Zygotic Associations and Multilocus Statistics in a Nonequilibrium Diploid Population

Rong-Cai Yang

Research Division, Alberta Agriculture, Food and Rural Development, Edmonton, Alberta T6H 5T6, Canada and Department of Renewable Resources, University of Alberta, Edmonton, Alberta T6G 2H1, Canada

Manuscript received July 14, 1999
Accepted for publication April 3, 2000

ABSTRACT

The usual approach to characterizing and estimating multilocus associations in a diploid population assumes that the population is in Hardy-Weinberg equilibrium. The purpose of this study is to develop a set of summary statistics that can be used to characterize and estimate the multilocus associations in a nonequilibrium population. The concept of "zygotic associations" is first expanded to facilitate the development. The summary statistics are calculated using the distribution of a random variable, the number of heterozygous loci (K) found in diploid individuals in the population. In particular, the variance of K consists of single-locus and multilocus components with the latter being the sum of zygotic associations between pairs of loci. Simulation results show that the multilocus associations in the variance of K are detectable in a sample of moderate size ($\geq$30) when the sum of all pairwise zygotic associations is greater than zero and when gene frequency is intermediate. The method presented here is a generalization of the well-known development for the Hardy-Weinberg equilibrium population and thus may be of more general use in elucidating the multilocus organizations in nonequilibrium and equilibrium populations.

THE extent and patterns of nonrandom associations between linked as well as independent loci provide important information about the history of a population, the evolutionary forces governing these loci, and the location of the loci on the chromosomes. Such multilocus associations may arise from many demographic and evolutionary events including epistatic selection, random drift due to population growth and decline, mixing of two or more distinct gene pools, nonrandom mating, and mutation, regardless of whether or not the loci are physically linked (*e.g.*, Hedrick *et al.* 1978; Brown 1979; Barton and Clark 1990).

A number of statistical measures have been proposed to characterize the multilocus associations, but the literature has focused on characterizing gametic disequilibria, *i.e.*, nonrandom associations of alleles at two loci ordered within gametes (*e.g.*, Hedrick 1987). While these measures are useful for analyzing haploid data or diploid data from a Hardy-Weinberg equilibrium population, they may not be appropriate for a nonequilibrium diploid population in which a complete characterization of two-locus associations also requires other types of disequilibria (Cockerham and Weir 1973; Weir 1979). For example, in a hybrid population arising from mixing of genes from two or more populations or species, alleles derived from the same populations or species tend to cluster together in the same individuals, either because of Wahlund's (1928) effect or because of strong selection against hybrids or both. The resulting Hardy-Weinberg disequilibria at individual loci and multilocus associations across loci may be persistent and may be detectable for a number of generations after an initial mixing of gene pools. Thus, the multilocus associations in the hybrid population need to be characterized at the zygote level.

A related issue about characterizing and testing the multilocus associations is that most of the proposed measures are defined for a pair of loci only. When there are a large number of loci, each having many alleles, pairwise measures may be too many to be readily manageable and interpretable. For example, for 20 loci, each with four alleles in a nonequilibrium population, there are 6 independent Hardy-Weinberg disequilibria for each of the 20 loci, 9 gametic disequilibria, 9 nongametic disequilibria, 54 trigenic disequilibria, and 45 quadrigenic disequilibria for each of 190 locus pairs. Furthermore, unless a stringent significance level is imposed, the large number of required pairwise tests under commonly used significance levels of 5 and 1% may produce spurious association realizations (Karlin and Piazza 1981; Weir 1996, pp. 133–135). Therefore, it is desirable to have a set of summary statistics that adequately describe the extent and patterns of multilocus structure in a nonequilibrium population.

The objective of this study is to develop such a set of summary statistics. The concept of "zygotic associations"

Address for correspondence: Research Division, Alberta Agriculture, Food and Rural Development, #202, 7000 - 113 St., Edmonton, Alberta T6H 5T6, Canada. E-mail: rongcai.yang@agric.gov.ab.ca

(Haldane 1949; Bennett and Binet 1956; Allard *et al.* 1968) is first expanded to facilitate the development. The summary statistics are calculated using the distribution of a random variable, the number of heterozygous loci (K) found in diploid individuals in the population. A similar method by Brown *et al.* (1980) has been used to analyze multilocus data collected from haploid, inbred, or random mating populations (*e.g.*, Brown *et al.* 1980; Whittam *et al.* 1983; Nevo and Beiles 1989; Maynard Smith *et al.* 1993; Yeh *et al.* 1994; Haubold *et al.* 1998), but it considers only gametic disequilibrium. Numerical analyses are also carried out to depict the dependence of the zygotic associations on gene frequencies and various disequilibria and to examine the sensitivity of our method for detecting the multilocus zygotic associations.

ZYGOTIC ASSOCIATIONS

Let us consider a diploid population in which individual genotypes are known at each of m loci. Two of these m loci are indexed by j and l with alleles j_u, $u = 1, 2, \ldots, r$ and l_y, $y = 1, 2, \ldots, s$, respectively. Frequencies of genotypes at loci j and l from the union of gametes $j_u l_y$ and $j_v l_z$ are written as ${}^{jl}P_{vz}^{uy} = {}^{jl}P_{uy}^{vz}$. Weir (1979) described various marginal totals that are sums of genotypic frequencies indicated by dots for the indices summed. For example, one-locus genotypic frequencies for $j_u j_v$ and $l_y l_z$ are denoted by

$$ {}^{j}P_{v.}^{u.} = \sum_{y=1}^{s}\sum_{z=1}^{s} {}^{jl}P_{vz}^{uy} \quad \text{and} \quad {}^{l}P_{.z}^{.y} = \sum_{u=1}^{r}\sum_{v=1}^{r} {}^{j}P_{vz}^{uy}, $$

and frequencies of alleles j_u and l_y are given by

$$ {}^{j}p_u = {}^{j}P_{..}^{u.} = \sum_{v=1}^{r}\sum_{y=1}^{s}\sum_{z=1}^{s} {}^{jl}P_{vz}^{uy} \quad \text{and} \quad {}^{l}p_y = {}^{l}P_{..}^{.y} = \sum_{u=1}^{r}\sum_{v=1}^{r}\sum_{z=1}^{s} {}^{j}P_{vz}^{uy}. $$

Following Bennett and Binet (1956) and Allard *et al.* (1968), we now define a zygotic association between loci j and l as a deviation of joint frequencies of double heterozygotes from products of frequencies of heterozygotes at the two loci:

$$ {}^{jl}\omega_{vz}^{uy} = {}^{jl}P_{vz}^{uy} - {}^{j}P_{v.}^{u.}{}^{l}P_{.z}^{.y}. \tag{1} $$

The other three zygotic associations, ${}^{jl}\omega_{uz}^{uy}$, ${}^{jl}\omega_{vy}^{uy}$, and ${}^{jl}\omega_{uy}^{uy}$, can be similarly defined by substituting appropriate allele indexes in (1). It is easy to find the ranges of these zygotic associations. For example, the range of ${}^{jl}\omega_{vz}^{uy}$ is

$$ -{}^{j}P_{v.}^{u.}{}^{l}P_{.z}^{.y} \le {}^{jl}\omega_{vz}^{uy} \le \min[{}^{j}P_{v.}^{u.}(1 - {}^{l}P_{.z}^{.y}), (1 - {}^{j}P_{v.}^{u.}){}^{l}P_{.z}^{.y}]. \tag{2a} $$

This dependence of the zygotic association on the marginal frequencies at single loci suggests a need to normalize the zygotic association ${}^{jl}\omega_{vz}^{uy}$,

$$ {}^{jl}\omega_{vz}^{uy'} = \begin{cases} \dfrac{{}^{jl}\omega_{vz}^{uy}}{{}^{j}P_{v.}^{u.}{}^{l}P_{.z}^{.y}}, & {}^{jl}\omega_{vz}^{uy} < 0 \\ \dfrac{{}^{jl}\omega_{vz}^{uy}}{\min[{}^{j}P_{v.}^{u.}(1 - {}^{l}P_{.z}^{.y}), (1 - {}^{j}P_{v.}^{u.}){}^{l}P_{.z}^{.y}]}, & {}^{jl}\omega_{vz}^{uy} > 0, \end{cases} \tag{2b} $$

which is analogous to Lewontin's (1964) normalized gametic disequilibrium.

When summing over all alleles at loci j and l, we obtain an overall measure of zygotic associations (ω_{jl}) and the following relations:

$$ \begin{aligned} \omega_{jl} &= \sum_{u=1}^{r}\sum_{y=1}^{s} {}^{jl}\omega_{uy}^{uy} = \sum\sum_{u\neq v}\sum\sum_{y\neq z} {}^{jl}\omega_{vz}^{uy} \\ &= -\sum_{u=1}^{r}\sum\sum_{y\neq z} {}^{jl}\omega_{uz}^{uy} = -\sum\sum_{u\neq v}\sum_{y=1}^{s} {}^{jl}\omega_{vy}^{uy}. \end{aligned} \tag{3} $$

Thus, the sum $\Sigma_{u=1}^{r}\Sigma_{v=1}^{r}\Sigma_{y=1}^{s}\Sigma_{z=1}^{s} {}^{jl}P_{vz}^{uy} = 1$ can be expanded into four classes of genotypic frequencies: (i) frequency of being homozygous at both loci; (ii) homozygous at locus j and heterozygous at locus l; (iii) heterozygous at locus j and homozygous at locus l; and (iv) heterozygous at both loci,

$$ \begin{aligned} \sum_{u=1}^{r}\sum_{y=1}^{s} {}^{jl}P_{uy}^{uy} &= (1 - H_j)(1 - H_l) + \omega_{jl} \\ \sum_{u=1}^{r}\sum\sum_{y\neq z} {}^{jl}P_{uz}^{uy} &= (1 - H_j)H_l - \omega_{jl} \\ \sum\sum_{u\neq v}\sum_{y=1}^{s} {}^{jl}P_{vy}^{uy} &= H_j(1 - H_l) - \omega_{jl} \\ \sum\sum_{u\neq v}\sum\sum_{y\neq z} {}^{jl}P_{vz}^{uy} &= H_j H_l + \omega_{jl}, \end{aligned} \tag{4} $$

where H_j and H_l are the population heterozygosities at loci j and l,

$$ \begin{aligned} H_j &= \sum\sum_{u\neq v} {}^{j}P_{v.}^{u.} = 1 - \sum_{u=1}^{r} {}^{j}P_{u.}^{u.} = h_j - \sum_{u=1}^{r} {}^{j}D_{u.}^{u.} \\ H_l &= \sum\sum_{y\neq z} {}^{l}P_{.z}^{.y} = 1 - \sum_{y=1}^{s} {}^{l}P_{.y}^{.y} = h_l - \sum_{y=1}^{s} {}^{l}D_{.y}^{.y}, \end{aligned} \tag{5} $$

with h_j ($= 1 - \Sigma_{u=1}^{r} {}^{j}p_u^2$) and ${}^{j}D_{u.}^{u.}$ ($= -\Sigma_{v\neq u} {}^{j}D_{v.}^{u.}$), for example, being the gene diversity (or expected heterozygosity under Hardy-Weinberg equilibrium) and Hardy-Weinberg disequilibrium for allele u at locus j, respectively.

MULTILOCUS HETEROZYGOSITY

Number of heterozygous loci (K): When a diploid individual is randomly taken from the population (defined above), it can be either homozygote or heterozygote at a given locus. If all m loci are evaluated, then the random variable K is simply the number of heterozygous loci found in the randomly chosen diploid individual from the population. Thus, K is the sum of m indicator variables, $K = \Sigma_{j=1}^{m} X_j$, where X_j takes either 1 or 0, depending on whether the jth locus is heterozygous or

homozygous. The probability that this locus is heterozygous is H_j, the population heterozygosity at the jth locus, and the probability that it is homozygous is $1 - H_j$. K can take any integer value from 0 to m. If $K = 0$, then all m loci are homozygous; if, on the other hand, $K = m$, then all m loci are heterozygous.

Moments of *K*: The expected value of K is

$$E(K) = E\left(\sum_{j=1}^{m} X_j\right) = \sum_{j=1}^{m} E(X_j) = \sum_{j=1}^{m} H_j, \quad (6)$$

and the second to fourth central moments are given by, letting $x_j = X_j - E(X_j)$,

$$\sigma_K^2 = E[K - E(K)]^2 = \sum_{j=1}^{m} E(x_j^2) + 2\sum\sum_{j<l} E(x_j x_l), \quad (7a)$$

$$E[K - E(K)]^3 = \sum_{j=1}^{m} E(x_j^3) + 3\sum_{j=1}^{m}\sum_{l\neq j} E(x_j^2 x_l) + 6\sum\sum\sum_{j<l<o} E(x_j x_l x_o), \quad (7b)$$

and

$$E[K - E(K)]^4 = \sum_{j=1}^{m} E(x_j^4) + 4\sum_{j=1}^{m}\sum_{l\neq j} E(x_j^3 x_l) + 6\sum\sum_{j<l} E(x_j^2 x_l^2) + 12\sum_{j=1}^{m}\sum\sum_{\substack{l,o\neq j \\ l<o}} E(x_j^2 x_l x_o) + 24\sum\sum\sum\sum_{j<l<o<q} E(x_j x_l x_o x_q), \quad (7c)$$

where, for example, $E(x_j^2 x_l)$ is the {21}th *central mixed moment* of variables X_j and X_l for loci j and l (Elandt-Johnson 1971, pp. 106–107). It is evident from (7a)–(7c) that evaluating the ith central moment of K requires a specification of joint genotypic frequencies for i loci, which include various associations for genes at up to i loci. For example, the variance (second central moment) of K is a function of single-locus heterozygosities and two-locus associations only and is independent of higher-order associations involving three or more loci. Similar arguments can be carried out for the third or higher central moments of K. If there is complete interlocus independence, (7a)–(7c) reduce to (3)–(5) of Brown *et al.* (1980) but we use heterozygosity $\{H_j\}$ instead of gene diversity $\{h_j\}$ to measure genetic variation at individual loci. When the population is in Hardy-Weinberg equilibrium, the heterozygosity equals to the gene diversity (*cf.* Equation 5).

Variance of *K*: The variance of K as given in (7a) has two components, one being the sum of variances at individual loci and the other being the sum of covariances between pairs of loci,

$$\sigma_K^2 = E[K - E(K)]^2 = \sum_{j=1}^{m} \mathrm{Var}(X_j) + 2\sum\sum_{j<l} \mathrm{Cov}(X_j, X_l), \quad (8)$$

where $\mathrm{Var}(X_j) = H_j - H_j^2$ and $\mathrm{Cov}(X_j, X_l) = \omega_{jl}$ as computed using the joint probability distribution between loci j and l (Table 1). Thus,

$$\sigma_K^2 = \sum_{j=1}^{m} H_j - \sum_{j=1}^{m} H_j^2 + 2\sum\sum_{j<l} \omega_{jl}. \quad (9a)$$

It is evident from (1) and (3) that $\omega_{jl} = \Sigma_{u=1}^{r}\Sigma_{y=1}^{s}[{}^{jl}P_{uy}^{uy} - {}^{j}P_{u.}^{u.\,l}P_{.y}^{.y}]$, for example. Following Cockerham and Weir (1973) and Weir (1979), the two-locus frequencies $\{{}^{jl}P_{uy}^{uy}\}$ are expressed in terms of gene frequencies and various genic disequilibria. Given these results and those in (5) for $\{H_j\}$, σ_K^2 in (9a) can be rewritten as

$$\sigma_K^2 = \sum_{j=1}^{m}\left[1 - \sum_{u=1}^{r}({}^{j}p_u^2 + {}^{j}D_{u.}^{u.})\right] - \sum_{j=1}^{m}\left[1 - \sum_{u=1}^{r}({}^{j}p_u^2 + {}^{j}D_{u.}^{u.})\right]^2 + 2\sum\sum_{j<l}\sum_{u=1}^{r}\sum_{y=1}^{s}[2{}^{j}p_u{}^{jl}D_{.y}^{uy} + 2{}^{l}p_y{}^{jl}D_{u.}^{uy} + {}^{jl}D_{uy}^{uy} + 2{}^{j}p_u{}^{l}p_y{}^{jl}D_{..}^{uy} + 2{}^{j}p_u{}^{l}p_y{}^{jl}D_{u.}^{.y} + ({}^{jl}D_{..}^{uy})^2 + ({}^{jl}D_{.y}^{u.})^2], \quad (9b)$$

where each genic disequilibrium (D) is the deviation of a frequency from that based on random association of genes and accounting for any lower-order disequilibria. Definitions and properties of these disequilibria are detailed in many places (*e.g.*, Weir 1979). Here it suffices to recognize that there are five types of disequilibria: (i) single-locus digenic disequilibria (*i.e.*, Hardy-Weinberg disequilibria, ${}^{j}D_{u.}^{u.}$ and ${}^{l}D_{.y}^{.y}$); (ii) two-locus digenic disequilibria for gametic genes (*i.e.*, gametic disequilibria, ${}^{jl}D_{..}^{uy}$); (iii) two-locus digenic disequilibria for nongametic genes (*i.e.*, nongametic disequilibria, ${}^{jl}D_{.y}^{u.}$); (iv) trigenic disequilibria (${}^{jl}D_{u.}^{uy}$ and ${}^{jl}D_{.y}^{uy}$); and (v) quadrigenic disequilibria (${}^{jl}D_{uy}^{uy}$).

Table 2 lists six special cases of σ_K^2 as given in (9a) or (9b). The first two cases assume that there are no zygotic associations between pairs of loci for all m loci ($\Sigma_{j<l}\,\omega_{jl} = 0$), but case 1 further assumes Hardy-Weinberg equilibrium in the population. When genotypes (zygotes) result from random union of gametes, all nongametic disequilibria including Hardy-Weinberg disequilibria at all loci disappear (*e.g.*, ${}^{j}D_{u.}^{u.} = {}^{jl}D_{.y}^{u.} = {}^{jl}D_{.y}^{uy} = {}^{jl}D_{uy}^{uy} = 0$). This leads to $\sigma_K^2(3)$ as given in case 3. $\sigma_K^2(3)$ was previously derived (*cf.* Equation 15 of Brown *et al.* 1980). Case 4 states a well-established fact that nonzero quadrigenic disequilibria occur under Hardy-Weinberg disequilibrium, even in a population that is in gametic equilibrium (*e.g.*, Haldane 1949; Bennett and Binet 1956; Weir and Cockerham 1973).

The last two cases in Table 2 are not directly obtainable from (9a) or (9b), but rather serve to illustrate the difficulty of finding the maximum value of σ_K^2 because the upper bound for ${}^{jl}\omega_{vz}^{uy}$ in (2a) is not unique. Case 5 portrays a scenario where all m loci are absolutely associated (Clegg *et al.* 1976). The final case constructs a population of hypothetical multilocus zygotes with maximum variance of heterozygosity by ranking the $\{H_j\}$ such that $H_1 > H_2 > H_3 > \ldots > H_m$. Similar expressions for these two cases were given by Brown *et al.* (1980) and Brown and Burdon (1983) for haploid and random mating populations.

NUMERICAL ANALYSIS

Relationships between zygotic associations and genic disequilibria: It is evident from (9a) and (9b) that the

TABLE 1

Joint frequency distribution of indicator variables X_j and X_l in terms of heterozygosities (H_j and H_l) and zygotic associations (ω_{jl}) at loci j and l

	X_l		
X_j	0	1	Total
0	$f_{00} = (1 - H_j)(1 - H_l) + \omega_{jl}$	$f_{01} = (1 - H_j)H_l - \omega_{jl}$	$f_{0.} = 1 - H_j$
1	$f_{10} = H_j(1 - H_l) - \omega_{jl}$	$f_{11} = H_jH_l + \omega_{jl}$	$f_{1.} = H_j$
Total	$f_{.0} = 1 - H_l$	$f_{.1} = H_l$	1.0

The overall measure of zygotic associations (ω_{jl}) can be expressed in one of five ways: $\omega_{jl} = f_{00}\,f_{11} - f_{01}\,f_{10} = f_{00} - f_{0.}\,f_{.0} = -(f_{01} - f_{0.}\,f_{.1}) = -(f_{10} - f_{1.}f_{.0}) = f_{11} - f_{1.}\,f_{.1}$.

overall measure of zygotic associations between a pair of loci is a complex function of gametic, nongametic, trigenic, and quadrigenic disequilibria weighted appropriately by gene frequencies. The range of values for each of these disequilibria is defined by gene frequencies and disequilibria of lower orders. To further explore such intricate interrelationships among zygotic associations, gene frequencies, and various genic disequilibria, numerical calculations are carried out. For simplicity, let us assume that there are two alleles (1 and 2) at each of the two loci. Frequencies of the ten possible genotypes are denoted as P_{11}^{11}, P_{12}^{11}, P_{12}^{12}, P_{21}^{11}, P_{22}^{11}, P_{21}^{12}, P_{22}^{12}, P_{21}^{21}, P_{22}^{21}, and P_{22}^{22}, dropping the identifiers for the two loci. These genotypic frequencies are grouped into four classes (f_{00}, f_{01}, f_{10}, and f_{11}) based on whether genotypes at individual loci are homozygous or heterozygous (Table 3). The marginal totals for the individual loci are, respectively, $f_{0.} = f_{00} + f_{01}$, $f_{1.} = f_{10} + f_{11}$, $f_{.0} = f_{00} + f_{10}$, and $f_{.1} = f_{01} + f_{11}$. Thus, the overall measure of zygotic associations (ω) can be calculated using the relations given in Table 1. To gauge the relationships between zygotic associations, gene frequencies, and various disequilibria, the two-locus genotypic frequencies are expressed in terms of disequilibrium functions (*cf.* Table 6.1 of Weir and Cockerham 1989). All types of disequilibria except for Hardy-Weinberg disequilibria affect the zygotic associations because they are genic disequilibria between the two loci.

We examine the effects of three genic disequilibria (gametic, trigenic, and quadrigenic disequilibria) on the distribution of zygotic associations. Since we assume equal gene frequencies (p) at both loci, the nongametic disequilibrium and gametic disequilibrium are equal, and so are the two trigenic disequilibria. To illustrate the three-way relationship, the effect of gene frequencies and gametic disequilibria on zygotic associations is depicted in Figure 1. In this case, the zygotic association is $\omega = 2(1 - 2p)^2D + 4D^2$, where D ($= D_{..}^{11} = -D_{..}^{12} =$

TABLE 2

Single-locus and multilocus components of variance of K, σ_K^2, under six special cases

	Variance of $K(\sigma_K^2)$	
Case	Single-locus component	Multilocus component
1. Hardy-Weinberg equilibrium with no zygotic associations	$\sum_{j=1}^{m}(h_j - h_j^2)$	0
2. Hardy-Weinberg disequilibrium with no zygotic associations	$\sum_{j=1}^{m}(H_j - H_j^2)$	0
3. Hardy-Weinberg equilibrium with gametic disequilibria	$\sum_{j=1}^{m}(h_j - h_j^2)$	$2\sum_{j<l}\sum\sum_{u=1y=1}^{r\ s}[2^jp_u^lp_y^{jl}D_{..}^{uy} + (^{jl}D_{..}^{uy})^2]$
4. Hardy-Weinberg disequilibrium with quadrigenic disequilibria	$\sum_{j=1}^{m}(H_j - H_j^2)$	$2\sum_{j<l}\sum\sum_{u=1y=1}^{r\ s}[^{jl}D_{uy}^{uy}]$
5. Maximum σ_K^2 with absolute associations	$\sum_{j=1}^{m}(H_j - H_j^2)$	$(m - 1)\sum_{j=1}^{m}(H_j - H_j^2)$
6. Maximum σ_K^2 with ranked single-locus heterozygosities	$\sum_{j=1}^{m}(H_j - H_j^2)$	$2\sum_{j=1}^{m}(j - 1)H_j - \sum_{j<l}\sum H_jH_l$

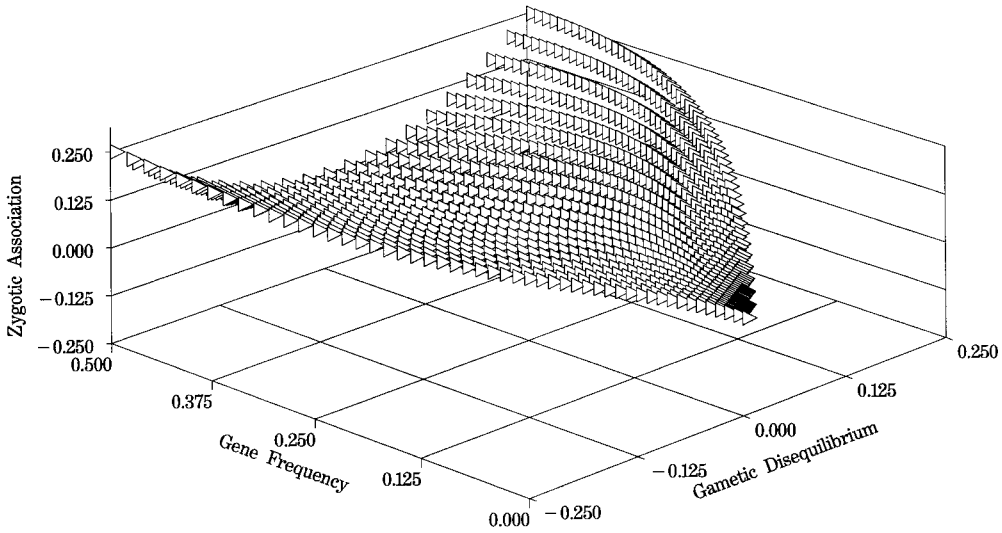

Figure 1.—Dependence of zygotic associations on gene frequency and gametic disequilibrium.

$-D_{\cdot\cdot}^{21} = D_{\cdot\cdot}^{22}$) is the gametic disequilibrium. The maximum zygotic association ($\omega = 0.25$) is obtained at $p = 0.5$ and $D = \pm 0.25$, but while ω always increases with $D > 0$, it can be negative with $D < 0$ for some gene frequencies as shown in Figure 1. The zygotic association is affected little by trigenic disequilibria, but increases with positive and decreases with negative quandrigenic disequilibria, respectively (the 3D plots for trigenic and quadrigenic disequilibria are not presented).

Estimating zygotic associations from variance of multilocus heterozygosity: The variance of K in (9a) suggests that the average zygotic associations ($\overline{\omega}$) may be obtained by

$$\overline{\omega} = \sum_{j<l}\sum \omega_{jl} = \frac{1}{2}[\sigma_K^2 - \sigma_K^2(2)], \qquad (10)$$

where $\sigma_K^2(2) = \Sigma_{j=1}^m (H_j - H_j^2)$ is for case 2 of Table 2. To estimate $\overline{\omega}$ from a sample of n diploid individuals with m polymorphic loci, one needs to estimate σ_K^2 and single-locus heterozygosities, $\{H_j\}$. There are several discussions of procedure for estimating these parameters from a sample taken from a random mating population or haploid population (*e.g.*, Brown *et al.* 1980; Brown and Burdon 1983; Chakraborty 1984). Essentially the same estimation procedure is used in the following simulation study.

The nonequilibrium population for two loci each with two alleles is constructed using the fact that each two-locus genotypic frequency can be written as a sum of the product of single-locus frequencies and its zygotic association (Table 4). For a given gene frequency (p) at a locus, Hardy-Weinberg disequilibrium ($D = D_{1\cdot}^{1\cdot} = -D_{2\cdot}^{1\cdot} = -D_{1\cdot}^{2\cdot} = D_{2\cdot}^{2\cdot}$) is bounded by

$$\max[-p^2, -(1-p)^2] \le D \le p(1-p) \qquad (11)$$

so that the frequencies of the three genotypes at this locus are completely described by p and D: $P_{1\cdot}^{1\cdot} = p^2 + D$, $P_{2\cdot}^{1\cdot} = 2p(1-p) - 2D$, and $P_{2\cdot}^{2\cdot} = (1-p)^2 + D$. We simulate three D values: zero and half the maximum and minimum possible values as given in (11). While bounds of nine individual zygotic associations can be computed from the single-locus genotypic frequencies using (2a), we choose to compute only the four associations (ω_{11}^{11}, ω_{21}^{21}, ω_{12}^{12}, and ω_{22}^{22}) since the remaining five (ω_{12}^{11}, ω_{21}^{11}, ω_{22}^{11}, ω_{22}^{12}, and ω_{22}^{21}) are simply the functions of those four associations as explained in Table 4. For simplicity, a further assumption in our simulation is that only one zygotic association is present in the population

TABLE 3

Frequencies of homozygotes and heterozygotes in terms of frequencies of 10 possible genotypes at two loci (j and l)

	Locus l		
Locus j	Homozygosity	Heterozygosity	Total
Homozygosity	$f_{00} = P_{11}^{11} + P_{12}^{12} + P_{21}^{21} + P_{22}^{22}$	$f_{01} = P_{12}^{11} + P_{22}^{21}$	$f_{0\cdot} = P_{1\cdot}^{1\cdot} + P_{2\cdot}^{2\cdot}$
Heterozygosity	$f_{10} = P_{21}^{11} + P_{22}^{12}$	$f_{11} = P_{22}^{11} + P_{21}^{12}$	$f_{1\cdot} = P_{2\cdot}^{1\cdot}$
Total	$f_{\cdot 0} = P_{\cdot 1}^{\cdot 1} + P_{\cdot 2}^{\cdot 2}$	$f_{\cdot 1} = P_{\cdot 2}^{\cdot 1}$	1

and the other three are zero. Under this assumption, the bounds of these four zygotic associations are

$$-\min(P^{1\cdot}_{1\cdot}P^{\cdot 1}_{\cdot 1},\ P^{1\cdot}_{2\cdot}P^{\cdot 1}_{\cdot 2}) \le \omega^{11}_{11} \le \min(P^{1\cdot}_{1\cdot}P^{\cdot 1}_{\cdot 2},\ P^{1\cdot}_{2\cdot}P^{\cdot 1}_{\cdot 1})$$
$$-\min(P^{1\cdot}_{1\cdot}P^{\cdot 2}_{\cdot 2},\ P^{1\cdot}_{2\cdot}P^{\cdot 1}_{\cdot 2}) \le \omega^{12}_{12} \le \min(P^{1\cdot}_{1\cdot}P^{\cdot 1}_{\cdot 2},\ P^{1\cdot}_{2\cdot}P^{\cdot 2}_{\cdot 2})$$
$$-\min(P^{2\cdot}_{2\cdot}P^{\cdot 1}_{\cdot 1},\ P^{1\cdot}_{2\cdot}P^{\cdot 1}_{\cdot 2}) \le \omega^{21}_{21} \le \min(P^{2\cdot}_{2\cdot}P^{\cdot 1}_{\cdot 2},\ P^{1\cdot}_{2\cdot}P^{\cdot 1}_{\cdot 1})$$
$$-\min(P^{2\cdot}_{2\cdot}P^{\cdot 2}_{\cdot 2},\ P^{1\cdot}_{2\cdot}P^{\cdot 1}_{\cdot 2}) \le \omega^{22}_{22} \le \min(P^{2\cdot}_{2\cdot}P^{\cdot 1}_{\cdot 2},\ P^{1\cdot}_{2\cdot}P^{\cdot 2}_{\cdot 2}). \quad (12)$$

We simulate three values of zygotic association: zero and half the maximum and minimum possible values as given in (12).

From each of 27 constructed populations [3 gene frequencies ($p = 0.1$, 0.3, and 0.5) × 3 values of Hardy-Weinberg disequilibrium × 3 values of zygotic association], 10,000 replicate samples of size $n = 30$ or $n = 100$ are drawn. For a sample of n diploid individuals, let $\tilde{X}_{tj}$ be 1 or 0 according to whether the tth individual in the sample is heterozygous or homozygous at the jth locus. Then the number of heterozygous loci for this individual is $\tilde{K}_t = \Sigma_{j=1}^{m}\tilde{X}_{tj}$. We compute the sample mean as $\tilde{K} = \Sigma_{t=1}^{n}\tilde{K}_t/n$ and the sample variance as

$$s^2_K = \frac{1}{n}\sum_{t=1}^{n}(\tilde{K}_t - \tilde{K})^2. \quad (13a)$$

Using various expectations of indicators defined for the sample (Weir *et al.* 1990; Weir 1996, pp. 142–144), it is easily seen that while the sample mean is an unbiased estimator of K, $[E(\tilde{K}) = K]$, the sample variance (13a) is not an unbiased estimator of σ^2_K, *i.e.*, $E(s^2_K) = [(n - 1)/n]\sigma^2_K$, because we have divided by n rather than the customary $(n - 1)$ in computing (13a). Clearly, the bias should be negligible unless the sample size is very small.

Under the null hypothesis of no zygotic association (H_0), we estimate $\sigma^2_K(2)$ by computing the sample variance, $s^2_K(2)$, as the sum of sample variances for m loci $\{s^2_j\}$,

$$s^2_K(2) = \sum_{j=1}^{m}s^2_j = \sum_{j=1}^{m}\left[\frac{1}{n}\sum_{t=1}^{n}(\tilde{X}_{tj} - \tilde{X}_j)^2\right], \quad (13b)$$

where $\tilde{X}_j = \Sigma_{t=1}^{n}\tilde{X}_{tj}/n$. While the estimator $s^2_K(2)$ in (13b) is slightly biased for the same reason as in computing s^2_K, its expectation and sampling variance can be readily calculated by inserting the appropriate results in (7) under interlocus independence (see also Equations 3–5 of Brown *et al.* 1980) into the well-known formulas of Kendall and Stuart (1977, Equations 10.8 and 10.9),

$$E(s^2_K|H_0) = \sum_j H_j - \sum_j H_j^2 \quad (14a)$$

and

$$\mathrm{Var}(s^2_K|H_0) = \frac{1}{n}\left\{\sum_j H_j - 7\sum_j H_j^2 + 12\sum_j H_j^3 - 6\sum_j H_j^4 + 2\left[\sum_j H_j - \sum_j H_j\right]^2\right\}. \quad (14b)$$

Two one-tailed tests are used to determine if the sample variance s^2_K is significantly greater than its expectation under zero zygotic association $\sigma^2_K(2)$. In the first test, assuming that the distribution of K under H_0 approximates a normal distribution, the statistic

$$X^2_{s^2_K} = ns^2_K/\sigma^2_K(2) \quad (15)$$

has a χ^2 distribution with n d.f., where n is the number of diploid individuals in the sample and $\sigma^2_K(2)$ is estimated using (13b) [The chi-square test (15) would have d.f. = $(n - 1)$ if the customary $(n - 1)$ is used to compute s^2_K]. The null hypothesis (H_0) is rejected if $X^2_{s^2_K}$ exceeds 43.77 or 124.34, the upper-tailed 5% critical value of χ^2 distribution with d.f. = 30 or d.f. = 100, respectively. Manly (1985, p. 331) defined a similar statistic for haploid data, but because s^2_K was computed from a sample of n^2 "dependent" gamete pairs (comparisons) for n haplotypes (Brown *et al.* 1980), the appropriate degrees of freedom for the chi-square test are yet

TABLE 4

Joint frequencies of nine genotypes at loci j and l in terms of their single-locus genotypic frequencies and zygotic associations

	Locus l			
Locus j	$l_1\ l_1$	$l_1\ l_2$	$l_2\ l_2$	Total
$j_1 j_1$	$P^{11}_{11} = P^{1\cdot}_{1\cdot}P^{\cdot 1}_{\cdot 1} + \omega^{11}_{11}$	$P^{11}_{12} = P^{1\cdot}_{1\cdot}P^{\cdot 1}_{\cdot 2} + \omega^{11}_{12}$	$P^{12}_{12} = P^{1\cdot}_{1\cdot}P^{\cdot 2}_{\cdot 2} + \omega^{12}_{12}$	$P^{1\cdot}_{1\cdot}$
$j_1 j_2$	$P^{11}_{21} = P^{1\cdot}_{2\cdot}P^{\cdot 1}_{\cdot 1} + \omega^{11}_{21}$	$P^{11}_{22} + P^{12}_{21} = P^{1\cdot}_{2\cdot}P^{\cdot 1}_{\cdot 2} + \omega^{11}_{22}$	$P^{12}_{22} = P^{1\cdot}_{2\cdot}P^{\cdot 2}_{\cdot 2} + \omega^{12}_{22}$	$P^{1\cdot}_{2\cdot}$
$j_2 j_2$	$P^{21}_{21} = P^{2\cdot}_{2\cdot}P^{\cdot 1}_{\cdot 1} + \omega^{21}_{21}$	$P^{21}_{22} = P^{2\cdot}_{2\cdot}P^{\cdot 1}_{\cdot 2} + \omega^{21}_{22}$	$P^{22}_{22} = P^{2\cdot}_{2\cdot}P^{\cdot 2}_{\cdot 2} + \omega^{22}_{22}$	$P^{2\cdot}_{2\cdot}$
Total	$P^{\cdot 1}_{\cdot 1}$	$P^{\cdot 1}_{\cdot 2}$	$P^{\cdot 2}_{\cdot 2}$	1

The zygotic associations are constrained by the single-locus frequencies such that only four of the nine zygotic associations need to be defined and the remaining five are entirely expressed in terms of the four defined zygotic associations. For example, if ω^{11}_{11}, ω^{21}_{21}, ω^{12}_{12}, and ω^{22}_{22} are defined, then the remaining five are expressed as follows: $\omega^{11}_{12} = -(\omega^{11}_{11} + \omega^{12}_{12})$, $\omega^{11}_{21} = -(\omega^{11}_{11} + \omega^{21}_{21})$, $\omega^{12}_{22} = -(\omega^{12}_{12} + \omega^{22}_{22})$, $\omega^{21}_{22} = -(\omega^{21}_{21} + \omega^{22}_{22})$, and $\omega^{11}_{22} = \omega^{11}_{11} + \omega^{12}_{12} + \omega^{21}_{21} + \omega^{22}_{22}$.

TABLE 5

Mean, standard deviation, skewness, and kurtosis of s_K^2 under zero zygotic association for two gene frequencies (p) and three Hardy-Weinberg disequilibria (D)

		Zygotic association		Mean of s_K^2		SD of s_K^2		Skewness	Kurtosis
p	D	Mean	SD	Obs.	Exp.	Obs.	Exp.	(g_1)	(g_2)
0.1	−0.005	0.0004	0.027	0.30	0.31	0.08	0.08	0.43	0.03
0.1	0.000	−0.0002	0.026	0.29	0.30	0.08	0.08	0.46	0.14
0.1	0.045	−0.0002	0.015	0.16	0.16	0.07	0.07	0.58	0.51
0.5	−0.125	0.0001	0.034	0.36	0.38	0.09	0.09	0.33	−0.19
0.5	0.000	−0.0005	0.045	0.48	0.50	0.09	0.09	0.03	−0.11
0.5	0.125	0.0004	0.034	0.36	0.38	0.09	0.09	0.36	−0.08

Expected mean and standard deviation of s_K^2 are computed using (14a) and (14b). Skewness (g_1) = $m_3/m_2\sqrt{m_2}$ and kurtosis (g_2) = $(m_4/m_2^2) - 3$, where m_2, m_3, and m_4 are the second, third, and fourth central moments of s_K^2 computed from 10,000 samples of $n = 30$. Obs., observed; Exp., expected.

to be determined. Furthermore, Haubold *et al.* (1998) recently provided a more appropriate formula to estimate $\sigma_K^2(2)$ for haploid data with an account of the interdependence between the gamete pairs. In the second test, assuming that the sampling distribution of s_K^2 approximates normality, Brown *et al.* (1980) suggested a test criterion of rejecting H_0 if $s_K^2 > L$, the upper-tailed 5% critical value for s_K^2. In our simulation, L is estimated by

$$L \cong s_K^2(2) + 1.645\sqrt{\mathrm{Var}(s_K^2|H_0)}. \qquad (16)$$

Statistical properties of sample zygotic association and s_K^2 are examined for the simulated samples of sizes $n = 30$ and $n = 100$. Despite the slight downward bias in the mean values of $s_K^2(2)$ by a factor of $(n - 1)/n$, its observed standard deviations are very close to their expected values even for $n = 30$ (Table 5), suggesting that (14b) is an adequate approximation to the sampling variance of $s_K^2(2)$. Table 5 also shows that Hardy-Weinberg disequilibrium (D) affects $\sigma_K^2(2)$ in an interesting way. Avoidance of mating between relatives ($D < 0$) increases heterozygosity whereas inbreeding ($D > 0$) decreases it. Thus, $\sigma_K^2(2)$ is expected to be greater for $D < 0$ or smaller for $D > 0$ than that for the equilibrium population ($D = 0$). However, this is not true when the gene frequency approaches $p = 0.5$. At $p = 0.5$, the maximum $\sigma_K^2(2)$ is obtained only when the population is in the Hardy-Weinberg equilibrium ($D = 0$) and any change in heterozygosity either due to avoidance of mating between relatives or to inbreeding would result in a smaller $\sigma_K^2(2)$. Negligible skewness and kurtosis suggest that the normality of the sampling distribution of $s_K^2(2)$ required for the test criterion (16) is probably adequate even though our simulation results are limited to the two loci only. As expected, the estimates of zygotic association in all simulated populations are zero or very close to zero. The increase of sample size from $n = 30$ to $n = 100$ (not presented) has improved the results only slightly.

The means of $\tilde{\omega}_{11}^{11}$ are close to their respective theoretical values and the sampling variances of $\tilde{\omega}_{11}^{11}$ increase with increasing gene frequencies at $n = 30$ (Table 6). The increase of sample sizes from 30 to 100 reduces the sampling variances and downward bias of estimated σ_K^2 (results not presented for $n = 100$). The $X_{s_K^2}^2$ test statistics are close to their expected values of 30.0 for $n = 30$ and 100.0 for $n = 100$ when zygotic association is small at low gene frequencies, but fluctuate with large positive or negative zygotic associations at more intermediate gene frequencies. The standard deviations of the chi-square statistics are also close to their expectations of 7.75 for $n = 30$ and 14.14 for $n = 100$ in most cases, but sizable discrepancies occur in the cases of large positive or negative zygotic associations. Similar patterns of sampling behaviors and properties are revealed for $\omega_{12}^{12} = \omega_{21}^{21}$ and ω_{22}^{22}.

Judging from the estimated powers of the two test statistics, the zygotic associations are detectable only when they are positive and when the gene frequencies are close to 0.5 (Table 6). Figure 2 further shows that the powers increase with the large, positive zygotic associations and that zero powers are obtained for the large, negative zygotic associations when $p = 0.5$ and $D = 0.125$. Similar patterns are observed for other values of p and D. It is of interest to note that, unlike the nonlinear relationship in Figure 2, a linear relationship of zygotic associations with the variances of K or with chi-square values is observed (results not shown). The power should be 0.05 for the cases of no zygotic associations as a 5% significance level is used to reject these null hypotheses. According to this criterion, both tests perform reasonably well. While test (16) is slightly more powerful than test (15) in most cases, the two tests essentially provide the same amount of power across the range of zygotic associations. The increase of sample size from 30 to 100 results in an increase in the power of detecting the zygotic associations. Hardy-Weinberg disequilibrium (D) has little effect on the detection. For

TABLE 6

Properties of sample variance of K, s_K^2, and its use to detect zygotic association ω_{11}^{11} with two gene frequencies (p), three Hardy-Weinberg disequilibria (D), and three zygotic associations (ω_{11}^{11}), as estimated from 10,000 samples of size $n = 30$

			$\tilde{\omega}_{11}^{11}$		s_K^2		$X^2_{s_K^2}$			
p	D	ω_{11}^{11}	Mean	SD	Mean	SD	Mean	SD	$P(C)$	$P(B)$
0.1	−0.005	0.000	0.000	0.028	0.30	0.08	29.9	7.6	0.05	0.06
0.1	−0.005	0.000	0.000	0.028	0.30	0.08	30.0	7.6	0.05	0.06
0.1	−0.005	0.001	0.001	0.028	0.30	0.09	30.3	7.8	0.06	0.06
0.1	0.000	0.000	0.000	0.026	0.29	0.08	30.0	7.6	0.05	0.05
0.1	0.000	0.000	0.000	0.026	0.28	0.08	30.0	7.6	0.05	0.05
0.1	0.000	0.002	0.002	0.027	0.29	0.08	30.4	7.8	0.06	0.06
0.1	0.045	−0.003	−0.003	0.012	0.15	0.06	28.7	5.8	0.04	0.01
0.1	0.045	0.000	0.000	0.015	0.16	0.07	30.0	7.1	0.08	0.03
0.1	0.045	0.005	0.005	0.018	0.17	0.08	32.3	8.6	0.14	0.05
0.5	−0.125	−0.016	−0.015	0.032	0.33	0.08	26.5	7.2	0.02	0.03
0.5	−0.125	0.000	0.001	0.034	0.36	0.09	30.1	7.7	0.05	0.09
0.5	−0.125	0.094	0.091	0.040	0.54	0.13	50.9	8.0	0.81	0.85
0.5	0.000	−0.063	−0.061	0.043	0.36	0.09	19.3	7.6	0.00	0.00
0.5	0.000	0.000	−0.001	0.045	0.48	0.09	29.9	7.9	0.04	0.12
0.5	0.000	0.125	0.121	0.039	0.73	0.08	51.2	6.8	0.87	0.95
0.5	0.125	−0.063	−0.061	0.023	0.24	0.01	16.2	3.7	0.00	0.00
0.5	0.125	0.000	−0.001	0.033	0.36	0.09	29.8	7.7	0.04	0.07
0.5	0.125	0.094	0.091	0.040	0.54	0.13	51.0	8.1	0.82	0.86

$P(C)$ is the proportion of time that chi square exceeds 43.8 or 124.3, the upper-tailed 5% critical value of $\chi^2_{\text{d.f.}=30}$ or $\chi^2_{\text{d.f.}=100}$, respectively (*cf.* Equation 15). $P(B)$ is the proportion of time that the sample variance of K, s_K^2, exceeds the upper-tailed 5% critical value as given in (16). $\tilde{\omega}_{11}^{11}$ is a sample estimate of ω_{11}^{11}.

example, with $p = 0.5$, $\omega_{11}^{11} = 0.0938$ for both $D = -0.125$ and $D = 0.125$. The power estimates with $n = 30$ are 0.810 for $D = -0.125$ and 0.816 for $D = 0.125$, according to the chi-square test criteria (15).

DISCUSSION

A wide range of molecular data, from isozymes to newly developed microsatellite markers, is now available for population genetic analysis. The average heterozygosity across all the loci scored has been routinely used to summarize the molecular data at hand. In the presence of nonrandom associations within and among loci, there is a need to characterize various genic disequilibria (*e.g.*, Cockerham and Weir 1973; Weir 1979; Weir and Cockerham 1989), but the number of disequilibria for multiple alleles and many loci quickly increases beyond comprehension. This article has expanded the earlier concept of zygotic associations to effectively summarize those disequilibria within and between pairs of loci [*cf.* (9a) and (9b)]. The measure of zygotic associations shares most of the properties by gametic disequilibrium, but at the zygote level (Table 4). Further, we have developed a method to compute a set of summary statistics that are used to characterize and estimate the multilocus associations in the nonequilibrium population. This development substantiates and complements the earlier development of Brown *et al.* (1980) for a Hardy-Weinberg equilibrium population in which the multilocus associations are the function of only one type of two-locus disequilibria, gametic disequilibria. For the equilibrium population, our method reduces to that of Brown *et al.* (1980) because, in this case, the gametic frequencies can be inferred from the zygotic frequencies at individual loci. However, our method should be of more general use in elucidating the multilocus organizations in nonequilibrium and equilibrium dip-

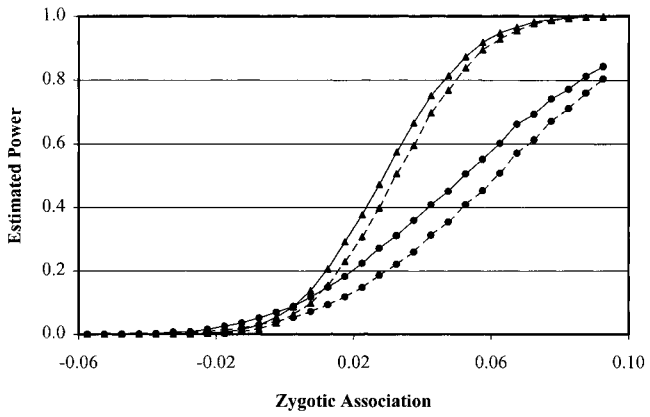

Figure 2.—The relationships between zygotic associations and the estimated powers of two tests as given in Equation 15 (dashed lines) and Equation 16 (solid lines). Each point represents the power estimated from 10,000 simulated samples of sizes $n = 30$ (●) and $n = 100$ (▲).

loid populations. For haploid data such as those from genetic assessment of bacterial or inbred plant populations, the procedures of Brown *et al.* (1980) and Haubold *et al.* (1998) should be used to construct the distribution of K through comparing all possible pairs of gametes in a population and to estimate different moments of K with an account of the interdependence between the gamete pairs for detecting multilocus associations.

Our method may be particularly useful for characterizing and estimating the multilocus associations in hybrid populations. Because these populations arise from the mixing of two or more distinct gene pools, strong Wahlund effect and selection against heterozygotes may frequently occur, thereby maintaining Hardy-Weinberg disequilibrium and zygotic associations for a long time. Given that alleles derived from the same parental populations or species tend to cluster together in the same individuals, the majority of pairwise zygotic associations should be positive, leading to an easier detection of the multilocus associations from our summary statistics. Barton and Gale (1993) have recently proposed a somewhat different method of estimating the multilocus associations from the variance of hybrid index for a hybrid population arising from the mixing of two parental gene pools. While their method is based on essentially the same strategy to summarize the multilocus data, it is of limited value in (i) detecting multilocus associations for hybrid populations arising from the mixing of more than two parental gene pools; (ii) using unfixed but informative markers for the multilocus analysis; and (iii) analyzing hybrid populations that are not in Hardy-Weinberg equilibrium.

The zygotic associations and multilocus statistics presented here are for a single nonequilibrium population. When many nonequilibrium populations are studied, the total variance of K may be partitioned into components due to the single-locus and multilocus effects of population subdivision. The method of partitioning the total variance of K among several haploid populations by Brown and Feldman (1981) can be conceivably extended to account for population subdivision related to Hardy-Weinberg disequilibrium and zygotic associations in nonequilibrium populations. However, the analysis of variance (ANOVA) of the K values may be considered for the multilocus diploid data with a complex hierarchical population structure. In this case, the ANOVA procedure of estimating hierarchical F-statistics by Yang (1998) may be extended to assess the effects of multilocus population subdivision at different levels of hierarchy.

The two sample sizes ($n = 30$ and $n = 100$) in our simulation probably represent the two ends of what may be used in most experimental population genetic studies for measuring multilocus heterozygosity. The sample of $n = 30$ appears to provide sufficient power to detect zygotic associations, agreeing with Brown *et al.*'s (1980) assertion that the multilocus statistics can be used with relatively small samples (in the order of 30). The increase in the sample size to $n = 100$ results only in slight to moderate reduction in the sampling variance of s_K^2 and an increase in the powers of detecting the multilocus associations (Figure 2). We have not simulated samples of very small sizes that may occur in practice. With small sample sizes, the validity of the assumed distributions for the sample variance of K as required by tests (15) and (16) may not be warranted. In this case, the recently developed permutation test (Guo and Thompson 1992) may be a preferred alternative to detect zygotic associations because it requires no assumptions about the distributions of multilocus statistics. In the permutation test, the null distribution [*i.e.*, the distribution of $s_K^2(2)$] is generated by randomly shuffling the single-locus zygotes among individuals in the sample. This is very similar to the randomization scheme described by Haubold *et al.* (1998) for haploid data, but it retains Hardy-Weinberg disequilibrium in the zygotes. However, the permutation test can be computationally intensive, particularly when the sample size is large. Thus, tests (15) and (16) should be useful for analyzing samples of moderate to large sizes.

I thank three reviewers for comments and constructive criticisms on earlier versions of the manuscript. This research has been supported in part by the Natural Sciences and Engineering Research Council of Canada grant OGP0183983.

LITERATURE CITED

Allard, R. W., S. K. Jain and P. L. Workman, 1968 The genetics of inbreeding populations. Adv. Genet. **14:** 55–131.

Barton, N. H., and A. G. Clark, 1990 Population structure and process in evolution, pp. 115–173 in *Population Biology*, edited by K. Wöhrmann and S. K. Jain. Springer-Verlag, Berlin.

Barton, N. H., and K. S. Gale, 1993 Genetic analysis of hybrid zones, pp. 13–45 in *Hybrid Zones and the Evolutionary Process*, edited by R. G. Harrison. Oxford University Press, New York.

Bennett, J. H., and F. E. Binet, 1956 Association between Mendelian factors with mixed selfing and random mating. Heredity **10:** 51–55.

Brown, A. H. D., 1979 Enzyme polymorphism in plant populations. Theor. Popul. Biol. **15:** 1–42.

Brown, A. H. D., and J. J. Burdon, 1983 Multilocus diversity in an outcrossing weed, *Echium plantagineum* L. Aust. J. Biol. Sci. **36:** 503–509.

Brown, A. H. D., and M. W. Feldman, 1981 Population structure of multilocus associations. Proc. Natl. Acad. Sci. USA **78:** 5913–5916.

Brown, A. H. D., M. W. Feldman and E. Nevo, 1980 Multilocus structure of natural populations of *Hordeum spontaneum*. Genetics **96:** 523–536.

Chakraborty, R., 1984 Detection of nonrandom association of alleles from the distribution of the number of heterozygous loci in a sample. Genetics **108:** 719–731.

Clegg, M. T., J. F. Kidwell, M. G. Kidwell and N. J. Daniel, 1976 Dynamics of correlated genetic systems. I. Selection in the region of the Glued locus of *Drosophila melanogaster*. Genetics **83:** 793–810.

Cockerham, C. C., and B. S. Weir, 1973 Descent measures for two loci with some applications. Theor. Popul. Biol. **4:** 300–330.

Elandt-Johnson, R. C., 1971 *Probability Models and Statistical Methods in Genetics*. Wiley, New York.

Guo, S. W., and E. A. Thompson, 1992 Performing the exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics **48:** 361–372.

Haldane, J. B. S., 1949 The association of characters as a result of inbreeding and linkage. Ann. Eugen. **15:** 15–23.
Haubold, B., M. Travisano, P. B. Rainey and R. R. Hudson, 1998 Detecting linkage disequilibrium in bacterial populations. Genetics **150:** 1341–1348.
Hedrick, P. W., 1987 Gametic disequilibrium measures: proceed with caution. Genetics **117:** 331–341.
Hedrick, P. W., S. K. Jain and L. Holden, 1978 Multilocus systems in evolution. Evol. Biol. **11:** 101–184.
Karlin, S., and A. Piazza, 1981 Statistical methods for assessing linkage disequilibrium at the HLA-A, B, C loci. Ann. Hum. Genet. **45:** 79–94.
Kendall, M., and A. Stuart, 1977 *The Advanced Theory of Statistics. Vol. 1. Distribution Theory*, Ed. 4. Griffin, London.
Lewontin, R. C., 1964 The interaction of selection and linkage. I. General considerations; heterotic models. Genetics **49:** 49–67.
Manly, B. F. J., 1985 *The Statistics of Natural Selection on Animal Populations*. Chapman and Hall, London.
Maynard Smith, J., N. H. Smith, M. O'Rouke and B. G. Spratt, 1993 How clonal are bacteria? Proc. Natl. Acad. Sci. USA **90:** 4384–4388.
Nevo, E., and A. Beiles, 1989 Genetic diversity of wild emmer wheat in Israel and Turkey: structure, evolution, and application in breeding. Theor. Appl. Genet. **77:** 421–455.
Wahlund, S., 1928 Zusammensetzung von population und korrelationserscheinung vom standpunkt der vererbungslehre aus betrachtet. Hereditas **11:** 65–106.
Weir, B. S., 1979 Inferences about linkage disequilibrium. Biometrics **35:** 235–254.
Weir, B. S., 1996 *Genetic Data Analysis II*. Sinauer Associates, Sunderland, MA.
Weir, B. S., and C. C. Cockerham, 1973 Mixed self and random mating at two loci. Genet. Res. **21:** 247–262.
Weir, B. S., and C. C. Cockerham, 1989 Complete characterization of disequilibrium at two loci, pp. 86–110 in *Mathematical Evolutionary Theory*, edited by M. W. Feldman. Princeton University Press, Princeton, NJ.
Weir, B. S., J. Reynolds and K. G. Dodds, 1990 The variance of sample heterozygosity. Theor. Popul. Biol. **37:** 235–253.
Whittam, T. S., H. Ochman and R. K. Selander, 1983 Multilocus genetic structure in natural populations of *Escherichia coli*. Proc. Natl. Acad. Sci. USA **80:** 1751–1755.
Yang, R.-C., 1998 Estimating hierarchical *F*-statistics. Evolution **52:** 950–956.
Yeh, F. C., J. Shi, R.-C. Yang, J. Hong and Z. Ye, 1994 Genetic diversity and multilocus associations in *Cunninghamia lanceolata* (Lamb.) Hook. from People's Republic of China. Theor. Appl. Genet. **88:** 465–471.

Communicating editor: A. H. D. Brown