_2\check{Z}_1$	$\sum_i \mathring{p}_i^2\check{p}_i$	$\sum_i \mathring{p}_i^2 - \sum_i \mathring{p}_i^2\check{p}_i$

The components of expectations with $\mathscr{E}_x\mathring{Z}^2$ are given by replacing $\check{}$ by $\mathring{}$ in all of the above expectations.

viduals for a locus r in a replicate population is defined as

$$\mathring{\sigma}^2_{wg_r} = 2\left[\sum_i \hat{p}_i x_i^2 - \sum_i \hat{p}_i^2 x_i^2 - \sum_{i\neq j} \hat{p}_i\hat{p}_j x_i x_j\right]$$
$$= 2[\mathring{Z}_1 - \mathring{Z}_2 - \mathring{Z}_3] = 2\mathring{Z}$$

where $\hat{p}_i$ is the expected frequency of allele A_i in gametes at reproduction in a replicate population. To analyze the variance of this variance, we let the genetic variance of another distinct replicate population be defined as

$$\check{\sigma}^2_{wg_r} = 2\left[\sum_i \check{p}_i x_i^2 - \sum_i \check{p}_i^2 x_i^2 - \sum_{i\neq j} \check{p}_i\check{p}_j x_i x_j\right]$$
$$= 2[\check{Z}_1 - \check{Z}_2 - \check{Z}_3] = 2\check{Z}.$$

Then the population component of variance of $\mathring{\sigma}^2_w$ is

$$\mathrm{Var}_b(\mathring{\sigma}^2_{wg_r}) = \mathscr{E}(\mathring{\sigma}^2_{wg_r})^2 - \mathscr{E}(\mathring{\sigma}^2_{wg_r}\check{\sigma}^2_{wg_r}) = 4[\mathscr{E}\mathring{Z}^2 - \mathscr{E}\mathring{Z}\check{Z}].$$

Next we expand the expectation $\mathscr{E} = \mathscr{E}_b\mathscr{E}_x$ to include the expectation $\mathscr{E}_b$ over replicate populations, and $\mathscr{E}_x$ with respect to the x's, assuming that allelic effects and frequencies are independent. First, in taking expectation with respect to the x's, we note that $\mathscr{E}_x x_i x_j x_k x_l = \mathscr{E}_x x_i^2 x_j x_k = \mathscr{E}_x x_i^3 x_j = 0$ and $\mathscr{E}_x x_i^4 = \mu_4$, $\mathscr{E}_x x_i^2 x_j^2 = \sigma_x^4$ for $i \neq j \neq k \neq l$. Thus, the nonzero expectations with respect to the x's in $\mathscr{E}_x\mathring{Z}\check{Z}$ are

$$\mathscr{E}_x\mathring{Z}\check{Z} = \mathscr{E}_x\mathring{Z}_1\check{Z}_1 + \mathscr{E}_x\mathring{Z}_2\check{Z}_2 + \mathscr{E}_x\mathring{Z}_3\check{Z}_3 - 2\mathscr{E}_x\mathring{Z}_2\check{Z}_1$$

and the associated gene frequency functions are listed in Table 2.

Now we take the expectation, $\mathscr{E}_b$, over replicate populations using identity by descent measures. In deriving $\mathscr{E}_b\sum_i \hat{p}_i^2$ we note that $\mathscr{E}_b\sum_i \hat{p}_i^2$ is the expected probability that two randomly chosen genes in a population are the same gene. They could be the same either because they are identical by descent (with probability θ_1) or they are not identical by descent (with probability $1 - \theta_1$) but are identical in state (with probability $\sum_i p_i^2$ which is $q_2 = \sum_i v_i^2/u^2$ if the initial population is at equilibrium). This gives

$$\mathscr{E}_b \sum_i \hat{p}_i^2 = \theta_1 + (1 - \theta_1)q_2.$$

By similar arguments, we can show that

$$\mathscr{E}_b\sum_i \hat{p}_i\check{p}_i = \theta_1' + (1-\theta_1')q_2$$

$$\mathscr{E}_b\sum_i \hat{p}_i^2\check{p}_i = \gamma_1' + 2(\theta_1' - \gamma_1')q_2 + (\theta_1 - \gamma_1')q_2 + (1 - \theta_1 - 2\theta_1' + 2\gamma_1')q_3$$

$$\mathscr{E}_b\sum_i \hat{p}_i^2\check{p}_i^2 = \delta_2' + 4(\gamma_1' - \delta_2')q_2 + 2(\delta_3' - \delta_2')q_2 + (\delta_1' - \delta_2')q_2 + 2(\theta_1 - 2\gamma_1' - \delta_1' + 2\delta_2')q_3 + 4(\theta_1' - 2\gamma_1' - \delta_3' + 2\delta_2')q_3 + (1 - 2\theta_1 - 4\theta_1' + 8\gamma_1' + \delta_1' + 2\delta_3' - 6\delta_2')q_4$$

$$\mathscr{E}_b\sum_{i\neq j} \hat{p}_i^2\check{p}_j^2 = (\delta_1' - \delta_2')(1 - q_2) + 2(\theta_1 - 2\gamma_1' - \delta_1' + 2\delta_2') \cdot (q_2 - q_3) + (1 - 2\theta_1 - 4\theta_1' + 8\gamma_1' + \delta_1' + 2\delta_3' - 6\delta_2')(q_2^2 - q_4)$$

$$\mathscr{E}_b\sum_{i\neq j} \hat{p}_i\check{p}_i\hat{p}_j\check{p}_j = (\delta_3' - \delta_2')(1 - q_2) + 2(\theta_1' - 2\gamma_1' - \delta_3' + 2\delta_2')(q_2 - q_3) + (1 - 2\theta_1 - 4\theta_1' + 8\gamma_1' + \delta_1' + 2\delta_3' - 6\delta_2') \cdot (q_2^2 - q_4)$$

where $q_r = \sum_i (v_i/u)^r = \sum_i p_{i*}^r$. It is noted that these results depend on the initial gene frequencies p_{i0}'s satisfying the condition $\mathscr{E}_b \sum_i p_{i*}^{r-1}(p_{i0} - p_{i*}) = 0$ or $r = 1, 2, 3$ and 4. This condition is satisfied when (i) the initial population is at equilibrium, or is fixed for allele A_i with probability p_{i*} ($\mathscr{E}_b p_{i0} = p_{i*}$); (ii) the mutation rate is equal among alleles ($p_{i*} = 1/K$); or (iii) the number of alleles is infinite ($K \to \infty$), in which case $q_2 = q_3 = q_4 = 0$ also. When the condition is not satisfied, the gene frequency functions become more complex. We avoid discussing these situations here.

The nonzero expectations in $\mathscr{E}_x\mathring{Z}^2$ are

$$\mathscr{E}_x\mathring{Z}^2 = \mathscr{E}_x\mathring{Z}_1^2 + \mathscr{E}_x\mathring{Z}_2^2 + \mathscr{E}_x\mathring{Z}_3^2 - 2\mathscr{E}_x\mathring{Z}_1\mathring{Z}_2.$$

Note that this amounts to replacing $\check{}$ by $\mathring{}$ in all of the expectations involving $\mathring{Z}\check{Z}$ including the $\check{p}$'s, e.g. $\mathscr{E}_b\sum_{i\neq j} \hat{p}_i\check{p}_i\hat{p}_j\check{p}_j$ becomes $\mathscr{E}_b\sum_{i\neq j} \hat{p}_i^2\hat{p}_j^2$, and replacing the between population descent measures δ_1' and δ_3' by the within population descent measure δ_2, and also θ_1', γ_1', δ_2' by θ_1, γ_1, δ_1 respectively.

Putting the analysis together,

$$\begin{aligned}\mathrm{Var}_b(\mathring{\sigma}^2_{wg_r}) = 4\{&(\theta_1 - 2\gamma_1 + \delta_1 - \theta'_1 + 2\gamma'_1 - \delta'_2)\\ &\cdot [(1 - 5q_2 + 8q_3 - 4q_4)\mu_4\\ &- (1 - 9q_2 + 8q_3 - 12q_2^2 + 12q_4)\sigma_x^4]\\ &- (3\delta_2 - 2\delta_1 - \delta'_1 + 2\delta'_2 - 2\delta'_3)\\ &\cdot [(q_2 - 2q_3 + q_4)\mu_4\\ &- (1 - 9q_2 + 8q_3 - 12q_2^2 + 12q_4)\sigma_x^4]\}.\end{aligned}$$

It is initially zero. At equilibrium ($\delta'_{1_\infty} = \theta^2_{1_\infty}$ and all other primed descent measures are zero)

$$\begin{aligned}\mathrm{Var}_b(\mathring{\sigma}^2_{wg_r}) = 4\{&(\theta_{1_\infty} - 2\gamma_{1_\infty} + \delta_{1_\infty})\\ &\cdot [(1 - 5q_2 + 8q_3 - 4q_4)\mu_4\\ &- (1 - 9q_2 + 8q_3 - 12q_2^2 + 12q_4)\sigma_x^4]\\ &- (3\delta_{2_\infty} - 2\delta_{1_\infty} - \theta^2_{1_\infty})[(q_2 - 2q_3 + q_4)\mu_4\\ &- (1 - 9q_2 + 8q_3 - 12q_2^2 + 12q_4)\sigma_x^4]\}.\end{aligned}$$

For the infinite allele mutation model

$$\begin{aligned}&\mathrm{Var}_b(\mathring{\sigma}^2_{wg_r})\\ &\quad = 4\{(\theta_1 - 2\gamma_1 + \delta_1 - \theta'_1 + 2\gamma'_1 - \delta'_2)(\mu_4 - 3\sigma_x^4)\\ &\qquad + (2\theta_1 - 4\gamma_1 + 3\delta_2 - 2\theta'_1 + 4\gamma'_1 - \delta'_1 - 2\delta'_3)\sigma_x^4\}\end{aligned} \tag{6}$$

and at equilibrium

$$\begin{aligned}\mathrm{Var}_b(\mathring{\sigma}^2_{wg_r}) = 4\{&(\theta_{1_\infty} - 2\gamma_{1_\infty} + \delta_{1_\infty})(\mu_4 - 3\sigma_x^4)\\ &+ (2\theta_{1_\infty} - 4\gamma_{1_\infty} + 3\delta_{2_\infty} - \theta^2_{1_\infty})\sigma_x^4\}.\end{aligned} \tag{7}$$

Identity disequilibrium: When we consider several loci, we sum up the above variance over loci. In addition, the variance component also contains joint product terms involving two loci, aside from the variance of linkage disequilibrium discussed below. This part of the variation as we will show is due to identity disequilibrium.

For a pair of loci r and s the non-zero expectations with respect to the x's in $\mathrm{Var}_b(\mathring{\sigma}^2_{wg})$ contain

$$\begin{aligned}&4\sum_i\sum_k \mathring{p}_i(1 - \mathring{p}_i)\mathring{q}_k(1 - \mathring{q}_k)\sigma^2_{x_r}\sigma^2_{x_s}\\ &\qquad - 4\sum_i\sum_k \mathring{p}_i(1 - \mathring{p}_i)\check{q}_k(1 - \check{q}_k)\sigma^2_{x_r}\sigma^2_{x_s}\end{aligned} \tag{8}$$

where p and q denote allelic frequencies for loci r and s respectively. By taking the expectation $\mathscr{E}_b$ we note that

$$\begin{aligned}\mathscr{E}_b\sum_i\sum_k \mathring{p}_i^2\mathring{q}_k^2 &= \tilde{\delta}_1 + (\theta_{1_r} - \tilde{\delta}_1)q_{2_s} + (\theta_{1_s} - \tilde{\delta}_1)q_{2_r}\\ &\quad + (1 - \theta_{1_r} - \theta_{1_s} + \tilde{\delta}_1)q_{2_r}q_{2_s}\end{aligned}$$

$$\begin{aligned}\mathscr{E}_b\sum_i\sum_k \mathring{p}_i^2\check{q}_k^2 &= \tilde{\delta}'_1 + (\theta_{1_r} - \tilde{\delta}'_1)q_{2_s} + (\theta_{1_s} - \tilde{\delta}'_1)q_{2_r}\\ &\quad + (1 - \theta_{1_r} - \theta_{1_s} + \tilde{\delta}'_1)q_{2_r}q_{2_s}.\end{aligned}$$

This gives (8) the value

$$4(\tilde{\delta}_1 - \tilde{\delta}'_1)(1 - q_{2_r})(1 - q_{2_s})\sigma^2_{x_r}\sigma^2_{x_s}$$

since $\mathscr{E}_b\sum_i\sum_k \mathring{p}_i\mathring{q}_k = \mathscr{E}_b\sum_i\sum_k \mathring{p}_i\check{q}_k = 1$ and $\mathscr{E}_b\sum_i\sum_k \mathring{p}_i^2\mathring{q}_k = \mathscr{E}_b\sum_i\sum_k \mathring{p}_i^2\check{q}_k = \theta_{1_r} + (1 - \theta_{1_r})q_{2_r}$. For a founder population at identity equilibrium, $\tilde{\delta}'_1 = \theta_{1_r}\theta_{1_s}$, however.

$\eta_b = \tilde{\delta}_1 - \theta_{1_r}\theta_{1_s}$ is one of the components (COCKERHAM 1984) of the total identity disequilibrium $\eta = \tilde{\theta}_1 - \theta_{1_r}\theta_{1_s}$ (WEIR and COCKERHAM 1969) which is the difference between double identity and the product of the probabilities for single identities. Other components are $\eta_a = 2(\tilde{\gamma}_1 - \tilde{\delta}_1)$ and $\eta_d = \tilde{\theta}_1 - 2\tilde{\gamma}_1 + \tilde{\delta}_1$ with $\eta_b + \eta_a + \eta_d = \eta$. Consequently, when summing over all pairs of loci (indexed by r and s) for an infinite allele mutation model, we have

$$\begin{aligned}\mathrm{Var}_b(\mathring{\sigma}^2_{wg}) = 4\sum_r\{&(\theta_1 - 2\gamma_1 + \delta_1 - \theta'_1 + 2\gamma'_1 - \delta'_2)\\ &\cdot (\mu_4 - \sigma_x^4) + (3\delta_2 - 2\delta_1 - \delta'_1\\ &+ 2\delta'_2 - 2\delta'_3)\sigma_x^4\} + 4\sum_{r\neq s}\sum \eta_b\sigma^2_{x_r}\sigma^2_{x_s}.\end{aligned} \tag{9}$$

Linkage disequilibrium: Because our populations are finite with effective size N, the genetic variance within populations also contains covariances between loci due to linkage disequilibrium. For a pair of loci r and s in a population this covariance is, from (4)

$$\mathring{\sigma}^2_{wL_{rs}} = 4\sum_i\sum_k (\mathring{P}^{ik}_{\cdot\cdot} + \mathring{P}^{i\cdot}_{\cdot k} - 2\mathring{p}_i\mathring{q}_k)x_ix_k.$$

This contains two types of linkage disequilibrium, distinguished according to whether the two genes, within an individual, are gametic or not

$$\mathring{D}^{ik}_{\cdot\cdot} = \mathring{P}^{ik}_{\cdot\cdot} - \mathring{p}_i\mathring{q}_k \text{ for gametic genes}$$

$$\mathring{D}^{i\cdot}_{\cdot k} = \mathring{P}^{i\cdot}_{\cdot k} - \mathring{p}_i\mathring{q}_k \text{ for nongametic genes.}$$

These are covariances representing differences between the joint frequencies and the products of single frequencies. For monoecious mating populations, $\mathring{D}^{i\cdot}_{\cdot k} = 0$. By expectation $\mathscr{E}(\mathring{\sigma}^2_{wL_{rs}}) = 0$ and the variance is

$$\begin{aligned}\mathrm{Var}_b(\mathring{\sigma}^2_{wL_{rs}}) &= \mathscr{E}(\mathring{\sigma}^2_{wL_{rs}})^2 - \mathscr{E}(\mathring{\sigma}^2_{wL_{rs}}\check{\sigma}^2_{wL_{rs}})\\ &= 16\,\mathscr{E}\left(\sum_i\sum_k((\mathring{D}^{ik}_{\cdot\cdot})^2 - \mathring{D}^{ik}_{\cdot\cdot}\check{D}^{ik}_{\cdot\cdot})x_i^2x_k^2\right).\end{aligned}$$

Expressed in terms of identity by descent measures and summed over all pairs of loci

$$\begin{aligned}\mathrm{Var}_b(\mathring{\sigma}^2_{wL}) = 8\sum_{r\neq s}\sum&(\tilde{\theta}_1 - 2\tilde{\gamma}_1 + \tilde{\delta}_1 - \tilde{\theta}'_1 + 2\tilde{\gamma}'_1 - \tilde{\delta}'_2)\\ &\cdot (1 - q_{2_r})(1 - q_{2_s})\sigma^2_{x_r}\sigma^2_{x_s}.\end{aligned}$$

which can also be expressed as

$$\mathrm{Var}_b(\mathring{\sigma}^2_{wL}) = 8\sum_{r\neq s}\sum(\eta_d - \eta'_d)(1 - q_{2_r})(1 - q_{2_s})\sigma^2_{x_r}\sigma^2_{x_s} \tag{10}$$

if we let $\eta_d' = \tilde{\theta}_1' - 2\tilde{\gamma}_1' + \tilde{\delta}_2'$. As pointed out by COCKERHAM (1984), the variance of linkage disequilibrium is actually a component of the identity disequilibrium.

The total population component of variance for the genetic variance within populations is then

$$\mathrm{Var}_b(\mathring{\sigma}_w^2) = \mathrm{Var}_b(\mathring{\sigma}_{wg}^2 + \mathring{\sigma}_{wL}^2) = \mathrm{Var}_b(\mathring{\sigma}_{wg}^2) + \mathrm{Var}_b(\mathring{\sigma}_{wL}^2).$$

Sampling variance: We also consider the effect of sampling on the variance of the genetic variance within populations. For samples from different populations, the variance of genetic variances among samples contains an additional source of variation due to sampling. The genetic variance among individuals in a sample from a replicate population with size n is defined as

$$\hat{\sigma}_w^2 = \frac{n}{n-1}\Big\{4\sum_r\sum_i \hat{P}_{i\cdot}^{i\cdot}x_i^2 + \sum_r\sum_{i<j}\hat{P}_{j\cdot}^{i\cdot}(x_i+x_j)^2 + 2\sum_{r\neq s}\sum_i\sum_k(\hat{P}_{\cdot\cdot}^{ik} + \hat{P}_{\cdot k}^{i\cdot})x_ix_k - (2\sum_r\sum_i \hat{p}_ix_i)^2\Big\} \tag{11}$$

where $n/(n-1)$ is the usual correction factor for bias, and $\hat{P}$'s (and $\hat{p}$'s) represent actual frequencies. This can be decomposed into components representing the sampling of gene frequencies and the effects of sample departures from Hardy-Weinberg and linkage equilibrium

$$\hat{\sigma}_w^2 = \hat{\sigma}_{wg}^2 + \sigma_{wHW}^2 + \hat{\sigma}_{wL}^2$$

with

$$\hat{\sigma}_{wg}^2 = \frac{n}{n-1}2\sum_r\Big\{\sum_i \hat{p}_ix_i^2 - \sum_i \hat{p}_i^2x_i^2 - \sum_{i\neq j}\sum \hat{p}_i\hat{p}_jx_ix_j\Big\}$$

$$= \frac{n}{n-1}2\sum_r\{\hat{Z}_{r1} - \hat{Z}_{r2} - \hat{Z}_{r3}\} \tag{12}$$

$$\hat{\sigma}_{wHW}^2 = \frac{n}{n-1}2\sum_r\Big\{\sum_i \hat{P}_{i\cdot}^{i\cdot}x_i^2 - \sum_i \hat{p}_i^2x_i^2 + \sum_{i<j}\sum \hat{P}_{j\cdot}^{i\cdot}x_ix_j - \sum_{i\neq j}\sum \hat{p}_i\hat{p}_jx_ix_j\Big\}$$

$$= \frac{n}{n-1}2\sum_r\{\hat{Z}_{r4} - \hat{Z}_{r2} + \hat{Z}_{r5} - \hat{Z}_{r3}\} \tag{13}$$

$$\sigma_{wL}^2 = \frac{n}{n-1}2\sum_{r\neq s}\sum\Big\{\sum_i\sum_k(\hat{P}_{\cdot\cdot}^{ik} + \hat{P}_{\cdot k}^{i\cdot})x_ix_k - 2\sum_i\sum_k \hat{p}_i\hat{q}_kx_ix_k\Big\} \tag{14}$$

$$= \frac{n}{n-1}2\sum_{r\neq s}\sum\{\hat{Z}_{rs6} - 2\hat{Z}_{rs7}\}.$$

In this setting the sample heterozygotes are represented only once in the summation and the gene frequencies are calculated as

$$\hat{p}_i = \hat{P}_{i\cdot}^{i\cdot} + \tfrac{1}{2}\sum_{i<j}\hat{P}_{j\cdot}^{i\cdot} \tag{15}$$

Throughout the rest of the paper, a sum such as that in this equation is meant to include every possible heterozygote for allele A_i, but to include it only once. The coefficients

$$\hat{D}_{i\cdot}^{i\cdot} = \hat{P}_{i\cdot}^{i\cdot} - \hat{p}_i^2 \text{ and } \hat{D}_{j\cdot}^{i\cdot} = \hat{P}_{j\cdot}^{i\cdot} - 2\hat{p}_i\hat{p}_j$$

are the Hardy-Weinberg disequilibrium coefficients which measure sample departures from Hardy-Weinberg genotypic frequencies. Since separate digenic frequencies $\hat{P}_{\cdot\cdot}^{ik}$ and $\hat{P}_{\cdot k}^{i\cdot}$ cannot be observed, it is always convenient to define the sum of sampling gametic and nongametic linkage disequilibria as a composite measure

$$\hat{\Delta}_{ik} = \hat{D}_{\cdot\cdot}^{ik} + \hat{D}_{\cdot k}^{i\cdot} = \hat{P}_{\cdot\cdot}^{ik} + \hat{P}_{\cdot k}^{i\cdot} - 2\hat{p}_i\hat{q}_k$$

with

$$\hat{P}_{\cdot\cdot}^{ik} + \hat{P}_{\cdot k}^{i\cdot} = 2\hat{P}_{ik}^{ik} + \sum_{i<j}\hat{P}_{jk}^{ik} + \sum_{k<l}\hat{P}_{il}^{ik} + \tfrac{1}{2}\sum_{i<j}\sum_{k<l}(\hat{P}_{jl}^{ik} + \hat{P}_{jk}^{il}). \tag{16}$$

Note that the expected value of $\hat{\sigma}_{wL}^2$ is zero, but that of $\hat{\sigma}_{wHW}^2$ is not. With sampling the expectation $\mathscr{E} = \mathscr{E}_b\mathscr{E}_w\mathscr{E}_x$ contains three level expectations: $\mathscr{E}_x$ with respect to the x's, $\mathscr{E}_w$ with respect to the sampling within populations and $\mathscr{E}_b$ over all replicate populations. If we utilize $\mathscr{E}_w\hat{p} = \hat{p}_i$, $\mathscr{E}_w\hat{P}_{i\cdot}^{i\cdot} = \hat{p}_i^2$, and $\mathscr{E}_w\hat{P}_{j\cdot}^{i\cdot} = 2\hat{p}_i\hat{p}_j$, we can show that

$$\mathscr{E}_w\mathscr{E}_x(\hat{\sigma}_{wg}^2) = \frac{2n-1}{2(n-1)}2\sum_r\Big(1 - \sum_i \hat{p}_i^2\Big)\sigma_{x_r}^2$$

$$\mathscr{E}_w\mathscr{E}_x(\hat{\sigma}_{wHW}^2) = -\frac{1}{2(n-1)}2\sum_r\Big(1 - \sum_i \hat{p}_i^2\Big)\sigma_{x_r}^2$$

$$\mathscr{E}_w\mathscr{E}_x(\hat{\sigma}_w^2) = 2\sum_r\Big(1 - \sum_i \hat{p}_i^2\Big)\sigma_{x_r}^2.$$

Thus the lack of sample Hardy-Weinberg equilibrium is to reduce the variance on the average slightly. As an average over replicate populations $\mathscr{E}_b\mathscr{E}_w\mathscr{E}_x(\hat{\sigma}_w^2) = 2\sum_r(1-\theta_1)(1-q_2)\sigma_{x_r}^2$.

The sampling component of variance is defined as

$$\mathrm{Var}_w(\hat{\sigma}_w^2) = \mathscr{E}_b\mathscr{E}_w\mathscr{E}_x(\hat{\sigma}_w^2)^2 - \mathscr{E}_b\mathscr{E}_w\mathscr{E}_x(\hat{\sigma}_w^2\tilde{\sigma}_w^2) = \mathscr{E}_b\mathscr{E}_x(\mathscr{E}_w(\hat{\sigma}_w^2)^2 - (\mathring{\sigma}_w^2)^2) \tag{17}$$

which is the difference between the variance of unrelated samples and the covariance of replicate samples within populations, where $\hat{\sigma}_w^2$ and $\tilde{\sigma}_w^2$ are used to denote for two distinct samples within a population, since $\mathscr{E}_b\mathscr{E}_w\mathscr{E}_x(\hat{\sigma}_w^2\tilde{\sigma}_w^2) = \mathscr{E}_b\mathscr{E}_x(\mathring{\sigma}_w^2)^2$ for an unbiased estimate of σ_w^2. Let $\tilde{\sigma}_w^2$ be defined similarly as $\hat{\sigma}_w^2$ above but with all $\hat{}$ being replaced by $\tilde{}$. Since we decompose $\hat{\sigma}_w^2$ into components and also in our decomposition the

components $\hat{\sigma}^2_{wg}$ and $\hat{\sigma}^2_{wHW}$ are correlated, the sampling component of variance of the variance comprises

$$\mathrm{Var}_w(\hat{\sigma}^2_w) = \mathrm{Var}_w(\hat{\sigma}^2_{wg}) + 2\,\mathrm{Cov}_w(\hat{\sigma}^2_{wg}, \hat{\sigma}^2_{wHW}) + \mathrm{Var}_w(\hat{\sigma}^2_{wHW}) + \mathrm{Var}_w(\hat{\sigma}^2_{wL})$$

with

$$\begin{aligned}\mathrm{Var}_w(\hat{\sigma}^2_{wg}) = \frac{4n^2}{(n-1)^2}\Big\{&\sum_r \mathscr{E}(\hat{Z}^2_{r1} + \hat{Z}^2_{r2} + \hat{Z}^2_{r3} \\ &- 2\hat{Z}_{r1}\hat{Z}_{r2} - \hat{Z}_{r1}\tilde{Z}_{r1} - \hat{Z}_{r2}\tilde{Z}_{r2} \\ &- \hat{Z}_{r3}\tilde{Z}_{r3} + 2\hat{Z}_{r1}\tilde{Z}_{r2}) \\ &+ \sum_{r\neq s}\sum \mathscr{E}(\hat{Z}_{r1}\hat{Z}_{s1} + \hat{Z}_{r2}\hat{Z}_{s2} - 2\hat{Z}_{r1}\hat{Z}_{s2} \\ &- \hat{Z}_{r1}\tilde{Z}_{s1} - \hat{Z}_{r2}\tilde{Z}_{s2} + 2\hat{Z}_{r1}\tilde{Z}_{s2})\Big\}\end{aligned}$$

$$\begin{aligned}\mathrm{Var}_w(\hat{\sigma}^2_{wHW}) = \frac{4n^2}{(n-1)^2}\Big\{&\sum_r \mathscr{E}(\hat{Z}^2_{r4} + \hat{Z}^2_{r2} + \hat{Z}^2_{r5} \\ &+ \hat{Z}^2_{r3} - 2\hat{Z}_{r2}\hat{Z}_{r4} - 2\hat{Z}_{r3}\hat{Z}_{r5} \\ &- \hat{Z}_{r4}\tilde{Z}_{r4} - \hat{Z}_{r2}\tilde{Z}_{r2} \\ &- \hat{Z}_{r5}\tilde{Z}_{r5} - \hat{Z}_{r3}\tilde{Z}_{r3} + 2\hat{Z}_{r2}\tilde{Z}_{r4} \\ &+ 2\hat{Z}_{r3}\tilde{Z}_{r5}) \\ &+ \sum_{r\neq s}\sum \mathscr{E}(\hat{Z}_{r4}\hat{Z}_{s4} + \hat{Z}_{r2}\hat{Z}_{s2} - 2\hat{Z}_{r2}\hat{Z}_{s4} \\ &- \hat{Z}_{r4}\tilde{Z}_{s4} - \hat{Z}_{r2}\tilde{Z}_{s2} + 2\hat{Z}_{r2}\tilde{Z}_{s4})\Big\}\end{aligned}$$

$$\begin{aligned}2\,\mathrm{Cov}_w(\hat{\sigma}^2_{wg}, \hat{\sigma}^2_{wHW}) = \frac{8n^2}{(n-1)^2}\Big\{&\sum_r \mathscr{E}(\hat{Z}_{r1}\hat{Z}_{r4} \\ &+ \hat{Z}^2_{r2} + \hat{Z}^2_{r3} - \hat{Z}_{r1}\hat{Z}_{r2} \\ &- \hat{Z}_{r2}\hat{Z}_{r4} - \hat{Z}_{r3}\hat{Z}_{r5} \\ &- \hat{Z}_{r1}\tilde{Z}_{r4} - \hat{Z}_{r2}\tilde{Z}_{r2} \\ &- \hat{Z}_{r3}\tilde{Z}_{r3} + \hat{Z}_{r1}\tilde{Z}_{r2} + \hat{Z}_{r2}\tilde{Z}_{r4} + \hat{Z}_{r3}\tilde{Z}_{r5}) \\ &+ \sum_{r\neq s}\sum \mathscr{E}(\hat{Z}_{r1}\hat{Z}_{s4} + \hat{Z}_{r2}\hat{Z}_{s2} - \hat{Z}_{r1}\hat{Z}_{s2} \\ &- \hat{Z}_{r2}\hat{Z}_{s4} - \hat{Z}_{r1}\tilde{Z}_{s4} \\ &- \hat{Z}_{r2}\tilde{Z}_{s2} + \hat{Z}_{r1}\tilde{Z}_{s2} + \hat{Z}_{r2}\tilde{Z}_{s4})\Big\}\end{aligned}$$

$$\begin{aligned}\mathrm{Var}_w(\hat{\sigma}^2_{wL}) = \frac{8n^2}{(n-1)^2}\sum_{r\neq s}\sum \mathscr{E}(&\hat{Z}^2_{rs6} + 4\hat{Z}^2_{rs7} \\ &- 4\hat{Z}_{rs6}\hat{Z}_{rs7} - \hat{Z}_{rs6}\tilde{Z}_{rs6} \\ &- 4\hat{Z}_{rs7}\tilde{Z}_{rs7} + 4\hat{Z}_{rs6}\tilde{Z}_{rs7}).\end{aligned}$$

These involve a series of complex sampling gene and genotype frequency functions. These functions are listed in Table 3 and analyzed in detail in

TABLE 3

Sampling gene and genotypic frequency functions associated with $\mathscr{L}_x\mathscr{L}_w(\hat{\sigma}^2_w\tilde{\sigma}^2_w)$

	μ_{4_r}	$\sigma^4_{x_r}$ or $\sigma^2_{x_r}\sigma^2_{x_s}$
$\mathscr{L}_x\hat{Z}_{r1}\tilde{Z}_{r1}$	$\sum_i \hat{p}_i\tilde{p}_i$	$\sum_{i\neq j} \hat{p}_i\tilde{p}_j$
$\mathscr{L}_x\hat{Z}_{r2}\tilde{Z}_{r2}$	$\sum_i \hat{p}^2_i\tilde{p}^2_i$	$\sum_{i\neq j} \hat{p}^2_i\tilde{p}^2_j$
$\mathscr{L}_x\hat{Z}_{r3}\tilde{Z}_{r3}$		$2\sum_{i\neq j} \hat{p}_i\hat{p}_j\tilde{p}_i\tilde{p}_j$
$\mathscr{L}_x\hat{Z}_{r4}\tilde{Z}_{r4}$	$\sum_i \hat{P}^i_{i\cdot}\tilde{P}^i_{i\cdot}$	$\sum_{i\neq j} \hat{P}^i_{i\cdot}\tilde{P}^j_{j\cdot}$
$\mathscr{L}_x\hat{Z}_{r5}\tilde{Z}_{r5}$		$\sum_{i<j} \hat{P}^i_{j\cdot}\tilde{P}^i_{j\cdot}$
$\mathscr{L}_x\hat{Z}_{r1}\tilde{Z}_{r2}$	$\sum_i \hat{p}_i\tilde{p}^2_i$	$\sum_{i\neq j} \hat{p}_i\tilde{p}^2_j$
$\mathscr{L}_x\hat{Z}_{r1}\tilde{Z}_{r4}$	$\sum_i \hat{p}_i\tilde{P}^i_{i\cdot}$	$\sum_{i\neq j} \hat{p}_i\tilde{P}^j_{j\cdot}$
$\mathscr{L}_x\hat{Z}_{r2}\tilde{Z}_{r4}$	$\sum_i \hat{p}^2_i\tilde{P}^i_{i\cdot}$	$\sum_{i\neq j} \hat{p}^2_i\tilde{P}^j_{j\cdot}$
$\mathscr{L}_x\hat{Z}_{r3}\tilde{Z}_{r5}$		$2\sum_{i<j} \hat{p}_i\hat{p}_j\tilde{P}^i_{j\cdot}$
$\mathscr{L}_x\hat{Z}_{r1}\tilde{Z}_{s1}$		$\sum_i\sum_k \hat{p}_i\tilde{q}_k$
$\mathscr{L}_x\hat{Z}_{r2}\tilde{Z}_{s2}$		$\sum_i\sum_k \hat{p}^2_i\tilde{q}^2_k$
$\mathscr{L}_x\hat{Z}_{r4}\tilde{Z}_{s4}$		$\sum_i\sum_k \hat{P}^i_{i\cdot}\tilde{P}^k_{\cdot k}$
$\mathscr{L}_x\hat{Z}_{r1}\tilde{Z}_{s2}$		$\sum_i\sum_k \hat{p}_i\tilde{q}^2_k$
$\mathscr{L}_x\hat{Z}_{r1}\tilde{Z}_{s4}$		$\sum_i\sum_k \hat{p}_i\tilde{P}^k_{\cdot k}$
$\mathscr{L}_x\hat{Z}_{r2}\tilde{Z}_{s4}$		$\sum_i\sum_k \hat{p}^2_i\tilde{P}^k_{\cdot k}$
$\mathscr{L}_x\hat{Z}_{rs6}\tilde{Z}_{rs6}$		$\sum_i\sum_k (\hat{P}^{ik}_{\cdot\cdot} + \hat{P}^i_{\cdot k})(\tilde{P}^{ik}_{\cdot\cdot} + \tilde{P}^i_{\cdot k})$
$\mathscr{L}_x\hat{Z}_{rs7}\tilde{Z}_{rs7}$		$\sum_i\sum_k \hat{p}_i\tilde{p}_i\hat{q}_k\tilde{q}_k$
$\mathscr{L}_x\hat{Z}_{rs6}\tilde{Z}_{rs7}$		$\sum_i\sum_k (\hat{P}^{ik}_{\cdot\cdot} + \hat{P}^i_{\cdot k})\tilde{p}_i\tilde{q}_k$

The components of expectations with $\mathscr{L}_x\mathscr{L}_w(\hat{\sigma}^2_w)^2$ are given by replacing ˜ by ˆ in all of the above expectations.

APPENDIX C. The results for the infinite allele mutation model are

$$\begin{aligned}\mathrm{Var}_w(\hat{\sigma}^2_{wg}) = \frac{4n^2}{(n-1)^2}\Big\{&\sum_r (\alpha(1-\alpha)^2(1-\theta_1) \\ &- 2\alpha(1-\alpha)(2-3\alpha)(\theta_1 - 2\gamma_1 + \delta_1))\mu_{4_r} \\ &+ \sum_r(3\alpha(1-\alpha)^2(1-\theta_1) \\ &- 2\alpha(1-\alpha)(2-3\alpha) \\ &\cdot (1 - 3\theta_1 + 2\gamma_1 - 3\delta_1 + 3\delta_2))\sigma^4_{x_r} \\ &+ \sum_{r\neq s}\sum(2\alpha^2(1-\alpha)(\tilde{\theta}_1 - \tilde{\gamma}_1) \\ &+ 2\alpha(1-\alpha)(2-3\alpha)(\tilde{\gamma}_1 - \tilde{\delta}_1))\sigma^2_{x_r}\sigma^2_{x_s}\Big\}\end{aligned}$$

$$\begin{aligned}&\mathrm{Var}_w(\hat{\sigma}^2_{wHW}) \\ &= \frac{\alpha^2}{(1-\alpha)^2}\mathrm{Var}_w(\hat{\sigma}^2_{wg}) + \frac{8\alpha}{(1-\alpha)(1-2\alpha)} \\ &\cdot\Big\{\sum_r((\theta_1 - 2\gamma_1 + \delta_1)\mu_{4_r} \\ &+ (1 - 3\theta_1 + 2\gamma_1 - 3\delta_1 + 3\delta_2)\sigma^4_{x_r}) \\ &+ \sum_{r\neq s}\sum(\tilde{\theta}_1 - 2\tilde{\gamma}_1 + \tilde{\delta}_1)\sigma^2_{x_r}\sigma^2_{x_s}\Big\}\end{aligned}$$

$$2\,\mathrm{Cov}_w(\hat{\sigma}^2_{wg}, \hat{\sigma}^2_{wHW}) = \frac{-2\alpha}{(1-\alpha)}\mathrm{Var}_w(\hat{\sigma}^2_{wg})$$

$$\mathrm{Var}_w(\hat{\sigma}^2_{wL}) = \frac{8n^2}{(n-1)^2}\sum_{r\neq s}\sum\Big\{2\alpha(1-2\alpha)(1-\theta_{1_r} - \theta_{1_s} + \tilde{\theta}_1) - 8\alpha^2(1-2\alpha)(\tilde{\theta}_1 - \tilde{\gamma}_1) - 2\alpha(1-2\alpha)(1-6\alpha)(\tilde{\theta}_1 - 2\tilde{\gamma}_1 + \tilde{\delta}_1)\Big\}\sigma^2_{x_r}\sigma^2_{x_s}$$

and

$$\mathrm{Var}_w(\hat{\sigma}^2_w) = 4\sum_r \alpha(1-\theta_1)(\mu_{4_r} + 3\sigma^4_{x_r}) + 24\sum_{r\neq s}\sum \alpha(1-\theta_{1_r} - \theta_{1_s} + \tilde{\theta}_1)\sigma^2_{x_r}\sigma^2_{x_s} - \frac{2\alpha(1-6\alpha)}{1-2\alpha}\mathscr{E}(\mathring{\sigma}^2_w)^2$$

where $\alpha = 1/2n$. To the order of $1/n$, these are approximately

$$\mathrm{Var}_w(\hat{\sigma}^2_{wg}) \simeq \frac{2}{n}\Big\{\sum_r(1-5\theta_1+8\gamma_1-4\delta_1)(\mu_{4_r}-3\sigma^4_{x_r}) + 2\sum_r(1-3\theta_1+8\gamma_1-6\delta_2)\sigma^4_{x_r} + 2\sum_{r\neq s}\sum\eta_a\sigma^2_{x_r}\sigma^2_{x_s}\Big\} \tag{18}$$

$$\mathrm{Var}_w(\hat{\sigma}^2_{wHW}) \simeq \frac{4}{n}\Big\{\sum_r(\theta_1-2\gamma_1+\delta_1)(\mu_{4_r}-3\sigma^4_{x_r}) + \sum_r(1-4\gamma_1+3\delta_2)\sigma^4_{x_r} + \sum_{r\neq s}\sum\eta_d\sigma^2_{x_r}\sigma^2_{x_s}\Big\} \tag{19}$$

$$2\,\mathrm{Cov}_w(\hat{\sigma}^2_{wg}, \hat{\sigma}^2_{wHW}) \simeq 0 \tag{20}$$

$$\mathrm{Var}_w(\hat{\sigma}^2_{wL}) \simeq \frac{8}{n}\sum_{r\neq s}\sum\{(1-\theta_{1_r})(1-\theta_{1_s}) + \eta_a + \eta_b\}\sigma^2_{x_r}\sigma^2_{x_s} \tag{21}$$

and

$$\mathrm{Var}_w(\hat{\sigma}^2_w) \simeq \frac{2}{n}\Big\{\sum_r((1-3\theta_1+4\gamma_1-2\delta_1)\cdot(\mu_{4_r}-3\sigma^4_{x_r}) + 2(2-3\theta_1+4\gamma_1-3\delta_1)\sigma^4_{x_r}) + 2\sum_{r\neq s}\sum(2-2\theta_{1_r}-2\theta_{1_s}+\tilde{\theta}_1+4\tilde{\gamma}_1-3\tilde{\delta}_1)\sigma^2_{x_r}\sigma^2_{x_s}\Big\}. \tag{22}$$

What is analyzed here is the component of variance of the variance among samples within populations. The total variance of $\hat{\sigma}^2_w$ contains both the components of variance among populations and among samples within populations

$$\mathrm{Var}(\hat{\sigma}^2_w) = \mathrm{Var}_w(\hat{\sigma}^2_w) + \mathrm{Var}_b(\mathring{\sigma}^2_w).$$

The advantage of partitioning the variance of variance into components is that variances of averages can be easily accommodated. For example, the average genetic variance for k equal sized samples from the same replicate population is $\bar{\sigma}^2_w = \sum_{i=1}^k \hat{\sigma}^2_{w_i}/k$. The variance among $\bar{\sigma}^2_w$'s from different replicate populations is

$$\mathrm{Var}(\bar{\sigma}^2_w) = \frac{\mathrm{Var}_w(\hat{\sigma}^2_w)}{k} + \mathrm{Var}_b(\mathring{\sigma}^2_w).$$

Numerical Analysis

Founder populations: We consider three founder populations in numerical analysis: an infinite equilibrium population, a random finite equilibrium population, and a random fixed population, *i.e.*, fixed for genes in proportion to their equilibrium frequencies. For an infinite equilibrium founder population all descent measures are zero initially, and all primed descent measures among populations will remain zero except $\delta'_1 = \theta^2_1$ and $\tilde{\delta}'_1 = \theta_{1_r}\theta_{1_s}$. For a finite equilibrium founder population, the initial values of the descent measures are the equilibrium values of the descent measures within the populations with size N_0. If the subpopulations in subsequent generations are of the same size as the finite equilibrium founder population (*i.e.*, $N = N_0$) the descent measures within populations will not change. Otherwise every descent measure will change. For a random fixed founder population all descent measures are initially one.

Linkage structure: Linkage affects descent measures involving two loci. For numerical analysis we assume that loci are randomly distributed in the genome with no interference within chromosomes. In this case the relation between recombination rate c and map distance d is $c = (1 - e^{-2d})/2$ for loci in the same chromosome. The recombination rate is 0.5 for loci in different chromosomes. In analysis we first sample locations for the m loci for a given number of chromosomes and map lengths and then calculate the pairwise recombination rates among m loci accordingly.

Without mutation: Let us first consider the dynamics of the variance of the genetic variances ignoring recurrent mutation. This discussion is of relevance to small populations in a relatively short time such as control populations in many experiments. Here we assume the allelic effects x's to be normally distributed so that $\mu_4 = 3\sigma^4_x$ and let $\sigma^2_x = 1$. Figure 1 depicts the relative magnitudes of $\mathrm{Var}_b(\mathring{\sigma}^2_{wb})$, $\mathrm{Var}_b(\mathring{\sigma}^2_{wL})$, $\mathrm{Var}_w(\hat{\sigma}^2_{wg})$, $\mathrm{Var}_w(\hat{\sigma}^2_{wHW})$ and $\mathrm{Var}_w(\hat{\sigma}^2_{wL})$ for two linkage structures with numbers of chromosomes $M = 3$ and 10 and each chromosome having map length $L = 1$ Morgan. The values in the figure depend very much on the number of loci m and sample size n. As m increases the relative weight of joint loci terms increases and as n increases the sampling variance decreases. We use $m = 100$ and $n = 100$ as representative values. Com-

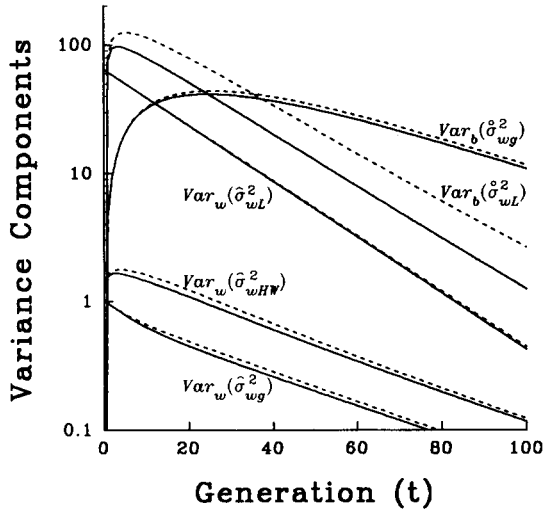

FIGURE 1.—The transient behaviors of the five components of variance of genetic variance within populations without mutation for two linkage structures: the solid lines for the number of chromosome $M = 10$ and the dashed lines for $M = 3$ with the length of each chromosome $L = 1$ Morgan. The founder population is at equilibrium with $N_0 = 1000$ and $u = 10^{-4}$. Other parameters are $N = 20$, $n = 100$, $m = 100$ and $\sigma_x^2 = 1$.

pared with the other components, $\mathrm{Var}_w(\hat{\sigma}^2_{wg})$ and $\mathrm{Var}_w(\hat{\sigma}^2_{wHW})$ are one or two orders smaller and $\mathrm{Var}_w(\hat{\sigma}^2_{wHW}) > \mathrm{Var}_w(\hat{\sigma}^2_{wg})$. Only $\mathrm{Var}_b(\mathring{\sigma}^2_{wL})$ increases significantly as linkage increases. This is because in other components the variances are dominated by terms other than identity disequilibrium coefficients. The identity disequilibrium coefficients η_a, η_b and η_d have been examined in detail by TACHIDA and COCKERHAM (1989). Depending on the population size and recombination rate, they are generally small in magnitude, start from initial value zeros (if there is no initial identity disequilibrium), quickly increase to their maximum values and then decrease. η_d is of order $1/N$; whereas η_a and η_b are of order $1/N^2$. As time goes on, the identity disequilibrium coefficients decrease and the whole variance of the genetic variance within populations is dominated by $\mathrm{Var}_b(\mathring{\sigma}^2_{wg})$. $\mathrm{Var}_w(\hat{\sigma}^2_{wL})$ is dominated by the coefficient $(1 - \theta_{1_r})(1 - \theta_{1_s})$ rather than η_a and η_b. So it can have an appreciable effect for n as large as 100 and is relatively independent of linkage structure. Since initially most of the variation is due to sampling rather than differentiation of populations, $\mathrm{Var}_w(\hat{\sigma}^2_{wL})$ can play a significant role in the first several generations.

Having made these qualitative discussions we now approximate the variance of the variance in terms of the variance. In APPENDIX B, we approximate the variance of linkage disequilibrium. For a pair of loci the dynamics of η_d can be approximated as

$$\eta_{d_t} \simeq \frac{(1-c)^2 + c^2}{2Nc(2-c)}[1 - (1-c)^{2t}](1 - \theta_{1_n})(1 - \theta_{1_s}),$$

ignoring recurrent mutation and letting $\eta_{d_0} = 0$, where c is the recombination rate. Thus in the absence of linkage,

$$\mathrm{Var}_b(\mathring{\sigma}^2_{wL})_t \simeq \frac{2}{3N}\left(1 - \left(\frac{1}{4}\right)^t\right)\sigma^4_{w_t}.$$

Since the sampling variance is dominated by $\mathrm{Var}_w(\hat{\sigma}^2_{wL})$ and $\mathrm{Var}_w(\hat{\sigma}^2_{wL})$ is dominated by the term with $(1 - \theta_{1_r})(1 - \theta_{1_s})$, the sampling variance can be approximated as

$$\mathrm{Var}_w(\hat{\sigma}^2_w)_t \simeq \frac{2}{n}\sigma^4_{w_t}.$$

This is in agreement with Equation 4 of WEIR and HILL (1980). $\mathrm{Var}_b(\mathring{\sigma}^2_{wg})$ depends on the distribution of allelic effects. For the normal distribution of allelic effects, it can be roughly approximated as

$$\mathrm{Var}_w(\mathring{\sigma}^2_{wg})_t \simeq \frac{24}{5}m(1 - \theta_{1_0}) \cdot \left\{\left(1 - \frac{1}{2N}\right)^t - \left(1 - \frac{1}{2N}\right)^{2t}\right\}\sigma^4_x.$$

Thus, the coefficient of variation of the genetic variance within populations can be approximated as

$$CV(\hat{\sigma}^2_w)_t = \frac{\sqrt{\mathrm{Var}(\hat{\sigma}^2_w)_t}}{\sigma^2_{w_t}} \simeq \sqrt{\frac{6}{5m(1-\theta_{1_0})}\left\{\left(1-\frac{1}{2N}\right)^{-t} - 1\right\} + \frac{2}{3N}\left(1 - \left(\frac{1}{4}\right)^t\right) + \frac{2}{n}} \quad (23)$$

in the absence of linkage or with a large number of chromosomes. As $\sigma^2_{w_t}$ decreases in the rate of $(1 - 1/2N)^t$, $CV(\hat{\sigma}^2_w)_t$ increases in the rate of $(1 - 1/2N)^{-t/2}$ after a few generations. It takes about

$$t > -\ln\left(1 + \frac{5m(1-\theta_{1_0})}{9N}\right) \Big/ \ln\left(1 - \frac{1}{2N}\right)$$

generations for $\mathrm{Var}_b(\mathring{\sigma}^2_{wg})_t$ to become larger than $Var_b(\mathring{\sigma}^2_{wL})_t$, which is about 23 generations for the parameters of Figure 1, ignoring linkage. When $N \gg m$, the condition is $t > 10m(1 - \theta_{1_0})/9$ generations. Figure 2 plots the dynamic behaviors of σ^2_w, $\mathrm{Var}(\hat{\sigma}^2_w)$ and $CV(\hat{\sigma}^2_w)$ for $M = 10$ chromosomes, along with the approximation (23) for free recombination. The approximation is reasonably good.

With mutation: With recurrent mutation the dynamics of variance of genetic variance within populations is a complex process. It depends very much on founder populations, linkage structures and equilibrium values. When founder populations contain large genetic variances the variance of the genetic variance can first increase to its maximum value and then decrease to the equilibrium value. On the other hand, when founder populations contain relatively small or

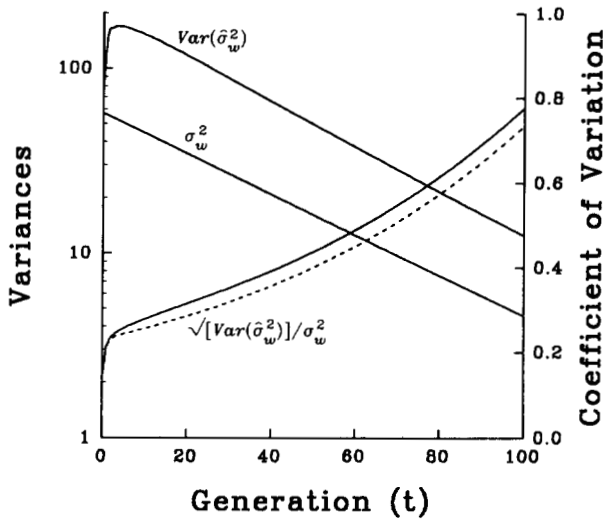

FIGURE 2.—The transient behaviors of the mean, variance and coefficient of variation of genetic variance within populations without mutation for $M = 10$. Other parameters are the same as Figure 1. The dashed line is the approximation of Equation 23 for free recombination.

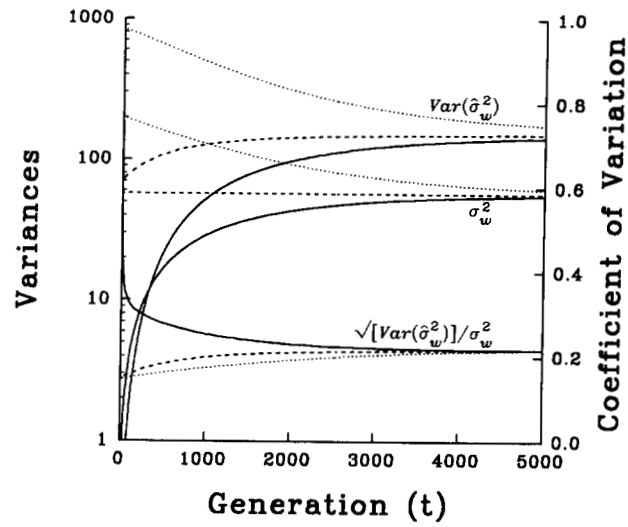

FIGURE 3.—The transient behaviors of the mean, variance and coefficient of variation of genetic variance within populations with mutation for three founder populations: the solid lines for the random fixed populations; the dashed lines for the finite equilibrium population with $N_0 = 1000$; and the dotted lines for the infinite equilibrium population. $N = 1000$, $n = 100$, $m = 100$, $M = 10$, $u = 10^{-4}$ and $\sigma_x^2 = 1$.

no genetic variances the dynamics is a monotone process with the variance of the genetic variance increasing from its initial value to the equilibrium. Initially most of variation is due to the sampling variance, $\mathrm{Var}_w(\hat{\sigma}_{wL}^2)$, [$\mathrm{Var}_w(\hat{\sigma}_{wg}^2)$ and $\mathrm{Var}_w(\hat{\sigma}_{wHW}^2)$ are always negligible] and the rapid build-up of the variance of linkage disequilibrium $\mathrm{Var}_b(\mathring{\sigma}_{wL}^2)$. However, gradually in about 30 generations (for the parameters of Figure 3) the variance becomes dominated by the component of genic variance $\mathrm{Var}_b(\mathring{\sigma}_{wg}^2)$. Figure 3 plots the transient behavior of the mean, variance and coefficient of variation of the genetic variance within populations for three founder populations.

At equilibrium the variance of the genetic variance within populations is essentially a function of the parameter $\phi = 4Nu$. If we ignore the identity disequilibrium coefficients η_b at equilibrium, from (7)

$$\mathrm{Var}_b(\mathring{\sigma}_{wg}^2)_\infty \simeq 4 \sum_r (2\theta_{1_\infty} - 4\gamma_{1_\infty} + 3\delta_{2_\infty} - \theta_{1_\infty}^2)\sigma_{x_r}^4$$

for a normal distribution of x. Approximately, the one-locus descent measures at equilibrium are

$$\theta_{1_\infty} = \frac{1}{1+\phi}$$

$$\gamma_{1_\infty} = \frac{2}{(1+\phi)(2+\phi)}$$

$$\delta_{2_\infty} = \frac{6+\phi}{(1+\phi)(2+\phi)(3+\phi)}.$$

Thus

$$\mathrm{Var}_b(\mathring{\sigma}_{wg}^2)_\infty \simeq 4m\phi\left\{\frac{1}{(1+\phi)^2} + \frac{\phi}{(1+\phi)(2+\phi)(3+\phi)}\right\}\sigma_x^4$$

Taking also the variance of linkage disequilibrium and the sampling variance into account

$$CV(\hat{\sigma}_w^2)_\infty \simeq \sqrt{\frac{1}{m\phi}\left\{1 + \frac{\phi(1+\phi)}{(2+\phi)(3+\phi)}\right\} + \frac{2}{3N} + \frac{2}{n}}$$

$$\simeq \sqrt{\frac{1}{4Num} + \frac{2}{3N} + \frac{2}{n}}$$

for ϕ small in the absence of linkage, which agrees with the principal terms of the approximation of LYNCH and HILL (1986) for two alleles under a different mutation model (they did not analyze the sampling variance). Figure 4 depicts the equilibrium values of σ_w^2, $\mathrm{Var}_b(\mathring{\sigma}_{wg}^2)$ and $CV(\mathring{\sigma}_{wg}^2)$ against ϕ for $m = 100$. For ϕ both very small and very large $\mathrm{Var}_b(\mathring{\sigma}_{wg}^2)$ is a small value. The maximum equilibrium value of $\mathrm{Var}_b(\mathring{\sigma}_{wg}^2)$ lies between $\phi = 1$ and 2. When ϕ is large, however, the variance of linkage disequilibrium and the sampling variance will be important at equilibrium.

VARIANCE OF GENETIC VARIANCE BETWEEN POPULATIONS

Unlike the genetic variance within populations, for which there is a real component of variance among populations, the genetic variance between populations is a parameter and variances arise only among sample estimates analogous to, but more complicated than, the sampling variance of the genetic variance within populations. Suppose that observations are taken on S replicate populations which provides an estimate of the genetic variance between populations. That is, however, just a single realization of the genetic vari-

ance between populations. If, for conceptual reasoning, we execute such an experiment many times in a way to keep everything identical except sampling, there will still be some variation among different estimates of the genetic variance between populations from different experiments. Clearly this sampling variance of the genetic variance between populations depends on the number of replicate populations of each experiment. If, for example, S is very large, approaching infinity, there will be no variance of the genetic variance between populations.

For S replicate populations, the genetic variance between populations is given by

$$\dot{\sigma}_b^2=\frac{1}{S-1}\left\{\sum_y\left(2\sum_r\sum_i \dot{p}_i x_i\right)^2-\frac{1}{S}\left(2\sum_y\sum_r\sum_i \dot{p}_i x_i\right)^2\right\} \quad (24)$$

which is the difference between the covariance of individuals within populations and that of individuals in distinct replicate populations, where the summation of y (and z below) is over S replicate populations. (For simplicity we have let the sample size n of each replicate population be very large in the above equation. Unless they are very small, finite sample sizes only trivially contribute to the variance of $\dot{\sigma}_b^2$.) Like the genetic variance within populations, this can also be decomposed into components corresponding to gene frequency differences between covariances of individuals within and between populations for within and between loci

$$\dot{\sigma}_b^2=\dot{\sigma}_{bg}^2+\dot{\sigma}_{bL}^2$$

with

$$\begin{aligned}\dot{\sigma}_{bg}^2&=\frac{4}{S}\sum_y\sum_r\left(\sum_i \dot{p}_i^2x_i^2+\sum_{i\neq j}\sum \dot{p}_i\dot{p}_jx_ix_j\right)\\&\quad-\frac{4}{S(S-1)}\sum_{y\neq z}\sum_r\sum\left(\sum_i \dot{p}_i\check{p}_ix_i^2+\sum_{i\neq j}\sum \dot{p}_i\check{p}_jx_ix_j\right) \quad (25)\\&=\frac{4}{S}\sum_y\sum_r\left(\dot{Z}_{r2}+\dot{Z}_{r3}\right)-\frac{4}{S(S-1)}\sum_{y\neq z}\sum_r\sum\left(\check{Z}_{r2}+\check{Z}_{r3}\right)\end{aligned}$$

$$\begin{aligned}\sigma_{bL}^2&=\frac{4}{S}\sum_y\sum_{r\neq s}\sum_i\sum_k \dot{p}_i\dot{q}_kx_ix_k\\&\quad-\frac{4}{S(S-1)}\sum_{y\neq z}\sum_{r\neq s}\sum_i\sum_k \dot{p}_i\check{q}_kx_ix_k \quad (26)\\&=\frac{4}{S}\sum_y\sum_{r\neq s}\dot{Z}_{rs7}-\frac{4}{S(S-1)}\sum_{y\neq z}\sum_{r\neq s}\check{Z}_{rs7}\end{aligned}$$

where $\dot{p}_i$ and $\check{p}_j$ denote gene frequencies from two distinct replicate populations.

The sampling variance of $\dot{\sigma}_b^2$ is defined as

$$\mathrm{Var}(\dot{\sigma}_b^2)=\mathscr{E}(\dot{\sigma}_b^2)^2-\mathscr{E}(\dot{\sigma}_b^2\grave{\sigma}_b^2) \quad (27)$$

where $\dot{\sigma}_b^2$ and $\grave{\sigma}_b^2$ are used to denote between population genetic variances from two distinct samples.

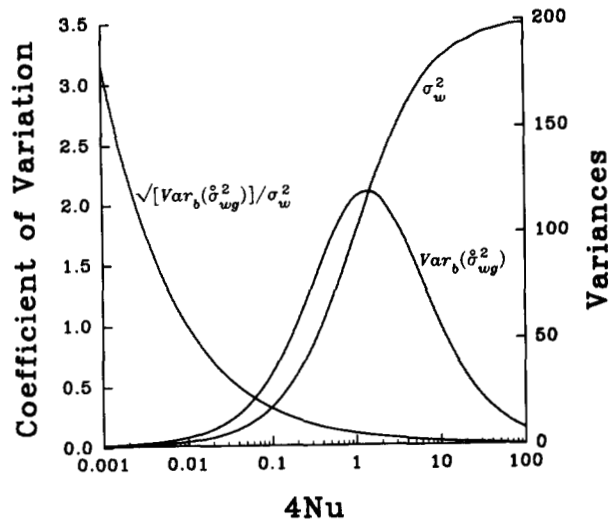

FIGURE 4.—The mean, variance and coefficient of variation of genic variance within populations at equilibrium are plotted against the parameter $4Nu$ for $m = 100$.

Let $\grave{\sigma}_{bg}^2$ and $\grave{\sigma}_{bL}^2$ be defined as in (25) and (26) but with all $\dot{}$ and $\check{}$ being replaced by $\grave{}$ and $\acute{}$. Then we have

$$\begin{aligned}\mathrm{Var}(\dot{\sigma}_{bg}^2)&=\frac{16}{S}\mathscr{E}\Big\{\sum_r(\dot{Z}_{r2}^2-\dot{Z}_{r2}\check{Z}_{r2}+\dot{Z}_{r3}^2-\dot{Z}_{r3}\check{Z}_{r3}\\&\quad-4\dot{Z}_{r2}\grave{Z}_{r2}+4\dot{Z}_{r2}\acute{Z}_{r2}-4\dot{Z}_{r3}\grave{Z}_{r3}+4\dot{Z}_{r3}\acute{Z}_{r3})\\&\quad+\sum_{r\neq s}\sum(\dot{Z}_{r2}\dot{Z}_{s2}-\dot{Z}_{r2}\check{Z}_{s2}-4\dot{Z}_{r2}\grave{Z}_{s2}+4\dot{Z}_{r2}\acute{Z}_{s2}\\&\quad+\frac{2}{S-1}\Big\{\sum_r(\grave{Z}_{r2}^2-\grave{Z}_{r2}\acute{Z}_{r2}+\grave{Z}_{r3}^2-\grave{Z}_{r3}\acute{Z}_{r3})\\&\quad+\sum_{r\neq s}\sum(\grave{Z}_{r2}\grave{Z}_{s2}-\grave{Z}_{r2}\acute{Z}_{s2})\Big\}\\&\quad+\frac{4(S-2)}{S-1}\Big\{\sum_r(\grave{Z}_{r2}\grave{Z}_{r2}-\grave{Z}_{r2}\acute{Z}_{r2}+\grave{Z}_{r3}\grave{Z}_{r3}\\&\quad-\grave{Z}_{r3}\acute{Z}_{r3})+\sum_{r\neq s}\sum(\grave{Z}_{r2}\grave{Z}_{s2}-\grave{Z}_{r2}\acute{Z}_{s2})\Big\}\Big\}\\\mathrm{Var}(\dot{\sigma}_{bL}^2)&=\frac{32}{S}\mathscr{E}\sum_{r\neq s}\sum\Big\{(\dot{Z}_{rs7}^2-\dot{Z}_{rs7}\check{Z}_{rs7}-4\dot{Z}_{rs7}\grave{Z}_{rs7}\\&\quad+4\dot{Z}_{rs7}\acute{Z}_{rs7})+\frac{2}{S-1}(\grave{Z}_{rs7}^2-\grave{Z}_{rs7}\acute{Z}_{rs7})\\&\quad+\frac{4(S-2)}{S-1}(\grave{Z}_{rs7}\grave{Z}_{rs7}-\grave{Z}_{rs7}\acute{Z}_{rs7})\Big\}.\end{aligned}$$

This involves some additional gene frequency functions listed in Table 4. Taking the expectation with respect to replicate populations

$$\begin{aligned}\mathscr{E}_b\sum_i \dot{p}_i^3\check{p}_i&=\delta_4'+(\gamma_1+3\gamma_1'+3\delta_5'-7\delta_4')q_2\\&\quad+3(\theta_1+\theta_1'-\gamma_1-3\gamma_1'-2\delta_5'+4\delta_4')q_3\\&\quad+(1-3\theta_1-3\theta_1'+2\gamma_1+6\gamma_1'\\&\quad+3\delta_5'-6\delta_4')q_4\end{aligned}$$

TABLE 4

Gene frequency functions associated with $Var_b(\mathring{\sigma}_b^2)$

	μ_{4_r}	$\sigma_{x_r}^4$ or $\sigma_{x_r}^2\sigma_{x_s}^2$
$\mathscr{L}_x\mathring{Z}_{r2}^2$	$\sum_i \mathring{p}_i^2\check{p}_i^2$	$\sum_{i\neq j}\mathring{p}_i\mathring{p}_j\check{p}_i\check{p}_j$
$\mathscr{L}_x\check{Z}_{r3}^2$		$\sum_{i\neq j}\mathring{p}_i^2\check{p}_j^2+\sum_{i\neq j}\mathring{p}_i\mathring{p}_j\check{p}_i\check{p}_j$
$\mathscr{L}_x\dot{Z}_{r2}\ddot{Z}_{r2}$	$\sum_i \mathring{p}_i^3\check{p}_i$	$\sum_{i\neq j}\mathring{p}_i^2\mathring{p}_j\check{p}_j$
$\mathscr{L}_x\dot{Z}_{r3}\ddot{Z}_{r3}$		$2\sum_{i\neq j}\mathring{p}_i^2\mathring{p}_j\check{p}_j$
$\mathscr{L}_x\dot{Z}_{r2}\ddot{Z}_{r2}$	$\sum_i \mathring{p}_i^2\dot{p}_i\acute{p}_i$	$\sum_{i\neq j}\mathring{p}_i^2\dot{p}_j\acute{p}_j$
$\mathscr{L}_x\dot{Z}_{r3}\ddot{Z}_{r3}$		$2\sum_{i\neq j}\mathring{p}_i\dot{p}_i\mathring{p}_j\acute{p}_j$
$\mathscr{L}_x\ddot{Z}_{r2}\dot{Z}_{r2}$	$\sum_i \mathring{p}_i^2\dot{p}_i\acute{p}_i$	$\sum_{i\neq j}\mathring{p}_i\dot{p}_i\mathring{p}_j\acute{p}_j$
$\mathscr{L}_x\check{Z}_{r3}\check{Z}_{r3}$		$\sum_{i\neq j}\mathring{p}_i\dot{p}_i\mathring{p}_j\acute{p}_j+\sum_{i\neq j}\mathring{p}_i^2\dot{p}_j\acute{p}_j$
$\mathscr{L}_x\check{Z}_{r2}\check{Z}_{r2}$	$\sum_i \mathring{p}_i\check{p}_i\dot{p}_i\acute{p}_i$	$\sum_{i\neq j}\mathring{p}_i\check{p}_i\dot{p}_j\acute{p}_j$
$\mathscr{L}_x\ddot{Z}_{r2}\ddot{Z}_{r3}$		$2\sum_{i\neq j}\mathring{p}_i\check{p}_i\dot{p}_j\acute{p}_j$
$\mathscr{L}_x\ddot{Z}_{r2}\ddot{Z}_{s2}$		$\sum_i\sum_k\mathring{p}_i\check{p}_i\dot{q}_k\acute{q}_k$
$\mathscr{L}_x\dot{Z}_{r2}\ddot{Z}_{s2}$		$\sum_i\sum_k\mathring{p}_i^2\dot{q}_k\acute{q}_k$
$\mathscr{L}_x\dot{Z}_{r2}\ddot{Z}_{s2}$		$\sum_i\sum_k\mathring{p}_i^2\dot{q}_k\acute{q}_k$
$\mathscr{L}_x\ddot{Z}_{r2}\ddot{Z}_{s2}$		$\sum_i\sum_k\mathring{p}_i\check{p}_i\dot{q}_k\acute{q}_k$
$\mathscr{L}_x\ddot{Z}_{r2}\ddot{Z}_{s2}$		$\sum_i\sum_k\mathring{p}_i\check{p}_i\dot{q}_k\acute{q}_k$
$\mathscr{L}_x\ddot{Z}_{rs7}^2$		$\frac{1}{2}\sum_i\sum_k\mathring{p}_i^2\dot{q}_k^2+\frac{1}{2}\sum_i\sum_k\mathring{p}_i\check{p}_i\dot{q}_k\acute{q}_k$
$\mathscr{L}_x\dot{Z}_{rs7}\ddot{Z}_{rs7}$		$\sum_i\sum_k\mathring{p}_i^2\dot{q}_k\acute{q}_k$
$\mathscr{L}_x\dot{Z}_{rs7}\dot{Z}_{rs7}$		$\sum_i\sum_k\mathring{p}_i\dot{p}_i\dot{q}_k\acute{q}_k$
$\mathscr{L}_x\ddot{Z}_{rs7}\dot{Z}_{rs7}$		$\frac{1}{2}\sum_i\sum_k\mathring{p}_i^2\dot{q}_k\acute{q}_k+\frac{1}{2}\sum_i\sum_k\mathring{p}_i\dot{p}_i\dot{q}_k\acute{q}_k$
$\mathscr{L}_x\ddot{Z}_{rs7}\ddot{Z}_{rs7}$		$\sum_i\sum_k\mathring{p}_i\check{p}_i\dot{q}_k\acute{q}_k$

$$\mathscr{L}_b\sum_i \mathring{p}_i^2\dot{p}_i\acute{p}_i=\delta_2''+(2\gamma_1'+2\gamma_1''+\delta_1''+2\delta_3''-7\delta_2'')q_2+(\theta_1+5\theta_1'-6\gamma_1'-6\gamma_1''-2\delta_1''-4\delta_3''+12\delta_2'')q_3+(1-\theta_1-5\theta_1'+4\gamma_1'+4\gamma_1''+\delta_1''+2\delta_3''-6\delta_2'')q_4$$

$$\mathscr{L}_b\sum_i \mathring{p}_i\check{p}_i\dot{p}_i\acute{p}_i=\delta_2'''+(4\gamma_1''+3\delta_1'''-7\delta_2''')q_2+6(\theta_1'-2\gamma_1''-\delta_1'''+2\delta_2''')q_3+(1-6\theta_1'+8\gamma_1''+3\delta_1'''-6\delta_2''')q_4$$

$$\mathscr{L}_b\sum_{i\neq j}\mathring{p}_i^2\mathring{p}_j\check{p}_j=(\delta_5'-\delta_4')(1-q_2)+(\theta_1+\theta_1'-\gamma_1-3\gamma_1'-2\delta_5'+4\delta_4')(q_2-q_3)+(1-3\theta_1-3\theta_1'+2\gamma_1+6\gamma_1'+3\delta_5'-6\delta_4')(q_2^2-q_4)$$

$$\mathscr{L}_b\sum_{i\neq j}\mathring{p}_i^2\dot{p}_j\acute{p}_j=(\delta_1''-\delta_2'')(1-q_2)+(\theta_1+\theta_1'-2\gamma_1'-2\gamma_1''-2\delta_1''+4\delta_2'')(q_2-q_3)+(1-\theta_1-5\theta_1'+4\gamma_1'+4\gamma_1''+\delta_1''+2\delta_3''-6\delta_2'')(q_2^2-q_4)$$

$$\mathscr{L}_b\sum_{i\neq j}\mathring{p}_i\dot{p}_i\mathring{p}_j\acute{p}_j=(\delta_3''-\delta_2'')(1-q_2)+2(\theta_1'-\gamma_1'-\gamma_1''-\delta_3''+2\delta_2'')(q_2-q_3)+(1-\theta_1-5\theta_1'+4\gamma_1'+4\gamma_1''+\delta_1''+2\delta_3''-6\delta_2'')(q_2^2-q_4)$$

$$\mathscr{L}_b\sum_{i\neq j}\mathring{p}_i\dot{p}_i\check{p}_j\acute{p}_j=(\delta_1'''-\delta_2''')(1-q_2)+2(\theta_1'-2\gamma_1''-\delta_1'''+2\delta_2''')(q_2-q_3)+(1-6\theta_1'+8\gamma_1''+3\delta_1'''-6\delta_2''')(q_2^2-q_4).$$

$$\mathscr{L}_b\sum_i\sum_k\mathring{p}_i^2\dot{q}_k\check{q}_k=\tilde{\delta}_3'+(\theta_{1_r}-\tilde{\delta}_3')q_{2_s}+(\theta_{1_s}'-\tilde{\delta}_3')q_{2_r}+(1-\theta_{1_r}-\theta_{1_s}'+\tilde{\delta}_3')q_{2_r}q_{2_s}$$

$$\mathscr{L}_b\sum_i\sum_k\mathring{p}_i\check{p}_i\dot{q}_k\check{q}_k=\tilde{\delta}_2'+(\theta_{1_r}'-\tilde{\delta}_2')q_{2_s}+(\theta_{1_s}'-\tilde{\delta}_2')q_{2_r}+(1-\theta_{2_r}'-\theta_{1_r}'-\theta_{1_s}'+\tilde{\delta}_2')q_{2_r}q_{2_s}$$

$$\mathscr{L}_b\sum_i\sum_k\mathring{p}_i^2\dot{q}_k\acute{q}_k=\tilde{\delta}_2''+(\theta_{1_r}-\tilde{\delta}_2'')q_{2_s}+(\theta_{1_s}'-\tilde{\delta}_2'')q_{2_r}+(1-\theta_{1_r}-\theta_{1_s}'+\tilde{\delta}_2'')q_{2_r}q_{2_s}$$

$$\mathscr{L}_b\sum_i\sum_k\mathring{p}_i\check{p}_i\dot{q}_k\acute{q}_k=\tilde{\delta}_1''+(\theta_{1_r}'-\tilde{\delta}_1'')q_{2_s}+(\theta_{1_s}'-\tilde{\delta}_1'')q_{2_r}+(1-\theta_{1_r}'-\theta_{1_s}'+\tilde{\delta}_1'')q_{2_r}q_{2_s}$$

$$\mathscr{L}_b\sum_i\sum_k\mathring{p}_i\check{p}_i\dot{q}_k\acute{q}_k=\tilde{\delta}_1'''+(\theta_{1_r}'-\tilde{\delta}_1''')q_{2_s}+(\theta_{1_s}'-\tilde{\delta}_1''')q_{2_r}+(1-\theta_{1_r}'-\theta_{1_s}'+\tilde{\delta}_1''')q_{2_r}q_{2_s}.$$

With these we have

$$\begin{aligned}\mathrm{Var}(\mathring{\sigma}_{bg}^2)=\frac{16}{S}\Bigg\{&\sum_r\Bigg((\delta_1-\delta_2'-4\delta_4'+8\delta_2''-4\delta_2''')+\frac{2}{S-1}(\delta_2'-2\delta_2''+\delta_2''')\Bigg)\\&\cdot\Bigg((1-7q_{2_r}+12q_{3_r}-6q_{4_r})\cdot(\mu_{4_r}-3\sigma_{x_r}^4)-6(q_{2_r}-4q_{3_r}+3q_{2_r}^2)\sigma_{x_r}^4\Bigg)\\&+\sum_r\Bigg((3\delta_2-\delta_1'-2\delta_3'-12\delta_5'+8\delta_1''+16\delta_3''-12\delta_1''')+\frac{2}{S-1}(\delta_1'+2\delta_3'-2\delta_1''-4\delta_3''+3\delta_1'')\Bigg)\Bigg((q_{2_r}-2q_{3_r}+q_{4_r})(\mu_{4_r}-3\sigma_{x_r}^4)+(1-4q_{3_r}+3q_{2_r}^2)\sigma_{x_r}^4\Bigg)\\&+\sum_{r\neq s}\sum\Bigg((\tilde{\delta}_1-\tilde{\delta}_1'-4\tilde{\delta}_3'+4\tilde{\delta}_1''+4\tilde{\delta}_2''-4\tilde{\delta}_1''')\end{aligned}$$

$$+\frac{2}{S-1}(\tilde{\delta}_2' - 2\tilde{\delta}_2'' + \tilde{\delta}_1''')\Big)$$

$$\cdot (1-q_{2_r})(1-q_{2_s})\sigma_{x_r}^2\sigma_{x_s}^2\Big\}$$

$$\mathrm{Var}(\mathring{\sigma}_{bL}^2) = \frac{32}{S}\sum_{r\neq s}\sum\Big((\tilde{\delta}_1 - \tilde{\delta}_2' - 4\tilde{\delta}_3'$$

$$+6\tilde{\delta}_1'' + 2\tilde{\delta}''_2 - 4\tilde{\delta}_1''') + \frac{1}{S-1}(\tilde{\delta}_1' + \tilde{\delta}_2' \qquad (29)$$

$$-2\tilde{\delta}_1'' - 2\tilde{\delta}_2'' + 2\delta_1''')\Big)(1-q_{2_r})(1-q_{2_s})\sigma_{x_r}^2\sigma_{x_s}^2.$$

At equilibrium, ($\delta_{1_\infty}' = \theta_{1_\infty}^2$, $\tilde{\delta}_{1_\infty} = \tilde{\delta}_{1_\infty}' = \theta_{1_{r\infty}}\theta_{1_{s\infty}}$ and all other primed descent measures are zero)

$$\mathrm{Var}(\sigma_{bg}^2) = \frac{16}{S}\sum_r\Big\{\delta_{1_\infty}\Big((1 - 7q_{2_r} + 12q_{3_r} - 6q_{4_r})$$

$$\cdot(\mu_{4_r} - 3\sigma_{x_r}^4) - 6(q_{2_r} - 4q_{3_r} + 3q_{2_r}^2)\sigma_{x_r}^4\Big)$$

$$+\Big(3\delta_{2_\infty} - \frac{S-3}{S-1}\theta_{1_\infty}^2\Big)\Big((q_{2_r} - 2q_{3_r} + q_{4_r})$$

$$\cdot(\mu_{4_r} - 3\sigma_{x_r}^4) + (1 - 4q_{3_r} + 3q_{2_r}^2)\sigma_{x_r}^4\Big)\Big\}$$

$$\mathrm{Var}(\mathring{\sigma}_{bL}^2) = \frac{32}{S-1}\sum_{r\neq s}\sum\theta_{1_{r\infty}}\theta_{1_{s\infty}}(1 - q_{2_r})(1 - q_{2_s})\sigma_{x_r}^2\sigma_{x_s}^2.$$

Although these functions are complex, they can be simplified. First we note what when the number of loci m is large the variance of genetic variance between populations is effectively dominated by $\mathrm{Var}(\mathring{\sigma}_{bL}^2)$. This is largely due to the difference between $\tilde{\delta}_1'$ and $\tilde{\delta}_2'$. For equilibrium founder populations the two-locus identity disequilibrium measures $\tilde{\theta}_{1_0} - \theta_{1_{r0}}\theta_{1_{s0}}$, $\tilde{\gamma}_{1_0} - \theta_{1_{r0}}\theta_{1_{s0}}$ and $\tilde{\delta}_{1_0} - \theta_{1_{r0}}\theta_{1_{s0}}$ are very small (SERANT 1974) and effectively we can approximate $\tilde{\theta}_{1_0} = \tilde{\gamma}_{1_0} = \tilde{\delta}_{1_0} = \theta_{1_{r0}}\theta_{1_{s0}}$. With this condition, we can see from APPENDIX A that $\tilde{\delta}_{1_t}' = \theta_{1_{rt}}\theta_{1_{st}}$, $\tilde{\gamma}_{2_t}' = \tilde{\delta}_{3_t}' = \tilde{\delta}_{2_t}'' = \theta_{1_{rt}}\theta_{1_{st}}'$, and $\tilde{\theta}_{1_t}' = \tilde{\gamma}_{1_t}' = \tilde{\delta}_{2_t}' = \tilde{\gamma}_{1_t}'' = \tilde{\delta}_{1_t}'' = \delta_{1_t}''' = \theta_{1_{rt}}'\theta_{1_{st}}'$. Thus the joint loci term of $\mathrm{Var}(\mathring{\sigma}_{bg}^2)$ is a function of $\eta_{b_t} = \tilde{\delta}_{1_t} - \theta_{1_{rt}}\theta_{1_{st}}$ and is small in magnitude and effectively $\mathrm{Var}(\mathring{\sigma}_{bg}^2)$ is of order m whereas $\mathrm{Var}(\mathring{\sigma}_{bL}^2)$ is of order m^2. Consequently, the whole variance of genetic variance between populations can be approximated as

$$\mathrm{Var}(\mathring{\sigma}_b^2)_t \simeq \mathrm{Var}(\mathring{\sigma}_{bL}^2)_t$$

$$\simeq \frac{32}{S-1}\sum_{r\neq s}\sum(\theta_{1_{rt}} - \theta_{1_{rt}}')(\theta_{1_{st}} - \theta_{1_{st}}')$$

$$\cdot (1 - q_{2_r})(1 - q_{2_s})\sigma_{x_r}^2\sigma_{x_s}^2.$$

Since $\sigma_{b_t}^2 = \mathscr{E}(\mathring{\sigma}_{b_t}^2) = 4\sum_r(\theta_{1_t} - \theta_{1_t}')(1 - q_{2_r})\sigma_{x_r}^2$, the coefficient of variation of genetic variance between populations can be approximated as

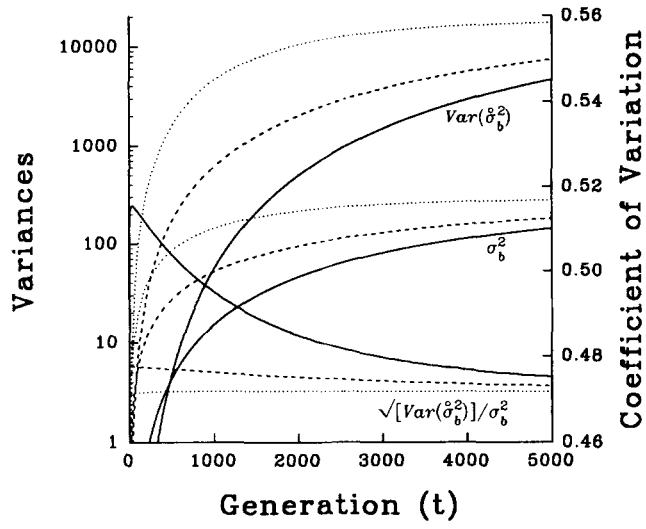

FIGURE 5.—The transient behaviors of the mean, variance and coefficient of variation of genetic variance between populations with mutation for three founder populations: the solid lines for the random fixed population; the dashed lines for the finite equilibrium population with $N_0 = 1000$; and the dotted lines for the infinite equilibrium population. $S = 10$. Other parameters are the same as Figure 3.

$$CV(\mathring{\sigma}_b^2)_t = \frac{\sqrt{\mathrm{Var}(\mathring{\sigma}_b^2)_t}}{\sigma_{b_t}^2} \simeq \sqrt{\frac{2}{S-1}} \qquad (30)$$

when the number of loci is large. This suggests that the sampling variance of $\mathring{\sigma}_b^2$ is *approximately* χ^2-distributed as LANDE (1977) and LYNCH (1988b) have assumed. This is however realized essentially by Central Limit Theorem for a very large number of loci. When the number of loci is finite the sampling distribution of $\mathring{\sigma}_b^2$ is not exactly χ^2 even though we assume that x_i's are independent and identical normal variables. This is because gene frequencies are variables, not constant (*i.e.*, $\sum_r\sum_i \hat{p}_i\hat{p}_i \neq \sum_r\sum_i p_ip_i$ although they are equal by expectation) so that the population means with indefinitely large samples, $(2\sum_r\sum_i \hat{p}_ix_i)$'s, are not symmetrically normally distributed (see RAO 1973, pp. 182–197). When m is small (30) tends to underestimate the coefficient of variation and the coefficient of variation will depend on the distribution of allelic effects and also the number of alleles for mutation at each locus. Figure 5 plots the mean, variance and coefficient of variation of genetic variance between populations for three founder populations for a normal distribution of allelic effects and the infinite allele mutation model with $m = 100$. When populations start from a random fixed founder population, the coefficient of variation is significantly larger than that approximated by (30) even for loci as large as 100. This would suggest that if we perform population divergence experiments for populations starting from a near fixed founder population and use genetic variance between populations

to estimate genetic parameters such as the rate of input of genetic variance from mutation (LYNCH 1988a), it would tend to underestimate the sampling variance of the estimate if the approximation (30) is used. However, it seems that unless m is very small the approximation (30) is sufficiently accurate for prediction for populations starting from a heterogeneous founder population.

DISCUSSION

Previous analyses of the variance of genetic variances within and between populations concerned mostly the establishment of simple and useful approximations. This is the first systematic treatment of the subject. Our results confirm some of previous approximations, point out assumptions of simplified approximations, and also reveal some problem of previous analyses.

There are many differences between our analysis and those by AVERY and HILL (1977) and BULMER (1980) on the variance of genetic variance within populations without mutation. Both AVERY and HILL (1977) and BULMER (1980) treated only two alleles and assume that allelic effects are fixed. We analyze a general multiple allele model and assume random allelic effects so that the results depend on the distribution of allelic effects. This contributes to the difference between our and their results. We use identity by descent measures in deriving the results. The advantage of that is that the exact solutions can be readily obtained and the results apply to any number of alleles and to any generation.

AVERY and HILL (1977) and BULMER (1980) did not distinguish between sample and population size. They derived results by analyzing transition equations for moments of the disequilibria rather than using transition arguments just for descent measures and translating these to observable quantities such as linkage disequilibrium only in the sampling generation. Also unlike BULMER (1976), BULMER (1980) included only the gametic linkage disequilibrium in the definition of the genetic variance due to linkage disequilibrium and grouped the non-gametic linkage disequilibrium into the Hardy-Weinberg disequilibrium component. Thus he obtained the approximation of $(5/3N)\sigma_w^2$ for the variance of the genetic variance due to linkage disequilibrium in the absence of linkage (which should be $(2/3N + 1/n)\sigma_w^4$ in our notation, including half of the sampling variance), and $(1/N)\sigma_w^4$ (should be $(1/n)\sigma_w^4$) for the variance of the genetic variance due to "Hardy-Weinberg disequilibrium." Whereas the variance of genetic variance due to linkage disequilibrium depends critically on linkage structure, the sampling variance, predominantly due to sampling linkage disequilibrium, is largely independent of linkage structure.

Hardy-Weinberg disequilibrium exists only in samples. The lack of sample Hardy-Weinberg equilibrium is to reduce the genetic variance on the average slightly. The squared coefficient of variation of the genetic variance due to sampling Hardy-Weinberg disequilibrium is small and of order $1/Nn$.

Initially the variance of genetic variance within populations is mostly due to linkage disequilibrium and sampling. Linkage disequilibrium is however a transient phenomenon. In the long-run the variance of the genetic variance is likely to be dominated by the differentiation of populations on gene frequencies since at equilibrium the squared coefficient of variation is $1/(4Num) + 2/3N + 2/n$ (LYNCH and HILL 1986; this study) where the first term is due to genic variance, the second due to linkage disequilibrium and the third due to sampling. Thus, unless the total mutation rate for the character mu is larger than $3/[8(1 + 3N/n)]$, the variance of genic variance will be an important component in the variance of genetic variance within populations in the long-run. Approximately it takes about $t > 10m(1 - \theta_{1_0})/9$ generations for $\mathrm{Var}_b(\mathring{\sigma}^2_{wg})$ to become a dominant factor for organisms with a large number of chromosomes and $N \gg m$. When N is of order m or smaller, the time needed is shorter. $(1 - \theta_{1_0})$ is the heterozygosity in the founder population. If m is of order hundreds, $m(1 - \theta_{1_0})$ is likely to be of order tens. This time scale may be too long for many short-term experiments. But there have been some long-term experiments which lasted for about 70 to 80 generations. In this time span the variance of genetic variance within populations can be dominated by the component due to genic variance. This feature of the dynamics of the variance of variance has been overlooked in the previous discussions.

The variance of genetic variances depends on the distribution of allelic effects however. This is true especially for the component of variance of genetic variance within populations due to genic variance. In the numerical analysis we used the normal distribution of allelic effects. If the distribution is leptokurtic the variance will be larger than that under the normal distribution; on the other hand if the distribution is platykurtic it will be smaller. It has been widely believed that the distribution of allelic effects is probably not normal, but highly leptokurtic (*e.g.*, ROBERTSON 1967). By using *P* element mutagenesis on *Drosophila melanogaster*, T. F. C. MACKAY, R. LYMAN and M. JACKSON (unpublished data) showed that the distributions of effects of *P* element inserts on abdominal and sternopleural bristle numbers are highly leptokurtic. If that is true in general, the relative importance of the component due to genic variance in the variance of genetic variance within populations will be greatly enhanced as the shape of the distribution affects only the terms within loci, not between loci.

The expected sampling variance of genetic variance between populations is approximately twice the expected variance squared divided by the sample size minus one when the number of loci is large, as expected if the means of populations are normally distributed (LANDE 1977; LYNCH 1988b). This is, however, a consequence of Central Limit Theorem that as the number of loci increases the distribution of sum of gene effects approaches normal, irrespective of underlying distributions of individual allelic effects and frequencies.

We have analyzed only additive effects of genes. An extension of the analysis to dominance is not a trivial matter. The complete description of genetic variances within and between populations with dominance for a general multiple allele model adds four additional components (COCKERHAM 1984; TACHIDA and COCKERHAM 1990). These components involve identity measures of up to four genes. An analysis of the variance of genetic variances with dominance would then involve numerous variances and covariances of different components and require identity measures of up to eight genes within and between populations. This analysis is currently being undertaken and will be presented elsewhere.

The current analysis is also only for genetic variances. Variance of phenotypic variances within and between populations will include environmental effects. This will then depend on the distribution of environmental effects, whether there is genotype-environment correlation and interaction, and whether there are common environmental effects between populations. For a very simple model in which the environmental effects are independently and normally distributed and there are no genotype-environment correlation and interaction and no common environmental effects, the expected variance of the phenotypic variance within populations will include twice the squared environmental variance and the variance of the between-population component is only trivially affected.

The analysis of variance of genetic variances within and between populations is central to many questions of quantitative genetics. It has important bearings on designing experiments and interpreting experimental results (HILL 1980; LYNCH 1988b); testing the neutral model of phenotypic evolution (LANDE 1977); and estimating genetic parameters, such as the rate of input of new genetic variance by mutation, using genetic variances within and between populations (LYNCH 1988a).

We thank HIDENORI TACHIDA, ANDY CLARK and MICHAEL LYNCH for useful comments on the manuscript. This investigation was supported in part by Research Grant GM 11546 from the National Institute of General Medical Sciences.

LITERATURE CITED

AVERY, P. J., 1978 The effect of finite population size on models of linked overdominant loci. Genet. Res. **31:** 239–254.

AVERY, P. J., and W. G. HILL, 1977 Variability in genetic parameters among small populations. Genet. Res. **29:** 193–213.

BULMER, M. G., 1976 The effect of selection on genetic variability: a simulation study. Genet. Res. **28:** 101–117.

BULMER, M. G., 1980 *The Mathematical Theory of Quantitative Genetics.* Oxford University Press, Oxford, U.K.

CHAKRABORTY, R., and M. NEI, 1982 Genetic differentiation of quantitative characters between populations or species. I. Mutation and random genetic drift. Genet. Res. **39:** 303–314.

COCKERHAM, C. C., 1984 Covariances of relatives for quantitative characters with drift, pp. 195–208 in *Human Population Genetics: The Pittsburgh Symposium*, edited by A. CHAKRAVARTI. Van Nostrand Reinhold, New York.

COCKERHAM, C. C., and H. TACHIDA, 1987 Evolution and maintenance of quantitative genetic variation by mutations. Proc. Natl. Acad. Sci. USA **84:** 6205–6209.

COCKERHAM, C. C., and B. S. WEIR, 1983 Variance of actual inbreeding. Theor. Popul. Biol. **23:** 85–109.

HILL, W. G. 1980 Design of quantitative genetic selection experiments, pp. 1–13 in *Selection Experiments in Laboratory and Domestic Animals*, edited by A. ROBERTSON. Common. Agric. Bur., Slough, England.

LANDE, R., 1977 Statistical tests for natural selection on quantitative characters. Evolution **31:** 442–444.

LYNCH, M., 1988a The rate of polygenic mutation. Genet. Res. **51:** 137–148.

LYNCH, M., 1988b Design and analysis of experiments on random drift and inbreeding depression. Genetics **120:** 791–807.

LYNCH, M., and W. G. HILL, 1986 Phenotypic evolution by neutral mutation. Evolution **40:** 915–935.

RAO, C. R., 1973 *Linear Statistical Inference and Its Applications*, Ed. 2. John Wiley & Sons, New York.

ROBERTSON, A., 1967 The nature of quantitative variation, pp. 265–280 in *Heritage from Mendel*, edited by R. A. BRINK and E. D. STYLES. University Wisconsin Press, Wisconsin.

SERANT, D., 1974 Linkage and inbreeding coefficients in a finite random mating population. Theor. Popul. Biol. **6:** 251–263.

TACHIDA, H., and C. C. COCKERHAM, 1989 Effects of identity disequilibrium and linkage on quantitative variation in finite populations. Genet. Res. **53:** 63–70.

TACHIDA, H., and C. C. COCKERHAM, 1990 Evolution of neutral quantitative characters with gene interaction and mutation, pp. 233–249 in *Population Biology of Genes and Molecules*, edited by M. TAKAHATA and J. F. CROW. Baifukan, Tokyo.

WEIR, B. S., P. J. AVERY, and W. G. HILL, 1980 Effect of mating structure on variation in inbreeding. Theor. Popul. Biol. **18:** 396–429.

WEIR, B. S., and C. C. COCKERHAM, 1969 Group inbreeding with two linked loci. Genetics **63:** 711–742.

WEIR, B. S., and W. G. HILL, 1980 Effect of mating structure on variation in linkage disequilibrium. Genetics **95:** 477–488.

WEIR, B. S., J. REYNOLDS, and K. G. DODDS, 1990 The variance of sample heterozygosity. Theor. Popul. Biol. **37:** 235–253.

WRIGHT, S., 1951 The genetic structure of populations. Ann. Eugenics **15:** 323–354.

WRIGHT, S., 1952 The theoretical variance within and among subdivisions of a population that is in a steady state. Genetics **37:** 312–321.

Communicating editor: A. G. CLARK

APPENDIX A: IDENTITY MEASURES

One locus

These identity by descent measures are dependent on the mutation rate, u, the effective population size, N, and the initial state of the populations. t indexes the generation number.

Within population:

$$\theta_{1_t} = \theta_{1_\infty} + A_1\lambda_1^t;\ \gamma_{1_t} = \gamma_{1_\infty} + A_2\lambda_1^t + A_2\lambda_2^t;$$

$$\delta_{1_t} = \delta_{1_\infty} + A_4\lambda_1^t + A_5\lambda_2^t + A_6\lambda_3^t;$$

$$\delta_{2_t} = \delta_{2_\infty} + A_7\lambda_1^t + A_8\lambda_2^t + A_9\lambda_3^t$$

where

$$\rho = 1 - u,\ \beta = 1/2N,\ \beta_i = \prod_{j=1}^{i}(1 - j\beta),\ \lambda_i = \rho^{i+1}\beta_i\ (i = 1, 2, 3),$$

and

$$\theta_{1_\infty} = \rho^2\beta/(1 - \lambda_1);\ \gamma_{1_\infty} = \rho^2\beta[\beta + 3\beta_1\theta_{1_\infty}]/(1 - \lambda_2);$$

$$\delta_{1_\infty} = \rho^4\beta[\beta^2 + 7\beta\beta_1\theta_{1_\infty} + 6\beta_2\gamma_{1_\infty}]/(1 - \lambda_3);$$

$$\delta_{2_\infty} = \rho^4\beta[\beta + 2(1 - \beta^2)\theta_{1_\infty} + 4\beta_2\gamma_{1_\infty}]/(1 - \lambda_3);$$

$$A_1 = \theta_{1_0} - \theta_{1_\infty};\ A_2 = 3\rho^3\beta\beta_1A_1/(\lambda_1 - \lambda_2);\ A_3 = \gamma_{1_0} - \gamma_{1_\infty} - A_2;$$

$$A_4 = \rho^4\beta[7\beta\beta_1A_1 + 6\beta_2A_2]/(\lambda_1 - \lambda_3);$$

$$A_5 = 6\rho^4\beta\beta_2A_3/(\lambda_2 - \lambda_3);\ A_6 = \delta_{1_0} - \delta_{1_\infty} - A_4 - A_5;$$

$$A_7 = 2\rho^4\beta[(1 - \beta^2)A_1 + 2\beta_2A_2]/(\lambda_1 - \lambda_3);$$

$$A_8 = 4\rho^4\beta\beta_2A_3/(\lambda_2 - \lambda_3);\ A_9 = \delta_{2_0} - \delta_{2_\infty} - A_7 - A_8.$$

Between two populations:

$$\theta'_{1_t} = B_1\omega_1^t;\ \gamma'_{1_t} = B_2\omega_1^t + B_3\omega_2^t;$$

$$\delta'_{1_t} = \theta_{1_\infty}^2 + B_4\lambda_1^t + B_5\omega_3^t;\ \delta'_{2_t} = B_6\omega_1^t + B_7\omega_2^t + B_8\omega_3^t;$$

$$\delta'_{3_t} = B_6\omega_1^t + B_7\omega_2^t + B_9\omega_3^t;$$

$$\delta'_{4_t} = B_{10}\omega_1^t + B_{11}\omega_2^t + B_{12}\omega_4^t;$$

$$\delta'_{5_t} = B_{13}\omega_1^t + B_{14}\omega_2^t + B_{15}\omega_4^t$$

where

$$\omega_1 = \rho^2,\ \omega_2 = \rho^3\beta_1,\ \omega_3 = \rho^4\beta_1^2,\ \omega_4 = \rho^4\beta_2,$$

and

$$B_1 = \theta_{1_0};\ B_2 = \rho^3\beta B_1/(\omega_1 - \omega_2);\ B_3 = \gamma_{1_0} - B_2;$$

$$B_4 = 2\rho^4\beta\beta_1A_1/(\lambda_1 - \omega_3);\ B_5 = \delta_{2_0} - \theta_{1_\infty}^2 - B_4;$$

$$B_6 = \rho^4\beta[\beta B_1 + 2\beta_1B_2]/(\omega_1 - \omega_3);\ B_7 = 2\rho^4\beta\beta_1B_3/(\omega_2 - \omega_3);$$

$$B_8 = \delta_{1_0} - B_6 - B_7;\ B_9 = \delta_{2_0} - B_6 - B_7;$$

$$B_{10} = \rho^4\beta[\beta B_1 + 3\beta_1B_2]/(\omega_1 - \omega_4);\ B_{11} = 3\rho^4\beta\beta_1B_3/(\omega_2 - \omega_4);$$

$$B_{12} = \delta_{1_0} - B_{10} - B_{11};\ B_{13} = \rho^4\beta[B_1 + 2\beta_1B_2]/(\omega_1 - \omega_4);$$

$$B_{14} = 2\rho^4\beta\beta_1B_3/(\omega_2 - \omega_4);\ B_{15} = \delta_{2_0} - B_{13} - B_{14}.$$

Among three populations:

$$\gamma''_{1_t} = C_1\nu_1^t;\ \delta''_{1_t} = C_2\omega_1^t + C_3\nu_2^t;\ \delta''_{2_t} = C_4\nu_1^t + C_5\nu_2^t;$$

$$\delta''_{3_t} = C_4\nu_1^t + C_6\nu_2^t$$

where

$$\nu_1 = \rho^3,\ \nu_2 = \rho^4\beta_1,$$

and

$$C_1 = \gamma_{1_0};\ C_2 = \rho^4\beta B_1/(\omega_1 - \nu_2);\ C_3 = \delta_{2_0} - C_2;$$

$$C_4 = \rho^4\beta C_1/(\nu_1 - \nu_2);\ C_5 = \delta_{1_0} - C_4;\ C_6 = \delta_{2_0} - C_4.$$

Among four populations:

$$\delta'''_{1_t} = \delta_{2_0}\rho^{4t};\ \delta'''_{2_t} = \delta_{1_0}\rho^{4t}.$$

Two loci

Two locus descent measures $\tilde{\theta}_1$, $\tilde{\gamma}_1$ and $\tilde{\delta}_1$ are discussed and approximated in APPENDIX B. Others are given here, which depend on the recombination rate, c, between a pair of loci i and j.

$$\tilde{\theta}'_{1_t} = E_1\phi_1^t - 2cE_2\phi_3^t + c^2E_3\phi_5^t;$$

$$\tilde{\gamma}'_{1_t} = E_1\phi_1^t + (\beta - c)E_2\phi_3^t + c\beta E_3\phi_5^t;$$

$$\tilde{\gamma}'_{2_t} = E_4\omega_{1_t}^t + E_5\phi_2^t + E_6\phi_4^t;$$

$$\tilde{\delta}'_{1_t} = \theta_{1_{i\infty}}\theta_{1_{j\infty}} + E_{7ij}\lambda_{1_i}^t + E_{7ji}\lambda_{1_j}^t + E_{8ij}\lambda_{1_i}^t\lambda_{1_j}^t;$$

$$\tilde{\delta}'_{2_t} = E_1\phi_1^t + 2\beta E_2\delta_3^t - \beta^2E_3\phi_5^t;$$

$$\tilde{\delta}'_{3_t} = E_4\omega_{1_i}^t + E_5\phi_2^t + E_9\phi_4^t;$$

$$\tilde{\gamma}''_{1_t} = E_{10}\phi_1^t + E_{11}\phi_3^t;\ \tilde{\delta}''_{1_t} = E_{10}\phi_1^t + E_{12}\phi_3^t;$$

$$\tilde{\delta}''_{2_t} = E_4\omega_{1_i}^t + E_{13}\phi_2^t;$$

$$\tilde{\delta}'''_{1_t} = \tilde{\delta}_{1_0}\phi_1^t$$

where

$$\phi_1 = \rho_i^2\rho_j^2,\ \phi_2 = \rho_i^2\rho_j^2(1 - \beta),\ \phi_3 = \rho_i^2\rho_j^2(1 - \beta - c),$$

$$\phi_4 = \rho_i^2\rho_j^2(1 - \beta - c)(1 - 2\beta),\ \phi_5 = \rho_i^2\rho_j^2(1 - \beta - c)^2,$$

and

$$E_1 = (\beta^2\tilde{\theta}_{1_0} + 2c\beta\tilde{\gamma}_{1_0} + c^2\tilde{\delta}_{1_0})/(c + \beta)^2;$$

$$E_2 = -(\beta\tilde{\theta}_{1_0} + (c - \beta)\tilde{\gamma}_{1_0} - c\tilde{\delta}_{1_0})/(c + \beta)^2;$$

$$E_3 = -(\tilde{\theta}_{1_0} - 2\tilde{\gamma}_{1_0} + \tilde{\delta}_{1_0})/(c + \beta)^2;$$

$$E_4 = \rho_i^2\rho_j^2\beta\theta_{1_{i0}}/(\omega_{1_i} - \phi_2);$$

$$E_5 = [(\beta_1 - \beta_2)\tilde{\gamma}_{1_0} + c(1 - 2\beta)\tilde{\delta}_{1_0}]/[\beta_1 - \beta_2 + c(1 - 2\beta)] - E_4;$$

$$E_6 = c(1 - 2\beta)(\tilde{\gamma}_{1_0} - \tilde{\delta}_{1_0})/[\beta_1 - \beta_2 + c(1 - 2\beta)];$$

$$E_{7ij} = \rho_i^2\rho_j^2\beta(1 - \beta)(\theta_{1_{i0}} - \theta_{1_{i\infty}})/[\lambda_{1i}(1 - \lambda_{1j})];$$

$$E_{8ij} = \tilde{\delta}_{1_0} - E_{7ij} - E_{7ji} - \theta_{1_{i\infty}}\theta_{1_{j\infty}};$$

$$E_9 = -(\beta_1 - \beta_2)(\tilde{\gamma}_{1_0} - \tilde{\delta}_{1_0})/[\beta_1 - \beta_2 + c(1 - 2\beta)];$$

$$E_{10} = (\beta\tilde{\gamma}_{1_0} + c\tilde{\delta}_{1_0})/(\beta + c);\ E_{11} = c(\tilde{\gamma}_{1_0} - \tilde{\delta}_{1_0})/(\beta + c);$$

$$E_{12} = -\beta(\tilde{\gamma}_{1_0} - \tilde{\delta}_{1_0})/(\beta + c);\ E_{13} = \tilde{\delta}_{1_0} - E_4.$$

APPENDIX B

An Approximation of variance of linkage disequilibrium: Here we approximate the variance of linkage disequilibrium $\eta_d = \tilde{\theta}_1 - 2\tilde{\gamma}_1 + \tilde{\delta}_1$. For monoecious mating population, the transition equations of identity disequilibrium measures

$$\eta_1 = \tilde{\theta}_1 - \theta_{1_i}\theta_{1_j},\ \eta_2 = \tilde{\gamma}_1 - \theta_{1_i}\theta_{1_j},\ \eta_3 = \tilde{\delta}_1 - \theta_{1_i}\theta_{1_j}$$

are (SERANT 1974)

$$\boldsymbol{\Omega}_{t+1} = \rho_i^2\rho_j^2[\boldsymbol{A}\boldsymbol{\Omega}_t + \boldsymbol{B}_t]$$

where

$$\Omega = \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{bmatrix}; \quad \boldsymbol{B}_t = (1-\theta_{1_{it}})(1-\theta_{1_{jt}})\beta((1-c)^2 + c^2 - \beta)\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

$$\boldsymbol{A} = \begin{bmatrix} (1-c)^2 - \beta + 2\beta c & 2c(1-c)(1-2\beta) & c^2(1-2\beta) \\ \beta(1-c-\beta+2\beta c) & (1-2\beta)(1-c-\beta+4\beta c) & c(1-2\beta)(1-3\beta) \\ 2\beta^2(1-\beta) & 4\beta(1-\beta)(1-2\beta) & (1-\beta)(1-2\beta)(1-3\beta) \end{bmatrix}.$$

SERANT found that the eigenvalues of matrix $\boldsymbol{A}$ are

$$e_1 = (1-2\beta) + O(\beta^2),$$
$$e_2 = 1 - c - \beta(5-2c) + O(\beta^2),$$
$$e_3 = (1-c)^2 - \beta(3-4c) + O(\beta^2).$$

This suggests that for large N we can approximate the eigenvalues as $e_1 \simeq (1-2\beta)$, $e_2 \simeq (1-2\beta)(1-c)$, and $e_3 \simeq (1-2\beta)(1-c)^2$ or equivalently the matrix $\boldsymbol{A}$ as

$$\boldsymbol{A} \simeq (1-2\beta)\begin{bmatrix} (1-c)^2 & 2c(1-c) & c^2 \\ 0 & (1-c) & c \\ 0 & 0 & 1 \end{bmatrix}$$

for a simple solution. With this we obtain the approximation

$$\eta_{d_t} \simeq \phi_6^t \eta_{d_0} + \frac{(1-c)^2 + c^2}{2Nc(2-c)}[(1-\theta_{1_{it}})(1-\theta_{1_{jt}}) - \phi_6^t(1-\theta_{1_{i0}})(1-\theta_{1_{j0}})]$$

where $\phi_6 = \rho_i^2\rho_j^2(1-2\beta)(1-c)^2$. This result agrees with the approximations of SERANT (1974), AVERY (1978) and WEIR and HILL (1980) at equilibrium. Ignoring recurrent mutation and assuming $\eta_{d_0} = 0$, this can be further approximated as

$$\eta_{d_t} \simeq \frac{(1-c)^2 + c^2}{2Nc(2-c)}[1 - (1-c)^{2t}](1-\theta_{1_{it}})(1-\theta_{1_{jt}}).$$

The accuracy of the approximation depends very much on N and c. When $N = 20$, the approximation is very good for $c = 0.5$ all time and is generally satisfactory for t up to one or two N generations. As c decreases, the approximation breaks down more quickly.

APPENDIX C

Analysis of sampling effects: With sampling the expectation of variance of genetic variance within populations contains three level expectations: $\mathscr{E}_x$ with respect to the x's, $\mathscr{E}_w$ with respect to the sampling within populations and $\mathscr{E}_b$ over all replicate populations. After taking the expectation $\mathscr{E}_x$, we take the expectation $\mathscr{E}_w$ on the sampling gene and genotypic frequencies. This is done based on the assumption of multinomial sampling of individuals from a replicate population. For example

$$\mathscr{E}_w(\hat{P}_{ij}^{ij}) = \mathring{P}_{ij}^{ij}, \quad \mathscr{E}_w(\hat{P}_{i\cdot}^{i\cdot}) = \mathring{P}_{i\cdot}^{i\cdot}, \quad \mathscr{E}_w(\hat{p}_i) = \mathring{p}_i$$

and

$$\mathrm{Var}_w(\hat{P}_{ij}^{ij}) = \mathring{P}_{ij}^{ij}(1 - \mathring{P}_{ij}^{ij})/n,$$
$$\mathrm{Var}_w(\hat{P}_{i\cdot}^{i\cdot}) = \mathring{P}_{i\cdot}^{i\cdot}(1 - \mathring{P}_{i\cdot}^{i\cdot})/n, \quad \mathrm{Var}_w(\hat{p}_i) = \mathring{p}_i(1-\mathring{p}_i)/2n$$

since linear combinations of multinomial variables are still multinomial. For functions like $\hat{P}_{i\cdot}^{i\cdot}\hat{p}_i^2$, however, we need to use the definition (15) to analyze the expectation of

$$\hat{P}_{i\cdot}^{i\cdot}\left(\hat{P}_{i\cdot}^{i\cdot} + \tfrac{1}{2}\sum_{i<j}\hat{P}_{j\cdot}^{j\cdot}\right)^2$$

based on multinomial theory. For functions involving gene frequencies from two loci, we need to use

$$\hat{p}_i = \hat{P}_{ik}^{ik} + \sum_{k<l}\hat{P}_{il}^{ik} + \sum_{l,m\neq k}\hat{P}_{il}^{im} + \tfrac{1}{2}\sum_{i<j}\left(\hat{P}_{jk}^{ik} + \sum_{k<l}\left(\hat{P}_{jl}^{ik} + \hat{P}_{jk}^{il}\right) + \sum_{l,m\neq k}\hat{P}_{jl}^{im}\right)$$

$$\hat{q}_k = \hat{P}_{ik}^{ik} + \sum_{i<j}\hat{P}_{jk}^{ik} + \sum_{j,n\neq i}\hat{P}_{jk}^{nk} + \tfrac{1}{2}\sum_{k<l}\left(\hat{P}_{il}^{ik} + \sum_{i<j}\left(\hat{P}_{jl}^{ik} + \hat{P}_{jk}^{il}\right) + \sum_{j,n\neq i}\hat{P}_{jl}^{nk}\right)$$

and

$$\hat{P}_{i\cdot}^{i\cdot} = \hat{P}_{ik}^{ik} + \sum_{k<l}\hat{P}_{il}^{ik} + \sum_{l,m\neq k}\hat{P}_{il}^{im}.$$

After taking the expectation $\mathscr{E}_w$, we then take the expectation $\mathscr{E}_b$ over all replicate populations and express the results in terms of identity by descent measures for the infinite allele mutation model. We utilize the fact that the populations are expected to be at the Hardy-Weinberg and linkage equilibria by the assumption, although samples from the populations are not. Thus, for example, in taking the expectation of $\hat{Z}_{r2}\hat{Z}_{r4}$, we perform the following analysis

$$\begin{aligned}
\mathscr{E}\hat{Z}_{r2}\hat{Z}_{r4} &= \mathscr{E}_b\mathscr{E}_w\mathscr{E}_x\sum_i\hat{P}_{i\cdot}^{i\cdot}x_i^2\sum_j\hat{p}_j^2x_j^2 \\
&= \mathscr{E}_b\mathscr{E}_w\left\{\sum_i\hat{P}_{i\cdot}^{i\cdot}\hat{p}_i^2\mu_4 + \sum_{i\neq j}\sum\hat{P}_{i\cdot}^{i\cdot}\hat{p}_j^2\sigma_x^4\right\} \\
&= \mathscr{E}_b\mathscr{E}_w\left\{\sum_i\hat{P}_{i\cdot}^{i\cdot}\left(\hat{P}_{i\cdot}^{i\cdot} + \tfrac{1}{2}\sum_{i<j}\hat{P}_{j\cdot}^{i\cdot}\right)^2\mu_4\right. \\
&\quad \left. + \sum_{i\neq j}\sum\hat{P}_{i\cdot}^{i\cdot}\left(\hat{P}_{j\cdot}^{j\cdot} + \tfrac{1}{2}\sum_{j<k}\hat{P}_{k\cdot}^{j\cdot}\right)^2\sigma_x^4\right\} \\
&= \mathscr{E}_b\mathscr{E}_w\left\{\left(\sum_i\hat{P}_{i\cdot}^{i\cdot 3} + \sum_{i<j}\sum\hat{P}_{i\cdot}^{i\cdot 2}\hat{P}_{j\cdot}^{i\cdot} + \tfrac{1}{4}\sum_{i<j}\sum\hat{P}_{i\cdot}^{i\cdot}\hat{P}_{j\cdot}^{i\cdot 2}\right.\right. \\
&\quad \left. + \tfrac{1}{4}\sum_{i<j,i<k,k\neq j}\sum\sum\hat{P}_{i\cdot}^{i\cdot}\hat{P}_{j\cdot}^{i\cdot}\hat{P}_{k\cdot}^{i\cdot}\right)\mu_4 \\
&\quad + \left(\sum_{i\neq j}\sum\hat{P}_{i\cdot}^{i\cdot}\hat{P}_{j\cdot}^{j\cdot 2} + \sum_{i\neq j,j<k}\sum\sum\hat{P}_{i\cdot}^{i\cdot}\hat{P}_{j\cdot}^{j\cdot}\hat{P}_{k\cdot}^{j\cdot} + \tfrac{1}{4}\sum_{i\neq j,j<k}\sum\sum\hat{P}_{i\cdot}^{i\cdot}\hat{P}_{k\cdot}^{j\cdot 2}\right. \\
&\quad \left.\left. + \tfrac{1}{4}\sum_{i\neq j,j<k,j<l,l\neq k}\sum\sum\sum\hat{P}_{i\cdot}^{i\cdot}\hat{P}_{k\cdot}^{j\cdot}\hat{P}_{l\cdot}^{j\cdot}\right)\sigma_x^4\right\} \\
&= \mathscr{E}_b\left\{\sum_i\left(\frac{1}{n^2}\mathring{P}_{i\cdot}^{i\cdot} + \frac{5}{2n}\left(1-\frac{1}{n}\right)\mathring{P}_{i\cdot}^{i\cdot}\right)\left(\hat{P}_{i\cdot}^{i\cdot} + \tfrac{1}{2}\sum_{i<j}\hat{P}_{j\cdot}^{i\cdot}\right)\right. \\
&\quad \left. + \frac{1}{2n}\left(1-\frac{1}{n}\right)\mathring{P}_{i\cdot}^{i\cdot 2}\right.
\end{aligned}$$

$$+\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\hat{p}_{i:}^{i}\left(\hat{P}_{i:}^{i}+\tfrac{1}{2}\sum_{i<j}\hat{P}_{j:}^{i}\right)^{2}\Big)\mu_4$$

$$+\sum\sum_{i\neq j}\Big(\frac{1}{2n}\left(1-\frac{1}{n}\right)\hat{P}_{i:}^{i}\hat{P}_{j:}^{j}$$

$$+\frac{1}{2n}\left(1-\frac{1}{n}\right)\hat{P}_{i:}^{i}\left(\hat{P}_{j:}^{j}+\tfrac{1}{2}\sum_{j<k}\hat{P}_{k:}^{j}\right)+\left(1-\frac{1}{n}\right)$$

$$\cdot\left(1-\frac{2}{n}\right)\hat{P}_{i:}^{i}\left(\hat{P}_{j:}^{j}+\tfrac{1}{2}\sum_{j<k}\hat{P}_{k:}^{j}\right)^{2}\Big)\sigma_x^4\Big\}$$

$$=\mathscr{E}_b\Big\{\sum_i\Big(\frac{1}{n^2}\hat{p}_i^2+\frac{5}{2n}\left(1-\frac{1}{n}\right)\hat{p}_i^3+\frac{1}{2n}\left(1-\frac{1}{n}\right)\hat{p}_i^4$$

$$+\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\hat{p}_i^4\Big)\mu_4$$

$$+\sum\sum_{i\neq j}\Big(\frac{1}{2n}\left(1-\frac{1}{n}\right)\hat{p}_i^2\hat{p}_j^2+\frac{1}{2n}\left(1-\frac{1}{n}\right)\hat{p}_i^2\hat{p}_j$$

$$+\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\hat{p}_i^2\hat{p}_j^2\Big)\sigma_x^4\Big\}$$

$$=\left(\frac{1}{n^2}\theta_1+\frac{5}{2n}\left(1-\frac{1}{n}\right)\gamma_1+\left(1-\frac{1}{n}\right)\left(1-\frac{3}{2n}\right)\delta_1\right)\mu_4$$

$$+\left(\frac{1}{2n}\left(1-\frac{1}{n}\right)(\theta_1-\gamma_1)+\left(1-\frac{1}{n}\right)\left(1-\frac{3}{2n}\right)(\delta_2-\delta_1)\right)\sigma_x^4.$$

Here we list the expectations of all the gene and genotype frequency functions required for analysis of sampling variance of genetic variance within populations.

One locus:

$$\mathscr{E}_b\mathscr{E}_w\sum_i\hat{p}_i^2=(1-\alpha)\theta_1+\alpha;$$

$$\mathscr{E}_b\mathscr{E}_w\sum_i\hat{p}_i\hat{P}_{i:}^{i}=(1-2\alpha)\gamma_1+2\alpha\theta_1;$$

$$\mathscr{E}_b\mathscr{E}_w\sum_i\hat{p}_i^3=(1-\alpha)(1-2\alpha)\gamma_1+3\alpha(1-\alpha)\theta_1+\alpha^2;$$

$$\mathscr{E}_b\mathscr{E}_w\sum_i\hat{P}_{i:}^{i:2}=(1-2\alpha)\delta_1+2\alpha\theta_1;$$

$$\mathscr{E}_b\mathscr{E}_w\sum_i\hat{p}_i^2\hat{P}_{i:}^{i}=(1-2\alpha)(1-3\alpha)\delta_1+5\alpha(1-2\alpha)\gamma_1+4\alpha^2\theta_1;$$

$$\mathscr{E}_b\mathscr{E}_w\sum_i\hat{p}_i^4=(1-\alpha)(1-2\alpha)(1-3\alpha)\delta_1$$
$$+6\alpha(1-\alpha)(1-2\alpha)\gamma_1+7\alpha^2(1-\alpha)\theta_1+\alpha^3;$$

$$\mathscr{E}_b\mathscr{E}_w\sum_{i\neq j}\hat{p}_i\hat{p}_j=(1-\alpha)(1-\theta_1);$$

$$\mathscr{E}_b\mathscr{E}_w\sum_{i\neq j}\hat{p}_i\hat{P}_{j:}^{j}=(1-2\alpha)(\theta_1-\gamma_1);$$

$$\mathscr{E}_b\mathscr{E}_w\sum_{i\neq j}\hat{p}_i\hat{p}_j^2=(1-\alpha)(1-2\alpha)(\theta_1-\gamma_1)+\alpha(1-\alpha)(1-\theta_1);$$

$$\mathscr{E}_b\mathscr{E}_w\sum_{i\neq j}\hat{P}_{i:}^{i}\hat{P}_{j:}^{j}=(1-2\alpha)(\delta_2-\delta_1);$$

$$\mathscr{E}_b\mathscr{E}_w\sum_{i<j}\hat{P}_{j:}^{i:2}=2(1-2\alpha)(\delta_2-\delta_1)+2\alpha(1-\theta_1);$$

$$\mathscr{E}_b\mathscr{E}_w\sum_{i\neq j}\hat{p}_i^2\hat{P}_{j:}^{j}=(1-2\alpha)(1-3\alpha)(\delta_2-\delta_1)+\alpha(1-2\alpha)(\theta_1-\gamma_1);$$

$$\mathscr{E}_b\mathscr{E}_w\sum_{i<j}\hat{p}_i\hat{p}_j\hat{P}_{j:}^{i}=(1-2\alpha)(1-3\alpha)(\delta_2-\delta_1)$$
$$+2\alpha(1-2\alpha)(\theta_1-\gamma_1)+\alpha^2(1-\theta_1);$$

$$\mathscr{E}_b\mathscr{E}_w\sum_{i\neq j}\hat{p}_i^2\hat{p}_j^2=(1-\alpha)(1-2\alpha)(1-3\alpha)(\delta_2-\delta_1)$$
$$+2\alpha(1-\alpha)(1-2\alpha)(\theta_1-\gamma_1)+\alpha^2(1-\alpha)(1-\theta_1);$$

$$\mathscr{E}_b\mathscr{E}_w\sum_i\hat{P}_{i:}^{i}\tilde{P}_{i:}^{i}=\delta_1;$$

$$\mathscr{E}_b\mathscr{E}_w\sum_i\hat{p}_i\tilde{P}_{i:}^{i}=\gamma_1;$$

$$\mathscr{E}_b\mathscr{E}_w\sum_i\hat{p}_i\tilde{p}_i=\theta_1;$$

$$\mathscr{E}_b\mathscr{E}_w\sum_i\hat{p}_i^2\tilde{P}_{i:}^{i}=(1-\alpha)\delta_1+\alpha\gamma_1;$$

$$\mathscr{E}_b\mathscr{E}_w\sum_i\hat{p}_i\tilde{p}_i^2=(1-\alpha)\gamma_1+\alpha\theta_1;$$

$$\mathscr{E}_b\mathscr{E}_w\sum_i\hat{p}_i^2\tilde{p}_i^2=(1-\alpha)^2\delta_1+2\alpha(1-\alpha)\gamma_1+\alpha^2\theta_1;$$

$$\mathscr{E}_b\mathscr{E}_w\sum_i\hat{p}_i^3\tilde{p}_i=(1-\alpha)(1-2\alpha)\delta_1+3\alpha(1-\alpha)\gamma_1+\alpha^2\theta_1;$$

$$\mathscr{E}_b\mathscr{E}_w\sum_{i\neq j}\hat{P}_{i:}^{i}\tilde{P}_{j:}^{j}=\delta_2-\delta_1;\quad\mathscr{E}_b\mathscr{E}_w\sum_{i<j}\hat{P}_{j:}^{i}\tilde{P}_{j:}^{i}=2(\delta_2-\delta_1);$$

$$\mathscr{E}_b\mathscr{E}_w\sum_{i\neq j}\hat{p}_i\tilde{P}_{j:}^{j}=\theta_1-\gamma_1;$$

$$\mathscr{E}_b\mathscr{E}_w\sum_{i\neq j}\hat{p}_i\tilde{p}_j=1-\theta_1;$$

$$\mathscr{E}_b\mathscr{E}_w\sum_{i\neq j}\hat{p}_i^2\tilde{P}_{j:}^{j}=(1-\alpha)(\delta_2-\delta_1)+\alpha(\theta_1-\gamma_1);$$

$$\mathscr{E}_b\mathscr{E}_w\sum_{i<j}\hat{p}_i\hat{p}_j\tilde{P}_{j:}^{i}=(1-\alpha)(\delta_2-\delta_1);$$

$$\mathscr{E}_b\mathscr{E}_w\sum_{i\neq j}\hat{p}_i\hat{p}_j\tilde{p}_i\tilde{p}_j=(1-\alpha)^2(\delta_2-\delta_1);$$

$$\mathscr{E}_b\mathscr{E}_w\sum_{i\neq j}\hat{p}_i\tilde{p}_j^2=(1-\alpha)(\theta_1-\gamma_1)+\alpha(1-\theta_1);$$

$$\mathscr{E}_b\mathscr{E}_w\sum_{i\neq j}\hat{p}_i^2\tilde{p}_j^2=(1-\alpha)^2(\delta_2-\delta_1)+2\alpha(1-\alpha)(\theta_1-\gamma_1)$$
$$+\alpha^2(1-\theta_1);$$

$$\mathscr{E}_b\mathscr{E}_w\sum_{i\neq j}\hat{p}_i^2\hat{p}_j\tilde{p}_j=(1-\alpha)(1-2\alpha)(\delta_2-\delta_1)$$
$$+\alpha(1-\alpha)(\theta_1-\gamma_1).$$

Two Loci:

$$\mathscr{E}_b\mathscr{E}_w\sum_i\sum_k\hat{p}_i\hat{q}_k=1;$$

$$\mathscr{E}_b\mathscr{E}_w\sum_i\sum_k\hat{p}_i\hat{P}_{\cdot:k}^{k}=\theta_1;$$

$$\mathscr{E}_b\mathscr{E}_w\sum_i\sum_k\hat{p}_i\hat{q}_k^2=(1-\alpha)\theta_1+\alpha;$$

$$\mathscr{E}_b\mathscr{E}_w\sum_i\sum_k\hat{P}_{i:}^{i}\hat{P}_{\cdot:k}^{k}=(1-2\alpha)\tilde{\delta}_1+2\alpha\tilde{\theta}_1;$$

$$\mathscr{E}_b\mathscr{E}_w\sum_i\sum_k\hat{p}_i^2\hat{P}_{\cdot:k}^{k}=(1-2\alpha)(1-3\alpha)\tilde{\delta}_1$$
$$+4\alpha(1-2\alpha)\tilde{\gamma}_1+2\alpha^2\tilde{\theta}_1+\alpha\theta_1;$$

$$\mathscr{E}_b\mathscr{E}_w\sum_i\sum_k(\hat{P}_{\cdot:}^{ik}+\hat{P}_{\cdot:k}^{i})^2$$
$$=(1-2\alpha)\tilde{\delta}_1+2(1-2\alpha)\tilde{\gamma}_1+\tilde{\theta}_1+4\alpha\theta_1+2\alpha;$$

$$\mathscr{E}_b\mathscr{E}_w\sum_i\sum_k(\hat{P}_{\cdot:}^{ik}+\hat{P}_{\cdot:k}^{i})\hat{p}_i\hat{q}_k=(1-2\alpha)(1-3\alpha)\tilde{\delta}_1$$
$$+(1-4\alpha^2)\tilde{\gamma}_1+\alpha\tilde{\theta}_1+4\alpha(1-\alpha)\theta_1+2\alpha^2;$$

$$\mathscr{E}_b\mathscr{E}_w\sum_i\sum_k\hat{p}_i^2\hat{q}_k^2=(1-\alpha)(1-2\alpha)(1-3\alpha)\tilde{\delta}_1$$
$$+4\alpha(1-\alpha)(1-2\alpha)\tilde{\gamma}_1$$
$$+2\alpha^2(1-\alpha)\tilde{\theta}_1+2\alpha(1-\alpha)\theta_1+\alpha^2;$$

$$\mathscr{E}_b\mathscr{E}_w\sum_i\sum_k \hat{p}_i\tilde{q}_k = 1; \quad \mathscr{E}_b\mathscr{E}_w\sum_i\sum_k \hat{p}_i\tilde{P}^{\cdot k}_{\cdot\cdot} = \theta_1;$$

$$\mathscr{E}_b\mathscr{E}_w\sum_i\sum_k \hat{p}_i\tilde{q}_k^2 = (1-\alpha)\theta_1 + \alpha;$$

$$\mathscr{E}_b\mathscr{E}_w\sum_i\sum_k \hat{P}^{i\cdot}_{\cdot\cdot}\tilde{P}^{\cdot k}_{\cdot\cdot} = \tilde{\delta}_1;$$

$$\mathscr{E}_b\mathscr{E}_w\sum_i\sum_k \hat{p}_i^2\tilde{P}^{\cdot k}_{\cdot\cdot} = (1-\alpha)\tilde{\delta}_1 + \alpha\theta_1;$$

$$\mathscr{E}_b\mathscr{E}_w\sum_i\sum_k (\hat{P}^{ik}_{\cdot\cdot} + \hat{P}^{i\cdot}_{\cdot k})(\tilde{P}^{ik}_{\cdot\cdot} + \tilde{P}^{i\cdot}_{\cdot k}) = \tilde{\delta}_1 + 2\tilde{\gamma}_1 + \tilde{\theta}_1;$$

$$\mathscr{E}_b\mathscr{E}_w\sum_i\sum_k (\hat{P}^{ik}_{\cdot\cdot} + \hat{P}^{i\cdot}_{\cdot k})\hat{p}_i\tilde{q}_k = (1-\alpha)\tilde{\delta}_1 + \tilde{\gamma}_1 + \alpha\tilde{\theta}_1;$$

$$\mathscr{E}_b\mathscr{E}_w\sum_i\sum_k \hat{p}_i\hat{p}_i\tilde{q}_k\tilde{q}_k = (1-\alpha)^2\tilde{\delta}_1 + 2\alpha(1-\alpha)\tilde{\gamma}_1 + \alpha^2\tilde{\theta}_1;$$

$$\mathscr{E}_b\mathscr{E}_w\sum_i\sum_k \hat{p}_i^2\tilde{q}_k^2 = (1-\alpha)^2\tilde{\delta}_1 + 2\alpha(1-\alpha)\theta_1 + \alpha^2;$$

$$\mathscr{E}_b\mathscr{E}_w\sum_i\sum_k \hat{p}_i^2\tilde{q}_k\tilde{q}_k = (1-\alpha)(1-2\alpha)\tilde{\delta}_1 + 2\alpha(1-\alpha)\tilde{\gamma}_1 + \alpha\theta_1$$

where $\alpha = 1/2n$, and n is the sample size in a replicate population.