Analysis of Multilocus Zygotic Associations

Rong-Cai Yang

Alberta Agriculture, Food and Rural Development, Edmonton, Alberta T6H 5T6, Canada, Department of Agricultural, Food and Nutritional Science, University of Alberta, Edmonton, Alberta T6G 2P5, Canada and Department of Renewable Resources, University of Alberta, Edmonton, Alberta T6G 2H1, Canada

Manuscript received April 25, 2001
Accepted for publication February 19, 2002

ABSTRACT

While nonrandom associations between zygotes at different loci (zygotic associations) frequently occur in Hardy-Weinberg disequilibrium populations, statistical analysis of such associations has received little attention. In this article, we describe the joint distributions of zygotes at multiple loci, which are completely characterized by heterozygosities at individual loci and various multilocus zygotic associations. These zygotic associations are defined in the same fashion as the usual multilocus linkage (gametic) disequilibria on the basis of gametic and allelic frequencies. The estimation and test procedures are described with details being given for three loci. The sampling properties of the estimates are examined through Monte Carlo simulation. The estimates of three-locus associations are not free of bias due to the presence of two-locus associations and vice versa. The power of detecting the zygotic associations is small unless different loci are strongly associated and/or sample sizes are large (>100). The analysis of zygotic associations not only offers an effective means of packaging numerous genic disequilibria required for a complete characterization of multilocus structure, but also provides opportunities for making inference about evolutionary and demographic processes through a comparative assessment of zygotic association vs. gametic disequilibrium for the same set of loci in nonequilibrium populations.

MULTILOCUS associations are most commonly studied at the gametic level. In this case, linkage disequilibrium or more appropriately gametic disequilibrium can be used to sufficiently describe the nonrandom associations of alleles at different loci ordered within gametes (Bennett 1954; Weir 1996). The evidence of gametic disequilibrium is important in inferring about the history of a population, the evolutionary forces governing these loci, and the location of the loci on the chromosomes. This approach to studying multilocus associations is appropriate for a haploid population where different gametes can be counted directly or for a Hardy-Weinberg equilibrium population where gametic frequencies can be inferred from genotypic (zygotic) frequencies. However, natural populations are rarely at equilibrium because of many disturbing forces such as inbreeding, population structure, and selection. In a nonequilibrium population, a complete characterization of multilocus associations requires gametic and many other genic disequilibria (Cockerham and Weir 1973; Weir 1979). Even with a moderate number of loci each with a few alleles, the number of genic disequilibria to be characterized and estimated can quickly increase beyond comprehension. Thus, with a large number of loci each with many alleles, it is necessary to have a single measure that is similar to gametic disequilibrium, but at zygote level.

Recently, Yang (2000) described characterization and estimation of such a measure for a pair of loci, which is called zygotic association. According to Yang (2000), the zygotic association is simply the deviation of two-locus zygotic frequencies from products of single-locus zygotic frequencies, but is composed of all nonallelic genic disequilibria at the two loci. Thus, in experimental population genetic studies, the zygotic association can be estimated directly by comparing the two- and single-locus zygotic frequencies observed in a sample of diploid individuals. Haldane (1949) was probably the first to recognize that the zygotic association can be generated as a result of partial inbreeding even in a linkage (gametic) equilibrium population. Subsequent studies have shown that such zygotic associations may arise from mixed selfing random mating (Bennett and Binet 1956; Allard et al. 1968; Weir and Cockerham 1973), associative overdominance (Ohta and Cockerham 1974; Charlesworth 1991), admixture of two or more distinct gene pools (Barton and Gale 1993), or heterotic selection (Mitton 1997). Thus, knowledge of extent and patterns of zygotic associations at two or more loci is essential for inferring about evolutionary and demographic processes. However, while there is substantial literature on gametic disequilibria at three or more loci (e.g., Bennett 1954; Brown 1975; Hill 1975; Thomson and Baur 1984; Barton 2000), equivalent development for multilocus zygotic associations is not

1Address for correspondence: Alberta Agriculture, Food and Rural Development, No. 301, J. G. O’Donoghue Bldg., 7000 - 113 St., Edmonton, Alberta T6H 5T6, Canada. E-mail: rongcai.yang@gov.ab.ca

yet available. In this article, we first describe the joint distributions of zygotes at multiple loci and their relationships with heterozygosities and zygotic associations. We then describe statistical procedures of estimating and testing multilocus zygotic associations from a sample of diploid individuals, with the details being given for the case of three loci. The sampling properties of the estimates are examined by computer simulation.

THEORY AND ANALYSIS

Consider a diploid population in which individual genotypes are known at m loci (e.g., codominant phenotypic markers such as the MN blood groups, allozymes, and microsatellites). Then, a genotype at a particular locus can be unambiguously recognized as a homozygote or heterozygote, depending on whether or not the two alleles at the locus are the same. Just as frequencies of genes or gametes at one or more loci are needed for defining and characterizing gametic disequilibria, frequencies of zygotes at one or more loci are required for defining and characterizing multilocus zygotic associations. These zygotic frequencies and their relationships with heterozygosities and multilocus zygotic associations are described below.

**One locus:** At a given locus, say locus $j$, the probability of an individual genotype being heterozygous or homozygous is defined as

$$f(X_j) = H_j^2(1-H_j)^{1-x_j},$$

(1)

where indicator $x_j$ takes either 1 or 0 to signal whether the genotype at the $j$th locus is a heterozygote or homozygote, and $H_j$ is the population heterozygosity at locus $j$. Thus, $f(1) = H_j$ and $f(0) = 1 - H_j$. If the population is in Hardy-Weinberg equilibrium, then the heterozygosity ($H_j$) is reduced to the gene diversity or expected heterozygosity under Hardy-Weinberg equilibrium ($h_j$). The relationship between $H_j$ and $h_j$ is given in Yang (2000, Equation 5). Such a relationship has been the basis for testing Hardy-Weinberg disequilibrium (Weir 1996).

**Two loci:** When two loci, say loci $j$ and $l$, are considered, the joint distribution of indicators, $X_j$ and $X_l$, is

$$f(X_jX_l) = f(X_j)f(X_l) + (-1)^{x_j+x_l}\omega_{jl},$$

(2)

where $f(X_j)$, for example, is given in (1) and $\omega_{jl}$ is the zygotic association between loci $j$ and $l$ (Yang 2000). Thus, $f(11) = H_j H_l + \omega_{jl}$, $f(01) = H_j (1 - H_l) - \omega_{jl}$, and $f(00) = (1 - H_j)(1 - H_l) + \omega_{jl}$. The marginal frequencies for the individual loci are: $f(0) = f(00) + f(01) = 1 - H_j$, $f(1) = f(10) + f(11) = H_j$, $f(-) = f(00) + f(10) = 1 - H_l$, and $f(-) = f(01) + f(11) = H_l$. These relationships enable $\omega_{jl}$ to be expressed in one of the following five ways:

$$\omega_{jl} = f(00) - f(0-)f(-) = -[f(10) - f(1-)f(0-)]$$

$$= -[f(01) - f(0-)f(1-)]$$

$$= f(11) - f(1-)f(1-) = f(00)f(11) - f(10)f(01).$$

Clearly, the zygotic association ($\omega_{jl}$) is bounded by the marginal zygotic frequencies at the two individual loci,

$$\max[-H_jH_l, - (1 - H_j)(1 - H_l)] \leq \omega_{jl} \leq \min[1 - H_j H_l, H_j(1 - H_l)],$$

(3)

Given that the variance of $X_j$, $\text{Var}(X_j) = H_j(1 - H_j)$, and the covariance between $X_j$ and $X_l$, $\text{Cov}(X_j, X_l) = \omega_{j},$ the correlation between heterozygosities at loci $j$ and $l$ is given by

$$\rho_{jl} = \frac{\omega_{jl}}{\sqrt{H_j(1 - H_j)H_l(1 - H_l)}}$$

To see how the zygotic association ($\omega_{jl}$) is related to different genic disequilibria including gametic disequilibrium, it is necessary to first identify the relationships between joint frequencies of homozygotes and heterozygotes in (2) and genotypic frequencies,

$$f(00) = \sum_{u=1}^s \sum_{v=1}^t \delta_{u,v} , f(01) = \sum_{u=1}^s \sum_{v=1}^t \delta_{u,v} , f(10) = \sum_{u=1}^s \sum_{v=1}^t \delta_{u,v} , f(11) = \sum_{u=1}^s \sum_{v=1}^t \delta_{u,v} .$$

(4a)

where, for example, $\delta_{u,v}$ is the frequency of genotypes at loci $j$ and $l$ from the union of gametes $j_u l_v$ and $j_v l_u$ ($u, v = 1, 2, \ldots, r, y, z = 1, 2, \ldots, s$). Then, using Cockerham and Weir’s (1973) disequilibrium functions for the two-locus frequencies ($\delta_{u,v}$), the zygotic association at loci $j$ and $l$ can be expressed in terms of individual genic disequilibria,

$$\omega_{jl} = f(00)f(11) - f(10)f(01)$$

$$= \sum_{u=1}^r \sum_{v=1}^s \left[ 2p_uq_v D_{uj} + 2p_uq_v D_{vu} + 2p_us D_{uv} \right]$$

$$+ 2p_uq_v D_{uv} + (\beta D_v) + (\beta D_u) + \gamma D_{uv},$$

(4b)

where, for example, $p_u$, is the frequency of allele $u$ at locus $j$. Clearly, each genic disequilibrium ($D$) in (4b) is the deviation of a frequency from that based on random association of genes and accounting for any lower order disequilibria. For example, the gametic disequilibrium ($\beta D_{uv}$) is the deviation of frequency of gamete $j_u l_v$ from the product of frequencies of alleles $u$ and $v$ at loci $j$ and $l$, $\beta D_{uv} = \delta_{u,v} - p_u q_v$. It is also evident from (4b) that even in a gametic equilibrium population $(\beta D_{uv} = 0)$, nonzero zygotic associations can arise from other forces such as partial inbreeding as a result of “identity disequilibrium” $(\gamma D_{uv} \neq 0)$. On the other hand, in a Hardy-Weinberg equilibrium population, the zygotic association is a function of gametic disequilibrium only. Yang (2000) has described in detail the interrelationships among gene frequencies, genic disequilibria, and two-locus zygotic associations.

**Three loci:** When three or more loci are considered jointly, two alternative approaches can be used to de-
scribe zygotic associations at these loci. The first is Bartlett’s (1935) multiplicative approach based on the multiway contingency table. In the case of three loci, the absence of three-locus zygotic association but the presence of all three pairwise associations implies that

\[ f(111)f(100)f(010)f(001) = f(110)f(101)f(011)f(000), \]

where \( f(111) \), for example, is the joint frequency of heterozygotes at the three loci. However, no explicit formulas for these joint zygotic frequencies can be given, and the numerical solutions are often sought. The second approach is the additive formulation of Bennett (1954) in which the joint frequencies of heterozygosities at three loci, for example, are linear functions of heterozygotes and two- and three-locus zygotic associations. Because of its relative simplicity relating to estimation and hypothesis testing, the additive approach is used for the subsequent development of multilocus zygotic associations. Thus, the joint distribution of indicators \( X_j, X_l, X_o \) for loci \( j, l \), and \( o \) is given by

\[
\begin{align*}
f(X_jX_lX_o) &= f(X_j)f(X_l)f(X_o) + f(X_j)(-1)^{X_j+X_l}w_{jlo} \\
&+ f(X_l)(-1)^{X_l+X_o}w_{lo} \\
&+ f(X_o)(-1)^{X_o+X_j}w_{jlo} \\
&+ f(X_j)(-1)^{X_j+X_l}w_{jlo} \\
&+ f(X_l)(-1)^{X_l+X_o}w_{lo} \\
&+ f(X_o)(-1)^{X_o+X_j}w_{jlo},
\end{align*}
\]

where \( f(X_j) \) and \( w_{jlo} \), for example, are given in (1) and (2), and \( w_{jlo} \) is the three-locus zygotic association. Three-locus independence is implied by zero three-locus association, but with the presence of all pairwise associations, \( i.e., \)

\[
f_0(X_jX_lX_o) = f(X_j)f(X_l)f(X_o) + f(X_j)(-1)^{X_j+X_l}w_{jlo} \\
+ f(X_l)(-1)^{X_l+X_o}w_{lo} \\
+ f(X_o)(-1)^{X_o+X_j}w_{jlo},
\]

Unlike the two-locus associations (\( \omega_{jl} \)) where the minimum is always zero for both \( \omega_j > 0 \) and \( \omega_l < 0 \) (cf. Equation 3), the three-locus zygotic associations may sometimes be bounded away from zero. In other words, both positive and negative \( \omega_{jlo} \) values may be constrained by their own minimum and maximum values. Let us first define two quantities,

\[
\begin{align*}
\beta_1 &= \min\{f_0(111), f_0(100), f_0(010), f_0(001)\} \\
\beta_2 &= \min\{f_0(110), f_0(101), f_0(011), f_0(000)\},
\end{align*}
\]

where \( f_0(111) = H_jH_lH_o + H_j\omega_o + H_l\omega_o + H_o\omega_j \), for example, is obtained using (6). Thus, following the development by Thomson and Baur (1984) for the three-locus gametic disequilibrium, the maximum and minimum values for \( \omega_{jlo} > 0 \) [denoted as \( \omega_{jlo}^{\text{max}} \)] and \( \omega_{jlo}^{\text{min}} \) are

\[
\begin{align*}
\omega_{jlo}^{\text{max}} &= \beta_2 \\
\omega_{jlo}^{\text{min}} &= \max[0, -\beta_1]
\end{align*}
\]

and those for \( \omega_{jlo} < 0 \) [denoted as \( \omega_{jlo}^{\text{max}} \) and \( \omega_{jlo}^{\text{min}} \)] are

\[
\omega_{jlo}^{\text{max}} = -\beta_1 \\
\omega_{jlo}^{\text{min}} = \max[0, -\beta_2].
\]

Since any of the eight \( f(X_jX_lX_o) \) values can be negative, neither \( \omega_{jlo}^{\text{max}} \) nor \( \omega_{jlo}^{\text{min}} \) is necessarily zero. Clearly, the \( \beta_1 \) and \( \beta_2 \) values can be used to determine the sign and range of \( \omega_{jlo} \),

\[
\omega_{jlo}^{\text{max}} \leq \omega_{jlo} \leq \omega_{jlo}^{\text{max}}, \quad \text{if } \beta_1 < 0, \beta_2 > 0 \text{ and } |\beta_1| > |\beta_2| \\
\omega_{jlo}^{\text{max}} \leq \omega_{jlo} \leq \omega_{jlo}^{\text{min}}, \quad \text{if } \beta_1 > 0, \beta_2 < 0 \text{ and } |\beta_1| > |\beta_2| \\
\omega_{jlo}^{\text{max}} \leq \omega_{jlo} \leq \omega_{jlo}^{\text{max}}, \quad \text{if } \beta_1 \geq 0 \text{ and } \beta_2 \leq 0,
\]

but some pairs of the \( \beta_1 \) and \( \beta_2 \) values (\( e.g., \beta_1 < 0 \) and \( \beta_2 < 0 \)) would not lead to the definable \( \omega_{jlo} \). Given \( 0 \leq H_j, H_l, H_o \leq 1 \), all profiles of heterozygosities \( (H_j, H_l, H_o) \), but with no two-locus associations (\( i.e., \omega_j = \omega_l = \omega_o = 0 \)), would produce the condition of \( \beta_1 \geq 0 \) and \( \beta_2 \geq 0 \) \( [i.e., f_0(X_jX_lX_o) = f(X_j)f(X_l)f(X_o) \geq 0] \) and thus definable \( \omega_{jlo} \) values. However, given a heterozygosity profile, not all configurations of two-locus zygotic associations \( \omega_{jlo}, \omega_{jlo}, \omega_{jlo} \) would lead to the conditions given in (8) under which \( \omega_{jlo} \) can be defined.

Table 1 shows some numerical examples to illustrate effects of heterozygosities, which are pairwise zygotic associations on the ranges of the three-locus association (\( \omega_{jlo} \)). For example, with \( H_j = H_l = 0.05 \) and \( H_o = 0.1 \), the ranges for \( \omega_j, \omega_l, \) and \( \omega_o \) are \([-0.0025 \leq \omega_j \leq 0.0475, -0.005 \leq \omega_l \leq 0.045, \) and \(-0.005 \leq \omega_o \leq 0.045 \), respectively. Each of these three ranges is divided by 19 to obtain 20 equally divided values from the minimum to the maximum. Thus, there are \( 8000 \) \( (20 \omega_j \times 20 \omega_l \times 20 \omega_o) \) configurations of the two-locus zygotic associations that can be used to define the ranges for \( \omega_{jlo} \). Of these 8000 configurations, 240 have the ranges with \( \omega_{jlo} < 0, 1496 \) have the ranges with \( \omega_{jlo} > 0, \) and 2158 have the ranges of \( \omega_{jlo}^{\text{max}} \leq \omega_{jlo} \leq \omega_{jlo}^{\text{max}}, \) but the remaining 4106 configurations do not lead to any definable \( \omega_{jlo} \). In each of these three cases, we identify a configuration that leads to the maximum range of \( \omega_{jlo} \) (it is noted that many other configurations may also lead to the same maximum range in each case).

As in the two-locus case, we wish to learn how the three-locus zygotic association (\( \omega_{jlo} \)) is related to genic disequilibria. To focus our interest in the relationships between zygotic associations and gametic disequilibria, we assume that a zygote is formed from random union of two gametes (\( i.e., \) the population is in Hardy-Weinberg equilibrium). Because, under this assumption, the zygotic frequencies are just products of the gametic frequencies, the three-locus zygotic and gametic frequencies are directly related. For convenience, we consider only the case of two alleles at each of the three loci and the notation in this case is varied to reduce the superscripts and subscripts. The frequencies of the two alleles \( f \) and \( j \) at locus \( j \) are \( p_j \) and \( p_l = (1 - p_j) \), those of the two alleles \( L \) and \( l \) at locus \( l \) are \( p_l \) and \( p_o = (1 - p_l) \), and those of the two alleles \( O \) and \( o \) at locus \( o \) are
TABLE 1
Ranges for three-locus zygotic associations

<table>
<thead>
<tr>
<th>$H_1$</th>
<th>$H_2$</th>
<th>$H_3$</th>
<th>Bounds on $\omega_{jkl}$</th>
<th>$\nu^*$</th>
<th>$\omega_{j}$</th>
<th>$\omega_{k}$</th>
<th>$\omega_{l}$</th>
<th>Range for $\omega_{jkl}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.05</td>
<td>0.1</td>
<td>$\omega_{jkl} &lt; 0$</td>
<td>240</td>
<td>0.0396</td>
<td>-0.0024</td>
<td>0.0003</td>
<td>-0.0041 -0.0015</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\omega_{jkl} &gt; 0$</td>
<td>1496</td>
<td>0.0238</td>
<td>0.0239</td>
<td>0.0229</td>
<td>0.0002 0.0213</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\omega_{jkl}^{(\max)} = \omega_{jkl} = \omega_{jkl}^{(\max)}$</td>
<td>2158</td>
<td>0.0212</td>
<td>0.0213</td>
<td>0.0213</td>
<td>-0.0045 0.0192</td>
</tr>
<tr>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>$\omega_{jkl} &lt; 0$</td>
<td>294</td>
<td>-0.0047</td>
<td>0.0742</td>
<td>0.0058</td>
<td>-0.0085 -0.0033</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\omega_{jkl} &gt; 0$</td>
<td>933</td>
<td>0.0479</td>
<td>0.0479</td>
<td>0.0479</td>
<td>0.0004 0.0425</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\omega_{jkl}^{(\max)} = \omega_{jkl} = \omega_{jkl}^{(\max)}$</td>
<td>1698</td>
<td>0.0426</td>
<td>0.0426</td>
<td>0.0426</td>
<td>-0.0085 0.0588</td>
</tr>
<tr>
<td>0.1</td>
<td>0.1</td>
<td>0.5</td>
<td>$\omega_{jkl} &lt; 0$</td>
<td>805</td>
<td>0.0584</td>
<td>-0.0237</td>
<td>-0.0237</td>
<td>-0.0295 -0.0032</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\omega_{jkl} &gt; 0$</td>
<td>805</td>
<td>0.0584</td>
<td>0.0237</td>
<td>0.0237</td>
<td>0.0032 0.0258</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\omega_{jkl}^{(\max)} = \omega_{jkl} = \omega_{jkl}^{(\max)}$</td>
<td>2288</td>
<td>0.0426</td>
<td>0.0026</td>
<td>0.0026</td>
<td>-0.0216 0.0258</td>
</tr>
<tr>
<td>0.1</td>
<td>0.3</td>
<td>0.5</td>
<td>$\omega_{jkl} &lt; 0$</td>
<td>1246</td>
<td>0.0384</td>
<td>-0.0132</td>
<td>0.0868</td>
<td>-0.0337 -0.0021</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\omega_{jkl} &gt; 0$</td>
<td>1284</td>
<td>0.0384</td>
<td>0.0184</td>
<td>-0.0395</td>
<td>0.0011 0.0326</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\omega_{jkl}^{(\max)} = \omega_{jkl} = \omega_{jkl}^{(\max)}$</td>
<td>2761</td>
<td>0.0226</td>
<td>0.0263</td>
<td>0.1026</td>
<td>-0.0321 0.0153</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>$\omega_{jkl} &lt; 0$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0 -0.0150 0.0150</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\omega_{jkl} &gt; 0$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0 -0.1250 0.1250</td>
</tr>
</tbody>
</table>

Shown are configurations of two-locus zygotic associations ($\omega_{jkl}$) that lead to definable three-locus associations and examples for resultant maximum ranges for three-locus zygotic association ($\omega_{jkl}$) for each heterozygosity profile ($H_1$, $H_2$, and $H_3$).

*The number of $\omega_{jkl}$ configurations that satisfy the bound set for $\omega_{123}$ as specified in column 4. The total number of $\omega_{jkl}$ configurations is 8000 ($20 \times 20 \times 20 \times 20$).

$p_0$ and $p_1 = 1 - p_0$. The gametic disequilibrium between the $j$-loci is denoted as $D_{j}$, between the $j$-loci as $D_{j}$, between the $l$-loci as $D_{l}$, and the three-locus gametic disequilibrium as $D_{jkl}$. These seven parameters are used to obtain the vector of eight three-locus gametic frequencies, $g = [g_{p0} g_{pl} g_{Po} g_{Po} g_{pl} g_{pl} g_{PL} g_{PL}]$, where $g_{p0}$, for example, is

$$g_{p0} = p_0 p_0 + p_1 D_{o} + p_0 D_{j} + p_0 D_{l}$$

(Brown 1975; Thomson and Baur 1984). Under random mating, the gametic frequencies are just appropriate sums of elements of the matrix $gg^T$, but the level of heterozygosity is not in any particular sorted order. Fortunately, an $(8 \times 8)$ matrix $G$ is found to enable the relationships between heterozygositites and gametic frequencies to be expressed directly, i.e., $f = Gg$, where $f = [f(000) f(001) f(010) f(011) f(100) f(101) f(110) f(111)]'$ and $G = [g_{p0} g_{pl} g_{Po} g_{Po} g_{pl} g_{pl} g_{PL} g_{PL}]$.

Note that $G$ is a symmetric matrix and the eight gametic frequencies in its first row or column are identical to those in $g$ whereas rows (columns) 2 to 8 are just different rearrangements of the eight gametic frequencies required to obtain the desired zygotic frequencies in $f$.

In the absence of two-locus gametic disequilibria (i.e., $D_{j} = D_{j} = D_{l} = 0$), the expressions of zygotic frequencies in $f$ are greatly simplified. For example, the frequency of homozygotes at all three loci, $j$, $l$, and $o$, is given by

$$f(000) = (p_0 p_0 D_o + p_1 p_1 + D_j)^2 + (p_0 p_0 D_o + D_j)^2 + (p_0 p_0 D_o + D_j)^2 + (p_0 p_0 D_o + D_j)^2$$

$$= (1 - H_j) (1 - H_l) - 2 (1 - 2 p_0) (1 - 2 p_0) D_{jkl} + 8 D_j^2$$

Here $H_j$, for example, is the same as the expected heterozygosity under Hardy-Weinberg equilibrium ($h_j = 2 p_0 p_1$). Given that $\omega_{jkl} = 2 (1 - 2 p_0) (1 - 2 p_0) D_{jkl} + 4 D_j^2$ (Yang 2000), for example, $\omega_{j} = \omega_{j} = \omega_{j} = 0$ in the absence of the pairwise gametic disequilibria. Thus, the three-locus zygotic association can be expressed in terms of gene frequencies and three-locus gametic disequilibrium

$$\omega_{jkl} = - [f(000) - (1 - H_j) (1 - H_l) (1 - H_j)]$$

$$= 2 (1 - 2 p_0) (1 - 2 p_0) D_{jkl} - 8 D_j^2$$
(cf. Equation 5). Table 2 lists the values of three-locus zygotic association ($\omega_{jlo}$) in the presence of three-locus gametic disequilibrium ($D_{jlo}$) but absence of the pairwise disequilibria ($D_j = D_o = D_o = 0$). The Hardy-Weinberg equilibrium population is assumed so that the values of $\omega_{jlo}$ are directly related to $D_{jlo}$ and gene frequencies ($p_1$, $p_2$, and $p_3$).

The combined effect of two- and three-locus gametic disequilibria on the values of $\omega_{jlo}$ is also examined (numerical results are not presented). The joint contribution of two- and three-locus gametic disequilibria to $\omega_{jlo}$ greatly cloaks their relationships with $\omega_{jlo}$. However, there are clearly cases where the $\omega_{jlo}$ values exceed the limits of $\omega_{jlo}$ under the cases of no pairwise disequilibria. For example, for $p_1 = p_2 = p_3 = 0.5$, the ranges for $D_{jlo}$, $D_{jlo}$, and $D_{jlo}$ are all from $-0.25$ to $0.25$, but permissible values of $D_{jlo}$ are determined by different combinations of these pairwise disequilibria with the given gene frequencies (Thomson and Baur 1984). It is found that

### Table 2

<table>
<thead>
<tr>
<th>$p_1$</th>
<th>$p_2$</th>
<th>$p_0$</th>
<th>$-p_1p_2p_0$</th>
<th>$p_1p_2(1-p_0)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>-0.001</td>
<td>0.009</td>
</tr>
<tr>
<td>0.1</td>
<td>0.1</td>
<td>0.2</td>
<td>-0.002</td>
<td>0.008</td>
</tr>
<tr>
<td>0.1</td>
<td>0.1</td>
<td>0.3</td>
<td>-0.003</td>
<td>0.007</td>
</tr>
<tr>
<td>0.1</td>
<td>0.1</td>
<td>0.4</td>
<td>-0.004</td>
<td>0.006</td>
</tr>
<tr>
<td>0.1</td>
<td>0.1</td>
<td>0.5</td>
<td>-0.005</td>
<td>0.005</td>
</tr>
<tr>
<td>0.1</td>
<td>0.2</td>
<td>0.2</td>
<td>-0.004</td>
<td>0.016</td>
</tr>
<tr>
<td>0.1</td>
<td>0.2</td>
<td>0.3</td>
<td>-0.006</td>
<td>0.014</td>
</tr>
<tr>
<td>0.1</td>
<td>0.2</td>
<td>0.4</td>
<td>-0.008</td>
<td>0.012</td>
</tr>
<tr>
<td>0.1</td>
<td>0.2</td>
<td>0.5</td>
<td>-0.010</td>
<td>0.010</td>
</tr>
<tr>
<td>0.1</td>
<td>0.3</td>
<td>0.3</td>
<td>-0.009</td>
<td>0.021</td>
</tr>
<tr>
<td>0.1</td>
<td>0.3</td>
<td>0.4</td>
<td>-0.012</td>
<td>0.018</td>
</tr>
<tr>
<td>0.1</td>
<td>0.3</td>
<td>0.5</td>
<td>-0.015</td>
<td>0.015</td>
</tr>
<tr>
<td>0.1</td>
<td>0.4</td>
<td>0.4</td>
<td>-0.016</td>
<td>0.024</td>
</tr>
<tr>
<td>0.1</td>
<td>0.4</td>
<td>0.5</td>
<td>-0.020</td>
<td>0.020</td>
</tr>
<tr>
<td>0.1</td>
<td>0.5</td>
<td>0.5</td>
<td>-0.025</td>
<td>0.025</td>
</tr>
<tr>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>-0.008</td>
<td>0.032</td>
</tr>
<tr>
<td>0.2</td>
<td>0.2</td>
<td>0.3</td>
<td>-0.012</td>
<td>0.028</td>
</tr>
<tr>
<td>0.2</td>
<td>0.2</td>
<td>0.4</td>
<td>-0.016</td>
<td>0.024</td>
</tr>
<tr>
<td>0.2</td>
<td>0.2</td>
<td>0.5</td>
<td>-0.020</td>
<td>0.020</td>
</tr>
<tr>
<td>0.2</td>
<td>0.3</td>
<td>0.3</td>
<td>-0.018</td>
<td>0.042</td>
</tr>
<tr>
<td>0.2</td>
<td>0.3</td>
<td>0.4</td>
<td>-0.024</td>
<td>0.036</td>
</tr>
<tr>
<td>0.2</td>
<td>0.3</td>
<td>0.5</td>
<td>-0.030</td>
<td>0.030</td>
</tr>
<tr>
<td>0.2</td>
<td>0.4</td>
<td>0.4</td>
<td>-0.032</td>
<td>0.048</td>
</tr>
<tr>
<td>0.2</td>
<td>0.4</td>
<td>0.5</td>
<td>-0.040</td>
<td>0.040</td>
</tr>
<tr>
<td>0.2</td>
<td>0.5</td>
<td>0.5</td>
<td>-0.050</td>
<td>0.050</td>
</tr>
<tr>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>-0.027</td>
<td>0.063</td>
</tr>
<tr>
<td>0.3</td>
<td>0.3</td>
<td>0.4</td>
<td>-0.036</td>
<td>0.054</td>
</tr>
<tr>
<td>0.3</td>
<td>0.3</td>
<td>0.5</td>
<td>-0.045</td>
<td>0.045</td>
</tr>
<tr>
<td>0.3</td>
<td>0.4</td>
<td>0.4</td>
<td>-0.048</td>
<td>0.072</td>
</tr>
<tr>
<td>0.3</td>
<td>0.4</td>
<td>0.5</td>
<td>-0.060</td>
<td>0.060</td>
</tr>
<tr>
<td>0.3</td>
<td>0.5</td>
<td>0.5</td>
<td>-0.075</td>
<td>0.075</td>
</tr>
<tr>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
<td>-0.064</td>
<td>0.096</td>
</tr>
<tr>
<td>0.4</td>
<td>0.4</td>
<td>0.5</td>
<td>-0.080</td>
<td>0.080</td>
</tr>
<tr>
<td>0.4</td>
<td>0.5</td>
<td>0.5</td>
<td>-0.100</td>
<td>0.100</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>-0.125</td>
<td>0.125</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$p_1$</th>
<th>$p_2$</th>
<th>$p_0$</th>
<th>$-p_1p_2p_0$</th>
<th>$p_1p_2(1-p_0)$</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Size of negative $D_{jlo}$</th>
<th>Size of positive $D_{jlo}$</th>
<th>Value of $\omega_{jlo}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>100%</td>
<td>50%</td>
<td>0.0010</td>
</tr>
<tr>
<td>50%</td>
<td>100%</td>
<td>0.0005</td>
</tr>
</tbody>
</table>

Shown are three-locus zygotic associations ($\omega_{jlo}$) in the presence of three-locus gametic disequilibrium ($D_{jlo}$) but absence of the pairwise disequilibria ($D_j = D_o = D_o = 0$). The Hardy-Weinberg equilibrium population is assumed so that the values of $\omega_{jlo}$ are directly related to $D_{jlo}$ and gene frequencies ($p_1$, $p_2$, and $p_3$).
when $D_{ij} = D_{ji} = D_{ik} = D_{ki} = -0.25$, $\omega_{ij}$ is $-0.5$, which exceeds the limit of $-0.125$ in the case of no two-locus disequilibria ($D_j = D_k = D_{ij} = 0$).

**More than three loci:** The extension to four or more loci following Bennett (1954) is straightforward. For example, the joint distribution of indicators $X_1$, $X_2$, $X_3$, and $X_4$ for loci $j$, $l$, $o$, and $q$, respectively, is given by

\begin{align}
 f(X_jX_lX_oX_q) &= f(X_j)f(X_l)f(X_o)f(X_q) \\
 &+ f(X_j)(-1)^{5\gamma_j\gamma_l\gamma_o\gamma_q}\omega_{ij} \\
 &+ f(X_l)(-1)^{5\gamma_j\gamma_l\gamma_o\gamma_q}\omega_{il} \\
 &+ f(X_o)(-1)^{5\gamma_j\gamma_l\gamma_o\gamma_q}\omega_{io} \\
 &+ f(X_q)(-1)^{5\gamma_j\gamma_l\gamma_o\gamma_q}\omega_{iq} \\
 &+ f(X_j)(-1)^{5\gamma_j\gamma_l\gamma_o}\omega_{ij} \\
 &+ f(X_l)(-1)^{5\gamma_j\gamma_l\gamma_o}\omega_{il} \\
 &+ f(X_o)(-1)^{5\gamma_j\gamma_l\gamma_o}\omega_{io} \\
 &+ f(X_q)(-1)^{5\gamma_j\gamma_l\gamma_o}\omega_{iq} \\
 &+ (1)^{5\gamma_j\gamma_l\gamma_o\gamma_q}[\omega_{ij}\omega_{il} + \omega_{io}\omega_{iq} + \omega_{ij}\omega_{iq} + \omega_{il}\omega_{iq}],
\end{align}

where $f(X_j)$, $\omega_{ij}$, and $\omega_{ij}$, for example, are given in (1), (2), and (5), and $\omega_{ij}$ is the four-locus zygotic association. In other words, the frequencies of 16 zygote classes for loci $j$, $l$, $o$, and $q$ can be uniquely defined in terms of the four heterozygosities for individual loci, the six pairwise zygotic associations, four three-locus zygotic associations, and one four-locus association. The three products of pairwise zygotic associations in the last term of (12) arise from the “two-locus” recombination, a distinct feature inherent in the associations for more than three linked loci (Bennett 1954; Lewontin 1964; Cockram and Tachida 1986). A set of functions $f_{ij}(X_jX_lX_oX_q)$, can be defined in a similar manner as (6) for $f_{ij}(X_jX_lX_o)$ to provide the basis for defining the range of $\omega_{ij}$. The higher order zygotic associations are required for deriving higher moments of the number of heterozygous loci (Yang 2000) or covariances of two-locus sample zygotic associations (Weir 1996, Chap. 4).

**STATISTICAL INFERENCE**

**Maximum-likelihood estimation:** For $m$ loci, there are $2^m$ possible classes of zygotes with two extreme classes being $m$-locus homozygotes ($00\cdots0$) and $m$-locus heterozygotes ($11\cdots1$). A total of $2^m - 1$ parameters can be estimated. Here we focus on the estimation for the case of three loci ($m = 3$), letting $j = 1$, $l = 2$, and $o = 3$ for convenience. Table 3 lists the eight classes of zygotes with the expected frequencies of $f(000)$, $f(001)$, $f(010)$, $f(011)$, $f(100)$, $f(101)$, $f(110)$, and $f(111)$ as obtained from (5). Seven parameters are estimable: three heterozygosities ($H_j$, $H_l$, and $H_o$), three two-locus zygotic associations ($\omega_{12}$, $\omega_{13}$, and $\omega_{23}$), and one three-locus zygotic association ($\omega_{123}$). If a sample of $n$ individuals is taken from a diploid population and if the numbers of each class in the sample are assumed to be multinomially distributed, frequencies of these classes can be estimated using the maximum-likelihood (ML) method. Let $n_{ijk}$ be the numbers of the $abch$ class of zygotes with $a$, $b$, and $c$ representing indicators $X_j$, $X_l$, and $X_o$, respectively. Thus the ML estimates of $\hat{f}$ are given by

\[ \hat{f}(abc) = \frac{n_{ijk}}{n}. \]

Various one- and two-locus marginal frequencies are given by sums of the three-locus frequencies as indicated by dots for the indices summed. For example, $\hat{f}(abc) = \sum_{i=0}^{1} \sum_{j=0}^{1} \sum_{k=0}^{1} n_{ijk}$. Note that the one-locus marginal frequencies, $\hat{f}(1\cdot\cdot) = H_j$, $\hat{f}(\cdot\cdot1) = H_l$, and $\hat{f}(\cdot\cdot\cdot) = H_o$ are the estimates of heterozygosities at loci 1, 2, and 3, respectively. The zygotic associations for two loci (e.g., $\omega_{12}$) and for all three loci ($\omega_{123}$) are estimated as

\[ \hat{\omega}_{12} = \hat{f}(1\cdot\cdot) - \hat{f}(1\cdot\cdot1) \]

\[ \hat{\omega}_{123} = \hat{f}(111) - \hat{f}(1\cdot\cdot)\hat{\omega}_{23} - \hat{f}(\cdot\cdot1)\hat{\omega}_{13} - \hat{f}(\cdot\cdot\cdot)\hat{\omega}_{12} - \hat{f}(\cdot\cdot1)\hat{f}(\cdot\cdot\cdot), \]

respectively. These ML estimates are biased as indicated from their expected values,

\[ E(\hat{\omega}_{12}) = \frac{(n - 1)}{n} \omega_{12} \]

\[ E(\hat{\omega}_{123}) = \frac{(n - 3)}{n} \omega_{123}. \]

Sampling variances of linear combinations of multinomial variables are known exactly. For example, $\text{var}(\hat{f}) = \frac{H_j}{n} (1 - H_j)/n$. The sampling variances of zygotic association estimates involve quadratic functions of observed heterozygosities and can be calculated using Fisher’s (1954) expression for the approximate variance of a function of multinomial observations $n_{ijk}$, for example, with expectations $E(n_{ijk}) = nf(abc)$. The sampling variances of $\hat{\omega}_{12}$ and $\hat{\omega}_{123}$ are

\[ \text{Var}(\hat{\omega}_{12}) = \frac{1}{n} [A_1A_2 + B_1B_2\omega_{12} - \omega_{12}^2] \]

\[ \text{Var}(\hat{\omega}_{123}) = \frac{1}{n} [A_1A_2A_3 + 6A_1A_2\omega_{12} + A_1B_2B_3 - \omega_{12}^3] \]

\[ + A_1B_2B_3 - 6A_1B_2\omega_{12} - 6B_2\omega_{12} - 2B_2\omega_{12}^2 - 2B_2\omega_{12}^3 - \omega_{123}^3. \]

where $A_i = H_i(1 - H_i)$ and $B_i = (1 - 2H_i)$ with $i = 1, 2, 3$. Equations 14a and 14b are essentially the same as Equations 3 and 13 of Brown (1975) for the sampling variances of two- and three-locus gametic disequilibria.

**Hypothesis testing:** Since the ML estimate $\hat{\omega}_i$ is approximately normally distributed, i.e., $\hat{\omega}_i \sim N(E(\hat{\omega}_i), \text{Var}(\hat{\omega}_i))$, a test statistic ($\chi^2$) that is constructed, after setting $\omega_i$ to zero in both $E(\hat{\omega}_i)$ and $\text{Var}(\hat{\omega}_i)$, is distributed as chi
square with 1 d.f., where subscript $i$ indexes for 12, 13, 23, and 123 for the three loci. For example, the test statistic for estimated zygotic association at loci 1 and 2 ($\hat{\omega}_{12}$),

$$\hat{X}^2 = \frac{\hat{\omega}_{12}^2}{\text{Var}(\hat{\omega}_{12}|\omega_{12} = 0)} = \frac{n\hat{\omega}_{12}^2}{H_1(1 - H_1) H_2(1 - H_2)},$$

is used to test for $\omega_{12} = 0$.

**Simulation:** Monte Carlo simulation is carried out to examine the performance of the estimators and test statistics for the four zygotic associations, $\omega_{12}$, $\omega_{13}$, $\omega_{23}$, and $\omega_{123}$. The eight frequencies of zygote classes, $J(X_i X_j X_k)$, can be constructed from given values of the four zygotic associations and three heterozygosities, $H_1$, $H_3$, and $H_3$ (cf. Table 3). For each of the 18 configurations given in Table 1, we consider three values (maximum, minimum, and zero) of three-locus zygotic association ($\omega_{123}$). Thus, there are a total of 54 populations constructed. From each population, 10,000 replicate samples of sizes $n = 30, 100, 300$ are drawn. Estimation and test are made for each simulated sample and descriptive statistics are calculated across all the samples.

Table 4 presents means and standard deviations (SD) of estimates from the simulated samples for 8 of the 54 constructed populations described above. The simulation results are given only for $n = 30$ and $n = 300$. It is evident that the averages of estimated zygotic associations are very close to their theoretical values when there is no or little association. In this case, bias is expected to be negligible as it arises only from the factor of $(n - 1) / n$. However, when such a case is not true, there can be a substantial amount of bias in the estimates. For example, for the case of $H_1 = 0.1$, $H_3 = 0.3$, and $H_3 = 0.5$ with $\omega_{12} = 0.023$, $\omega_{13} = 0.026$, $\omega_{23} = 0.103$, and $\omega_{123} = -0.032$, the respective averaged estimates of $\omega_{12}$, $\omega_{13}$, $\omega_{23}$, and $\omega_{123}$ are 0.023, 0.012, 0.099, and $-0.025$ for $n = 30$ and 0.024, 0.012, 0.103, and $-0.027$ for $n = 300$.

While $\omega_{12}$ and $\omega_{23}$ are almost identical to their theoretical values, $\omega_{13}$ is only less than one-half of its true value and $\omega_{123}$ is also a downwardly biased estimate of $\omega_{123}$. However, when $\omega_{123}$ is set to zero, the estimates of all three two-locus zygotic associations are unbiased; conversely $\omega_{123}$ is also an unbiased estimate of $\omega_{123}$ when there are no two-locus associations.

While means of estimated zygotic associations for the two sample sizes in Table 4 are similar, the larger sample leads to a much smaller SD. It is thus no surprise to see that the larger sample leads to a much greater power of detecting nonzero zygotic associations. The estimated powers for the cases of no zygotic associations in Table 4 are close to 0.05 as expected because a 5% significance level is used to reject these null hypotheses. To further explore the effect of sample sizes on the power, we calculate the powers of detecting three-locus associations in the presence of two-locus associations (i) $\omega_{12} = \omega_{13} = \omega_{23} = 0.0213$ and (ii) $\omega_{12} = 0.0226$, $\omega_{13} = 0.0263$, and $\omega_{23} = 0.0263$ (cf. Table 1) for sample sizes of 30, 100, 300, 500, and 1000 (the three loci are indexed as $j = 1$, $l = 2$, and $o = 3$). The critical value with a 5% significance level, $c_{0.05}$, which determines the rejection region for the hypothesis $H_1$, $\omega_{123} = 0$, is

$$c_{0.05} = 1.96\sqrt{\text{Var}(\hat{\omega}_{123}|\omega_{123} = 0)}.$$

Thus, the power (the probability of rejecting the false $H_1$) is given by

$$P(\hat{\omega}_{123} > c_{0.05}) = \Phi\left(-\frac{(\omega_{123} - c_{0.05})}{\sqrt{\text{Var}(\hat{\omega}_{123})}}\right) + \Phi\left(\frac{(\omega_{123} - c_{0.05})}{\sqrt{\text{Var}(\hat{\omega}_{123})}}\right),$$

where $\Phi(x)$ is the cumulative density function of normal variate $x$. The results of power calculations are displayed in Figure 1. The power is very small when zygotic associations are close to zero and when sample sizes are small ($<100$). These results corroborate those by Brown (1975) and Thompson et al. (1988) on detecting genetic disequilibria at two or three loci. On the other hand, Brown et al. (1980) and Yang (2000) have concluded that the multilocus association in the variance of the number of heterozygous loci ($\sigma^2_3$) is detectable in a sample of moderate size ($\approx 50$). However, the magnitude of such association in $\sigma^2_3$ may be appreciably larger than an individual association examined here.
**TABLE 4**

Powers of detecting three-locus zygotic associations

<table>
<thead>
<tr>
<th>$\omega_{12}$</th>
<th>$\omega_{13}$</th>
<th>$\omega_{23}$</th>
<th>$\omega_{123}$</th>
<th>$n$</th>
<th>Mean</th>
<th>SD</th>
<th>Mean</th>
<th>SD</th>
<th>Mean</th>
<th>SD</th>
<th>Mean</th>
<th>SD</th>
<th>$\hat{p}(\chi^2_{1})$</th>
<th>$\hat{p}(\chi^2_{1})$</th>
<th>$\hat{p}(\chi^2_{2})$</th>
<th>$\hat{p}(\chi^2_{3})$</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.002</td>
<td>30</td>
<td>0.000</td>
<td>0.099</td>
<td>0.000</td>
<td>0.012</td>
<td>0.000</td>
<td>0.012</td>
<td>0.002</td>
<td>0.007</td>
<td>0.069</td>
<td>0.069</td>
<td>0.069</td>
<td>0.284</td>
<td></td>
</tr>
<tr>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.002</td>
<td>300</td>
<td>0.000</td>
<td>0.003</td>
<td>0.000</td>
<td>0.004</td>
<td>0.000</td>
<td>0.004</td>
<td>0.002</td>
<td>0.002</td>
<td>0.049</td>
<td>0.036</td>
<td>0.040</td>
<td>0.847</td>
<td></td>
</tr>
<tr>
<td>0.021</td>
<td>0.021</td>
<td>0.021</td>
<td>-0.005</td>
<td>30</td>
<td>0.021</td>
<td>0.025</td>
<td>0.021</td>
<td>0.025</td>
<td>0.021</td>
<td>0.024</td>
<td>-0.004</td>
<td>0.005</td>
<td>0.680</td>
<td>0.522</td>
<td>0.523</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>0.021</td>
<td>0.021</td>
<td>0.021</td>
<td>0.000</td>
<td>300</td>
<td>0.021</td>
<td>0.008</td>
<td>0.021</td>
<td>0.008</td>
<td>0.021</td>
<td>0.008</td>
<td>-0.005</td>
<td>0.002</td>
<td>0.995</td>
<td>0.983</td>
<td>0.982</td>
<td>0.025</td>
<td></td>
</tr>
<tr>
<td>0.021</td>
<td>0.021</td>
<td>0.021</td>
<td>0.019</td>
<td>30</td>
<td>0.021</td>
<td>0.025</td>
<td>0.021</td>
<td>0.025</td>
<td>0.021</td>
<td>0.025</td>
<td>0.000</td>
<td>0.010</td>
<td>0.682</td>
<td>0.539</td>
<td>0.532</td>
<td>0.056</td>
<td></td>
</tr>
<tr>
<td>0.021</td>
<td>0.021</td>
<td>0.021</td>
<td>0.019</td>
<td>300</td>
<td>0.021</td>
<td>0.008</td>
<td>0.021</td>
<td>0.008</td>
<td>0.021</td>
<td>0.008</td>
<td>0.000</td>
<td>0.003</td>
<td>0.995</td>
<td>0.981</td>
<td>0.981</td>
<td>0.036</td>
<td></td>
</tr>
</tbody>
</table>

$H_k = 0.05; H_k = 0.05; H_k = 0.1$

<table>
<thead>
<tr>
<th>$\omega_{12}$</th>
<th>$\omega_{13}$</th>
<th>$\omega_{23}$</th>
<th>$\omega_{123}$</th>
<th>$n$</th>
<th>Mean</th>
<th>SD</th>
<th>Mean</th>
<th>SD</th>
<th>Mean</th>
<th>SD</th>
<th>Mean</th>
<th>SD</th>
<th>$\hat{p}(\chi^2_{1})$</th>
<th>$\hat{p}(\chi^2_{1})$</th>
<th>$\hat{p}(\chi^2_{2})$</th>
<th>$\hat{p}(\chi^2_{3})$</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.015</td>
<td>30</td>
<td>0.000</td>
<td>0.025</td>
<td>0.000</td>
<td>0.027</td>
<td>-0.001</td>
<td>0.042</td>
<td>0.014</td>
<td>0.012</td>
<td>0.038</td>
<td>0.030</td>
<td>0.057</td>
<td>0.161</td>
<td></td>
</tr>
<tr>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.015</td>
<td>300</td>
<td>0.000</td>
<td>0.008</td>
<td>0.000</td>
<td>0.009</td>
<td>0.000</td>
<td>0.013</td>
<td>0.015</td>
<td>0.004</td>
<td>0.049</td>
<td>0.051</td>
<td>0.052</td>
<td>0.992</td>
<td></td>
</tr>
<tr>
<td>0.023</td>
<td>0.026</td>
<td>0.103</td>
<td>-0.032</td>
<td>30</td>
<td>0.023</td>
<td>0.029</td>
<td>0.012</td>
<td>0.027</td>
<td>0.099</td>
<td>0.038</td>
<td>-0.025</td>
<td>0.017</td>
<td>0.188</td>
<td>0.052</td>
<td>0.738</td>
<td>0.552</td>
<td></td>
</tr>
<tr>
<td>0.023</td>
<td>0.026</td>
<td>0.103</td>
<td>0.000</td>
<td>300</td>
<td>0.024</td>
<td>0.009</td>
<td>0.012</td>
<td>0.009</td>
<td>0.103</td>
<td>0.012</td>
<td>-0.027</td>
<td>0.006</td>
<td>0.839</td>
<td>0.289</td>
<td>1.000</td>
<td>0.999</td>
<td></td>
</tr>
<tr>
<td>0.023</td>
<td>0.026</td>
<td>0.103</td>
<td>0.015</td>
<td>30</td>
<td>0.024</td>
<td>0.023</td>
<td>0.016</td>
<td>0.026</td>
<td>0.096</td>
<td>0.037</td>
<td>0.012</td>
<td>0.011</td>
<td>0.197</td>
<td>0.061</td>
<td>0.695</td>
<td>0.022</td>
<td></td>
</tr>
<tr>
<td>0.023</td>
<td>0.026</td>
<td>0.103</td>
<td>0.015</td>
<td>300</td>
<td>0.025</td>
<td>0.009</td>
<td>0.017</td>
<td>0.008</td>
<td>0.099</td>
<td>0.012</td>
<td>0.013</td>
<td>0.003</td>
<td>0.872</td>
<td>0.525</td>
<td>1.000</td>
<td>0.935</td>
<td></td>
</tr>
</tbody>
</table>
Figure 1.—Power to detect three-locus zygotic associations with samples of sizes \( n = 30 \) (●), \( n = 100 \) (○), \( n = 300 \) (●), \( n = 500 \) (○), and \( n = 1000 \) (▲). for two cases: (A) when heterozygosities at three loci are \( H_1 = H_2 = 0.05 \) and \( H_3 = 0.1 \), and three pairwise zygotic associations are \( \omega_{12} = \omega_{13} = \omega_{23} = 0.0213 \); (B) when heterozygosities at three loci are \( H_1 = 0.1 \), \( H_2 = 0.3 \), and \( H_3 = 0.5 \), and three pairwise zygotic associations are \( \omega_{12} = 0.0226 \), \( \omega_{13} = 0.0263 \), and \( \omega_{23} = 0.00 \).
individuals without distinguishing among different homozygotes in large and predominantly outcrossing populations (Mitton 1997). The assessment would be most sensitive with quantitative trait loci (QTL) that directly affect components of fitness. However, a lack of zygotic associations may also mean that selection discriminates among different homozygotes (e.g., favoring common homozygotes, but selecting against rare homozygotes). Thus, extra care is needed to choose homozygous QTL with similar selection advantages for such an analysis.

There are a variety of methods of estimating and interpreting multilocus gametic disequilibria from haploid data or diploid data from a Hardy-Weinberg equilibrium population (e.g., Bennett 1954; Brown et al. 1980; Barton 2000). In contrast, with the diploid data from a Hardy-Weinberg disequilibrium population, a complete characterization of multilocus associations also requires the exact test, the probability of the observed multilocus zygotic association is evaluated to determine whether or not the study population is in Hardy-Weinberg equilibrium. The estimation and hypothesis testing are quite straightforward as they are merely the direct adoption of the procedures used for diallelic haploid data. Thus, our method presents a simple solution to the analysis of complex multilocus structures in diploid populations.

Of course, such a highly compacted summary in the multilocus zygotic associations represents a severe loss of information. In particular, since the analysis is based on the frequencies of zygote classes, it completely ignores haplotype information such as linkages between different loci. Thus, when significant zygotic associations are detected, there is a need to determine which genic disequilibria are important. In light of great current interest in the linkage (gametic) disequilibrium approach to fine-scale QTL mapping (e.g., Pritchard and Przeworski 2001; Reich et al. 2001), it is essential to determine if gametic disequilibrium is important in the presence of significant zygotic associations. As shown earlier, if the study population is in Hardy-Weinberg equilibrium, then there are direct relationships between zygotic associations and various orders of gametic disequilibria (cf. Equations 4b and 11). In this case, it is definitely more informative to work directly with the raw genotypic data instead of the collapsed data based on zygote classes so that haplotype frequencies and gametic disequilibrium can be inferred. However, in the presence of Hardy-Weinberg disequilibrium, which may often be the case in natural populations, gametic disequilibrium may be inflated because many other types of nonallelic disequilibria may also cause the multilocus associations. The knowledge about the inflation may be gained through the comparative assessment of gametic vs. zygotic associations mentioned above.

In estimating and testing for multilocus zygotic associations, we adopt Bennett’s (1954) additive approach, with frequencies of different zygote classes being expressed as a linear function of the zygotic associations and heterozygosities (Table 3). This approach enables us to explicitly give estimates and to elucidate the sampling properties of these estimates. However, our tests for two- and three-locus associations are not independent as shown in the simulation results (Table 3). Hill (1975) discussed the use of the multiplicative approach (or log-linear model analysis) for developing an independent test for no three-locus association, but with the presence of two-locus associations. Another possibility is the exact test as suggested by Zaykin et al. (1995). In the exact test, the probability of the observed multilocus genotypic (zygotic) array conditional on the genotypic arrays expected under an appropriate hypothesis of zero zygotic association is evaluated to determine whether or not it lies in the tail of the empirical distribution generated by permutation. For example, the conditional probability required for testing if \( \omega_{21} = 0 \), given the presence of all three two-locus associations, is given by

\[
\Pr(n_{ab}, n_{ac}, n_{bc}) = \frac{\prod_{i=1}^{3} n_{i}^{n_{i}}}{n^{n_{abc}}} \prod_{i=1}^{3} \prod_{j=1}^{n_{i}} \prod_{k=1}^{n_{ij}} \binom{n_{ij}}{n_{ij}}
\]

where \( n_{ab}, n_{ac}, \) and \( n_{bc} \) are marginal total counts of the \( ab, ac, \) and \( bc \) classes of zygotes at locus pairs 12, 13, and 23, respectively. However, both log-linear model analysis and exact test do not allow for the explicit expression of the multilocus zygotic associations.

I thank Dr. Yun-Xin Fu and a reviewer for helpful comments. This research was partially supported by the Natural Sciences and Engineering Research Council of Canada grant OGP0183983.

**LITERATURE CITED**


Weir, B. S., 1996  Genetic Data Analysis II. Sinauer Associates, Sunderland, MA.


Communicating editor: Y.-X. Fu