Analysis of Multilocus Zygotic Associations

Rong-Cai Yang[1]

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

[1]*Address 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

$$f(X_j) = H_j^{X_j}(1 - H_j)^{1-X_j}, \tag{1}$$

where indicator X_j takes either 1 or 0 to signal whether the genotype at the jth 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 detecting 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

$$f(X_jX_l) = f(X_j)f(X_l) + (-1)^{X_j+X_l}\omega_{jl}, \tag{2}$$

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

$$\begin{aligned}\omega_{jl} &= f(00) - f(0\cdot)f(\cdot 0) = -[f(10) - f(1\cdot)f(\cdot 0)]\\ &= -[f(01) - f(0\cdot)f(\cdot 1)]\\ &= f(11) - f(1\cdot)f(\cdot 1) = f(00)f(11) - f(10)f(01).\end{aligned}$$

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

$$\begin{aligned}\max[-H_jH_l, -(1 - H_j)(1 - H_l)] \le \omega_{jl} < 0, &\quad \text{if } \omega_{jl} < 0\\ 0 < \omega_{jl} \le \min[(1 - H_j)H_l, H_j(1 - H_l)], &\quad \text{if } \omega_{jl} > 0.\end{aligned} \tag{3}$$

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

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

To see how the zygotic association (ω_{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,

$$\begin{aligned}f(00) &= \sum_{u=1}^{r}\sum_{y=1}^{s} {}^{jl}P_{uy}^{uy}, \quad f(01) = \sum_{u=1}^{r}\sum_{y\neq z}^{} {}^{jl}P_{uz}^{uy}\\ f(10) &= \sum_{u\neq v}^{}\sum_{y=1}^{s} {}^{jl}P_{vy}^{uy}, \quad f(11) = \sum_{u\neq v}^{}\sum_{y\neq z}^{} {}^{jl}P_{vz}^{uy},\end{aligned} \tag{4a}$$

where, for example, ${}^{jl}P_{vz}^{uy}$ is the frequency of genotypes at loci j and l from the union of gametes $j_u\ l_y$ and $j_v\ l_z$ ($u, v = 1, 2, \ldots, r$; $y, z = 1, 2, \ldots, s$). Then, using Cockerham and Weir's (1973) disequilibrium functions for the two-locus frequencies (*e.g.*, ${}^{jl}P_{uy}^{uy}$), the zygotic association at loci j and l can be expressed in terms of individual genic disequilibria,

$$\begin{aligned}\omega_{jl} &= f(00)f(11) - f(10)f(01)\\ &= \sum_{u=1}^{r}\sum_{y=1}^{s}\Big[2{}^{j}p_u{}^{jl}D_{\cdot y}^{uy} + 2{}^{l}p_y{}^{jl}D_{u\cdot}^{uy} + 2{}^{j}p_u{}^{l}p_y{}^{jl}D_{\cdot\cdot}^{uy}\\ &\qquad + 2{}^{j}p_u{}^{l}p_y{}^{jl}D_{u\cdot}^{\cdot y} + ({}^{jl}D_{\cdot\cdot}^{uy})^2 + ({}^{jl}D_{u\cdot}^{\cdot y})^2 + {}^{jl}D_{uy}^{uy}\Big],\end{aligned} \tag{4b}$$

where ${}^{j}p_u$, for example, 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 (${}^{jl}D_{\cdot\cdot}^{uy}$) is the deviation of frequency of gamete j_ul_y from the product of frequencies of alleles u and y at loci j and l, ${}^{jl}D_{\cdot\cdot}^{uy} = {}^{jl}P_{\cdot\cdot}^{uy} - {}^{j}p_u{}^{l}p_y$. It is also evident from (4b) that even in a gametic equilibrium population (${}^{jl}D_{\cdot\cdot}^{uy} = 0$), nonzero zygotic associations can arise from other forces such as partial inbreeding as a result of "identity disequilibrium" (${}^{jl}D_{vz}^{uy} \neq 0$). 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 heterozygosities 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, and X_o for loci j, l, and o is given by

$$\begin{aligned} f(X_jX_lX_o) &= f(X_j)\,f(X_l)\,f(X_o) + f(X_j)(-1)^{X_l+X_o}\omega_{lo} \\ &\quad + f(X_l)(-1)^{X_j+X_o}\omega_{jo} \\ &\quad + f(X_o)(-1)^{X_j+X_l}\omega_{jl} + (-1)^{X_j+X_l+X_o+1}\omega_{jlo}, \end{aligned} \tag{5}$$

where $f(X_j)$ and ω_{jl}, for example, are given in (1) and (2), and ω_{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.*,

$$\begin{aligned} f_0(X_jX_lX_o) &= f(X_j)\,f(X_l)\,f(X_o) + f(X_j)(-1)^{X_l+X_o}\omega_{lo} \\ &\quad + f(X_l)(-1)^{X_j+X_o}\omega_{jo} \\ &\quad + f(X_o)(-1)^{X_j+X_l}\omega_{jl}. \end{aligned} \tag{6}$$

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

$$\beta_1 = \min[\,f_0(111),\, f_0(100),\, f_0(010),\, f_0(001)]$$

$$\beta_2 = \min[\,f_0(110),\, f_0(101),\, f_0(011),\, f_0(000)],$$

where $f_0(111) = H_j\,H_l\,H_o + H_j\,\omega_{lo} + H_l\,\omega_{jo} + H_o\,\omega_{jl}$, 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 $\omega^+_{jlo}(\max)$ and $\omega^+_{jlo}(\min)$] are

$$\begin{aligned} \omega^+_{jlo}(\max) &= \beta_2 \\ \omega^+_{jlo}(\min) &= \max[0, -\beta_1] \end{aligned} \tag{7a}$$

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

$$\begin{aligned} \omega^-_{jlo}(\max) &= -\beta_1 \\ \omega^-_{jlo}(\min) &= \max[0, -\beta_2]. \end{aligned} \tag{7b}$$

Since any of the eight $f_0(X_jX_lX_o)$ values can be negative, neither $\omega^+_{jlo}(\min)$ nor $\omega^-_{jlo}(\min)$ is necessarily zero. Clearly, the β_1 and β_2 values can be used to determine the sign and range of ω_{jlo},

$$\begin{aligned} &\omega^+_{jlo}(\min) \le \omega_{jlo} \le \omega^+_{jlo}(\max), \quad \text{if } \beta_1 < 0,\ \beta_2 > 0 \text{ and } |\beta_1| \le |\beta_2| \\ &\omega^-_{jlo}(\max) \le \omega_{jlo} \le \omega^-_{jlo}(\min), \quad \text{if } \beta_1 > 0,\ \beta_2 < 0 \text{ and } |\beta_1| \ge |\beta_2| \\ &\omega^-_{jlo}(\max) \le \omega_{jlo} \le \omega^+_{jlo}(\max), \quad \text{if } \beta_1 \ge 0 \text{ and } \beta_2 \ge 0, \end{aligned} \tag{8}$$

but some pairs of the β_1 and β_2 values (*e.g.*, $\beta_1 < 0$ and $\beta_2 < 0$) would not lead to the definable ω_{jlo}. Given $0 \le H_j, H_l, H_o \le 1$, all profiles of heterozygosities $\{H_j\,H_l\,H_o\}$, but with no two-locus associations (*i.e.*, $\omega_{jl} = \omega_{jo} = \omega_{lo} = 0$), would produce the condition of $\beta_1 \ge 0$ and $\beta_2 \ge 0$ [*i.e.*, $f_0(X_jX_lX_o) = f(X_j)f(X_l)f(X_o) \ge 0$] and thus definable ω_{jlo} values. However, given a heterozygosity profile, not all configurations of two-locus zygotic associations $\{\omega_{jl}\ \omega_{jo}\ \omega_{lo}\}$ would lead to the conditions given in (8) under which ω_{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 (ω_{jlo}). For example, with $H_j = H_l = 0.05$ and $H_o = 0.1$, the ranges for ω_{jl}, ω_{jo}, and ω_{lo} are $-0.0025 \le \omega_{jl} \le 0.0475$, $-0.005 \le \omega_{jo} \le 0.045$, and $-0.005 \le \omega_{lo} \le 0.045$, 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 ω_{jl}'s $\times$ 20 ω_{jo}'s $\times$ 20 ω_{lo}'s) configurations of the two-locus zygotic associations that can be used to define the ranges for ω_{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 $\omega^-_{jlo}(\max) \le \omega_{jlo} \le \omega^+_{jlo}(\max)$, but the remaining 4106 configurations do not lead to any definable ω_{jlo}. In each of these three cases, we identify a configuration that leads to the maximum range of ω_{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 (ω_{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 J and j at locus j are p_J and p_j ($=1 - p_J$), those of the two alleles L and l at locus l are p_L and p_l ($=1 - p_L$), and those of the two alleles O and o at locus o are

TABLE 1

Ranges for three-locus zygotic associations

H_j	H_l	H_o	Bounds on ω_{jlo}	v^a	Example ω_{jl}	ω_{jo}	ω_{lo}	Range for ω_{jlo}	
0.1	0.05	0.1	$\omega_{jlo} < 0$	240	0.0396	−0.0024	0.0003	−0.0041	−0.0015
			$\omega_{jlo} > 0$	1496	0.0238	0.0239	0.0239	0.0002	0.0213
			$\omega_{jlo}^-(\max) \le \omega_{jlo} \le \omega_{jlo}^+(\max)$	2158	0.0212	0.0213	0.0213	−0.0045	0.0192
					0	0	0	−0.0003	0.0023
0.1	0.1	0.1	$\omega_{jlo} < 0$	294	−0.0047	0.0742	0.0058	−0.0085	−0.0033
			$\omega_{jlo} > 0$	933	0.0479	0.0479	0.0479	0.0004	0.0425
			$\omega_{jlo}^-(\max) \le \omega_{jlo} \le \omega_{jlo}^+(\max)$	1698	0.0426	0.0426	0.0426	−0.0085	0.0388
					0	0	0	−0.0010	0.0090
0.1	0.1	0.5	$\omega_{jlo} < 0$	805	0.0584	−0.0237	−0.0237	−0.0295	−0.0032
			$\omega_{jlo} > 0$	805	0.0584	0.0237	0.0237	0.0032	0.0295
			$\omega_{jlo}^-(\max) \le \omega_{jlo} \le \omega_{jlo}^+(\max)$	2288	0.0426	0.0026	0.0026	−0.0216	0.0258
					0	0	0	−0.0050	0.0050
0.1	0.3	0.5	$\omega_{jlo} < 0$	1246	0.0384	−0.0132	0.0868	−0.0337	−0.0021
			$\omega_{jlo} > 0$	1284	0.0384	0.0184	−0.0395	0.0011	0.0326
			$\omega_{jlo}^-(\max) \le \omega_{jlo} \le \omega_{jlo}^+(\max)$	2761	0.0226	0.0263	0.1026	−0.0321	0.0153
					0	0	0	−0.0150	0.0150
0.5	0.5	0.5	$\omega_{jlo} < 0$	0					
			$\omega_{jlo} > 0$	0					
			$\omega_{jlo}^-(\max) \le \omega_{jlo} \le \omega_{jlo}^+(\max)$	2440	0.0132	0.0132	0.0132	−0.1184	0.1184
					0	0	0	−0.1250	0.1250

Shown are configurations of two-locus zygotic associations (ω_{jl}) that lead to definable three-locus associations and examples for resultant maximum ranges for three-locus zygotic association (ω_{jlo}) for each heterozygosity profile (H_j, H_l, and H_o).

[a] The number of ω_{jl} configurations that satisfy the bound set for ω_{123} as specified in column 4. The total number of ω_{jl} configurations is 8000 (20 $\omega_{jl} \times$ 20 $\omega_{jo} \times$ 20 ω_{lo}).

p_O and p_o (=1 − p_O). The gametic disequilibrium between the *j*-*l* loci is denoted as D_{jl}, between the *j*-*o* loci as D_{jo}, between the *l*-*o* loci as D_{lo}, and the three-locus gametic disequilibrium as D_{jlo}. These seven parameters are used to obtain the vector of eight three-locus gametic frequencies, $\mathbf{g} = [g_{JLO}\ g_{JLo}\ g_{JlO}\ g_{Jlo}\ g_{jLO}\ g_{jLo}\ g_{jlO}\ g_{jlo}]'$, where g_{JLO}, for example, is

$$g_{JLO} = p_J p_L p_O + p_J D_{lo} + p_L D_{jo} + p_O D_{jl} + D_{jlo}$$

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

$$\mathbf{G} = \begin{bmatrix} g_{JLO} & g_{JLo} & g_{JlO} & g_{Jlo} & g_{jLO} & g_{jLo} & g_{jlO} & g_{jlo} \\ g_{JLo} & g_{JLO} & g_{Jlo} & g_{JlO} & g_{jLo} & g_{jLO} & g_{jlo} & g_{jlO} \\ g_{JlO} & g_{Jlo} & g_{JLO} & g_{JLo} & g_{jlO} & g_{jlo} & g_{jLO} & g_{jLo} \\ g_{Jlo} & g_{JlO} & g_{JLo} & g_{JLO} & g_{jlo} & g_{jlO} & g_{jLo} & g_{jLO} \\ g_{jLO} & g_{jLo} & g_{jlO} & g_{jlo} & g_{JLO} & g_{JLo} & g_{JlO} & g_{Jlo} \\ g_{jLo} & g_{jLO} & g_{jlo} & g_{jlO} & g_{JLo} & g_{JLO} & g_{Jlo} & g_{JlO} \\ g_{jlO} & g_{jlo} & g_{jLO} & g_{jLo} & g_{JlO} & g_{Jlo} & g_{JLO} & g_{JLo} \\ g_{jlo} & g_{jlO} & g_{jLo} & g_{jLO} & g_{Jlo} & g_{JlO} & g_{JLo} & g_{JLO} \end{bmatrix}. \quad (9)$$

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.*, $D_{jl} = D_{jo} = D_{lo} = 0$), 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

$$\begin{aligned} f(000) &= (p_J p_L p_O + D_{jlo})^2 + (p_J p_L p_O - D_{jlo})^2 \\ &\quad + (p_J p_l p_O - D_{jlo})^2 + (p_J p_l p_o + D_{jlo})^2 \\ &\quad + (p_j p_L p_O - D_{jlo})^2 + (p_j p_L p_o + D_{jlo})^2 \\ &\quad + (p_j p_l p_O + D_{jlo})^2 + (p_j p_l p_o - D_{jlo})^2 \\ &= (1 - H_j)(1 - H_l)(1 - H_o) \\ &\quad - 2(1 - 2p_j)(1 - 2p_l)(1 - 2p_O)D_{jlo} + 8D_{jlo}^2. \end{aligned} \quad (10)$$

Here H_j, for example, is the same as the expected heterozygosity under Hardy-Weinberg equilibrium ($h_j = 2p_J\ p_j$). Given that $\omega_{jl} = 2(1 - 2p_j)(1 - 2p_L)D_{jl} + 4D_{jl}^2$ (Yang 2000), for example, $\omega_{jl} = \omega_{jl} = \omega_{jl} = 0$ 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

$$\begin{aligned} \omega_{jlo} &= -[f(000) - (1 - H_j)(1 - H_l)(1 - H_o)] \\ &= 2(1 - 2p_j)(1 - 2p_L)(1 - 2p_O)D_{jlo} - 8D_{jlo}^2 \end{aligned} \quad (11)$$

TABLE 2

Three-locus gametic and zygotic associations

					Value of ω_{jlo}			
			Range of D_{jlo}		Size of negative D_{jlo}		Size of positive D_{jlo}	
p_J	p_L	p_O	$-p_J p_L p_O$	$p_J p_L(1 - p_O)$	100%	50%	50%	100%
0.1	0.1	0.1	−0.001	0.009	−0.0010	−0.0005	0.0044	0.0086
0.1	0.1	0.2	−0.002	0.008	−0.0016	−0.0008	0.0029	0.0056
0.1	0.1	0.3	−0.003	0.007	−0.0016	−0.0008	0.0017	0.0032
0.1	0.1	0.4	−0.004	0.006	−0.0012	−0.0005	0.0007	0.0012
0.1	0.1	0.5	−0.005	0.005	−0.0002	−0.0001	−0.0001	−0.0002
0.1	0.2	0.2	−0.004	0.016	−0.0024	−0.0012	0.0041	0.0072
0.1	0.2	0.3	−0.006	0.014	−0.0026	−0.0012	0.0023	0.0038
0.1	0.2	0.4	−0.008	0.012	−0.0020	−0.0009	0.0009	0.0012
0.1	0.2	0.5	−0.010	0.010	−0.0008	−0.0002	−0.0002	−0.0008
0.1	0.3	0.3	−0.009	0.021	−0.0030	−0.0013	0.0018	0.0018
0.1	0.3	0.4	−0.012	0.018	−0.0027	−0.0011	0.0005	−0.0003
0.1	0.3	0.5	−0.015	0.015	−0.0018	−0.0005	−0.0005	−0.0018
0.1	0.4	0.4	−0.016	0.024	−0.0031	−0.0010	−0.0004	−0.0031
0.1	0.4	0.5	−0.020	0.020	−0.0032	−0.0008	−0.0008	−0.0032
0.1	0.5	0.5	−0.025	0.025	−0.0050	−0.0013	−0.0013	−0.0050
0.2	0.2	0.2	−0.008	0.032	−0.0040	−0.0019	0.0049	0.0056
0.2	0.2	0.3	−0.012	0.028	−0.0046	−0.0020	0.0025	0.0018
0.2	0.2	0.4	−0.016	0.024	−0.0044	−0.0017	0.0006	−0.0012
0.2	0.2	0.5	−0.020	0.020	−0.0032	−0.0008	−0.0008	−0.0032
0.2	0.3	0.3	−0.018	0.042	−0.0060	−0.0024	0.0005	−0.0060
0.2	0.3	0.4	−0.024	0.036	−0.0069	−0.0023	−0.0009	−0.0069
0.2	0.3	0.5	−0.030	0.030	−0.0072	−0.0018	−0.0018	−0.0072
0.2	0.4	0.4	−0.032	0.048	−0.0097	−0.0028	−0.0035	−0.0161
0.2	0.4	0.5	−0.040	0.040	−0.0128	−0.0032	−0.0032	−0.0128
0.2	0.5	0.5	−0.050	0.050	−0.0200	−0.0050	−0.0050	−0.0200
0.3	0.3	0.3	−0.027	0.063	−0.0093	−0.0032	−0.0039	−0.0237
0.3	0.3	0.4	−0.036	0.054	−0.0127	−0.0037	−0.0041	−0.0199
0.3	0.3	0.5	−0.045	0.045	−0.0162	−0.0041	−0.0041	−0.0162
0.3	0.4	0.4	−0.048	0.072	−0.0200	−0.0054	−0.0092	−0.0392
0.3	0.4	0.5	−0.060	0.060	−0.0288	−0.0072	−0.0072	−0.0288
0.3	0.5	0.5	−0.075	0.075	−0.0450	−0.0113	−0.0113	−0.0450
0.4	0.4	0.4	−0.064	0.096	−0.0338	−0.0087	−0.0177	−0.0722
0.4	0.4	0.5	−0.080	0.080	−0.0512	−0.0128	−0.0128	−0.0512
0.4	0.5	0.5	−0.100	0.100	−0.0800	−0.0200	−0.0200	−0.0800
0.5	0.5	0.5	−0.125	0.125	−0.1250	−0.0313	−0.0313	−0.1250

Shown are three-locus zygotic associations (ω_{jlo}) in the presence of three-locus gametic disequilibrium (D_{jlo}) but absence of the pairwise disequilibria ($D_{jl} = D_{jo} = D_{lo} = 0$). The Hardy-Weinberg equilibrium population is assumed so that the values of ω_{jlo} are directly related to D_{jlo} and gene frequencies (p_J, p_L, and p_O).

(*cf.* Equation 5). Table 2 lists the values of three-locus zygotic association (ω_{jlo}) in the presence of three-locus gametic disequilibrium (D_{jlo}) but absence of the pairwise disequilibria ($D_{jl} = D_{jo} = D_{lo} = 0$) for various gene frequencies ($p_J \leq p_L \leq p_O$). In this case, the range of D_{jlo} is defined by the gene frequencies $-p_J\, p_L\, p_O \leq D_{jlo} \leq p_J\, p_L(1 - p_O)$. The values of ω_{jlo} are calculated for two sizes of negative D_{jlo} ($-p_J\, p_L\, p_O$ and $-0.5 p_J\, p_L\, p_O$) and two sizes of positive D_{jlo} [$p_J\, p_L(1 - p_O)$ and $0.5 p_J\, p_L(1 - p_O)$]. The higher the absolute values that D_{jlo} can take, the larger the absolute values of ω_{jlo}. It is evident from (11) that ω_{jlo} is always negative if $D_{jlo} < 0$, but can be either positive or negative if $D_{jlo} > 0$ with ω_{jlo} being positive only if $0 < D_{jlo} < (1 - 2p_J)(1 - 2p_L)(1 - 2p_O)/4$. As p_J, p_L, and p_O are approaching 0.5, the range of D_{jlo} is expanded and the absolute values of ω_{jlo} are increased. In the cases of $p_J = 0.5$ or $p_L = 0.5$ or $p_O = 0.5$, ω_{jlo} is always negative because $\omega_{jlo} = -8D^2_{jlo}$.

The combined effect of two- and three-locus gametic disequilibria on the values of ω_{jlo} is also examined (numerical results are not presented). The joint contribution of two- and three-locus gametic disequilibria to ω_{jlo} greatly cloaks their relationships with ω_{jlo}. However, there are clearly cases where the ω_{jlo} values exceed the limits of ω_{jlo} under the cases of no pairwise disequilibria. For example, for $p_J = p_L = p_O = 0.5$, the ranges for D_{jl}, D_{jo}, and D_{lo} 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

when $D_{jl} = D_{jo} = D_{lo} = D_{jlo} = -0.25$, ω_{jlo} is -0.5, which exceeds the limit of -0.125 in the case of no two-locus disequilibria ($D_{jl} = D_{jo} = D_{lo} = 0$).

More than three loci: The extension to four or more loci following Bennett (1954) is straightforward. For example, the joint distribution of indicators X_j, X_l, X_o, and X_q for loci j, l, o, and q, respectively, is given by

$$\begin{aligned} f(X_jX_lX_oX_q) = {} & f(X_j)f(X_l)f(X_o)f(X_q) + f(X_j)(-1)^{X_l+X_o+X_q+1}\omega_{loq} \\ & + f(X_l)(-1)^{X_j+X_o+q+1}\omega_{joq} + f(X_o)(-1)^{X_jX_lX_q+1}\omega_{jlq} \\ & + f(X_q)(-1)^{X_jX_lX_o+1}\omega_{jlo} + f(X_j)f(\mathrm{X}_l)(-1)^{X_o+X_q}\omega_{oq} \\ & + f(X_j)f(X_o)(-1)^{X_l+X_q}\omega_{lq} + f(X_j)f(X_q)(-1)^{X_l+X_o}\omega_{lo} \\ & + f(X_l)f(X_o)(-1)^{X_j+X_q}\omega_{jq} + f(X_l)f(X_q)(-1)^{X_j+X_o}\omega_{jo} \\ & + f(X_o)f(X_q)(-1)^{X_j+X_l}\omega_{jl} \\ & + (-1)^{X_j+X_l+X_o+X_q}[\omega_{jl}\omega_{oq} + \omega_{jo}\omega_{lq} + \omega_{jq}\omega_{lo} + \omega_{jloq}], \end{aligned} \tag{12}$$

where $f(X_j)$, ω_{jl}, and ω_{jlo}, for example, are given in (1), (2), and (5), and ω_{jloq} 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; Cockerham and Tachida 1986). A set of functions, $f_0(X_jX_lX_oX_q)$, can be defined in a similar manner as (6) for $f_0(X_jX_lX_o)$ to provide the basis for defining the range of ω_{jloq}. 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_1, H_2, and H_3), three two-locus zygotic associations (ω_{12}, ω_{13}, and ω_{23}), and one three-locus zygotic association (ω_{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_{abc} be the numbers of the abcth class of zygotes with a, b, and c representing indicators X_1, X_2, and X_3, respectively. Thus the ML estimates of f's are given by $\hat{f}(abc) = n_{abc}/n$ to satisfy the maximized multinomial likelihood,

$$\max L = \frac{n!}{\prod_{a=0}^{1}\prod_{b=0}^{1}\prod_{c=0}^{1} n_{abc}!}\prod_{a=0}^{1}\prod_{b=0}^{1}\prod_{c=0}^{1}[\hat{f}(abc)]^{n_{abc}}.$$

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, $\hat{f}(ab\cdot) = \Sigma_{c=0}^{1}\hat{f}(abc)$ and $\hat{f}(a\cdot\cdot) = \Sigma_{b=0}^{1}\Sigma_{c=0}^{1}\hat{f}(abc)$. Note that the one-locus marginal frequencies, $\hat{f}(1\cdot\cdot) = \hat{H}_1$, $\hat{f}(\cdot1\cdot) = \hat{H}_2$, and $\hat{f}(\cdot\cdot1) = \hat{H}_3$, are the estimates of heterozygosities at loci 1, 2, and 3, respectively. The zygotic associations for two loci (*e.g.*, ω_{12}) and for all three loci (ω_{123}) are estimated as

$$\hat{\omega}_{12} = \hat{f}(11\cdot) - \hat{f}(1\cdot\cdot)f(\cdot1\cdot) \tag{13a}$$

and

$$\begin{aligned} \hat{\omega}_{123} = {} & \hat{f}(111) - \hat{f}(1\cdot\cdot)\hat{\omega}_{23} - \hat{f}(\cdot1\cdot)\hat{\omega}_{13} \\ & - \hat{f}(\cdot\cdot1)\hat{\omega}_{12} - \hat{f}(1\cdot\cdot)\hat{f}(\cdot1\cdot)\hat{f}(\cdot\cdot1), \end{aligned} \tag{13b}$$

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

$$E(\hat{\omega}_{12}) = \frac{(n-1)}{n}\omega_{12} \quad \text{and} \quad E(\hat{\omega}_{123}) = \frac{(n-3)}{n}\omega_{123}.$$

Sampling variances of linear combinations of multinomial variables are known exactly. For example, $\mathrm{var}(\hat{H}_1) = H_1(1 - H_1)/n$. 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_{abc}, for example, with expectations $E(n_{abc}) = nf(abc)$. The sampling variances of $\hat{\omega}_{12}$ and $\hat{\omega}_{123}$ are

$$\mathrm{Var}(\hat{\omega}_{12}) = \frac{1}{n}[A_1A_2 + B_1B_2\omega_{12} - \omega_{12}^2] \tag{14a}$$

and

$$\begin{aligned} \mathrm{Var}(\hat{\omega}_{123}) = {} & \frac{1}{n}[A_1A_2A_3 + 6\omega_{12}\omega_{13}\omega_{23} + A_1(B_2B_3\omega_{23} - \omega_{23}^2) \\ & + A_2(B_1B_3\omega_{13} - \omega_{13}^2) + A_3(B_1B_2\omega_{12} - \omega_{12}^2) \\ & + \omega_{123}(B_1B_2B_3 - 2B_1\omega_{23} - 2B_2\omega_{13} - 2B_3\omega_{12}) - \omega_{123}^2], \end{aligned} \tag{14b}$$

where $A_t = H_t(1 - H_t)$ and $B_t = (1 - 2H_t)$ with $t = 1$, 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.*, $\hat{\omega}_i \sim N[E(\hat{\omega}_i), \mathrm{Var}(\hat{\omega}_i)]$, a test statistic ($\chi_i^2$) that is constructed, after setting ω_i to zero in both $E(\hat{\omega}_i)$ and $\mathrm{Var}(\hat{\omega}_i)$, is distributed as chi

TABLE 3

Joint frequencies of zygotes at three loci

X_1	X_2	X_3	Frequency
1	1	1	$f(111) = H_1H_2H_3 + H_1\omega_{23} + H_2\omega_{13} + H_3\omega_{12} + \omega_{123}$
1	1	0	$f(110) = H_1H_2(1 - H_3) - H_1\omega_{23} - H_2\omega_{13} + (1 - H_3)\omega_{12} - \omega_{123}$
1	0	1	$f(101) = H_1(1 - H_2)H_3 - H_1\omega_{23} + (1 - H_2)\omega_{13} - H_3\omega_{12} - \omega_{123}$
1	0	0	$f(100) = H_1(1 - H_2)(1 - H_3) + H_1\omega_{23} - (1 - H_2)\omega_{13} - (1 - H_3)\omega_{12} + \omega_{123}$
0	1	1	$f(011) = (1 - H_1)H_2H_3 + (1 - H_1)\omega_{23} - H_2\omega_{13} - H_3\omega_{12} - \omega_{123}$
0	1	0	$f(010) = (1 - H_1)H_2(1 - H_3) - (1 - H_1)\omega_{23} + H_2\omega_{13} - (1 - H_3)\omega_{12} + \omega_{123}$
0	0	1	$f(001) = (1 - H_1)(1 - H_2)H_3 - (1 - H_1)\omega_{23} - (1 - H_2)\omega_{13} + H_3\omega_{12} + \omega_{123}$
0	0	0	$f(000) = (1 - H_1)(1 - H_2)(1 - H_3) + (1 - H_1)\omega_{23} + (1 - H_2)\omega_{13} + (1 - H_3)\omega_{12} - \omega_{123}$

Shown are the joint frequency distributions of indicator variables, X_1, X_2, and X_3 in terms of their heterozygosities (H_1, H_2, and H_3) and zygotic associations (ω_{12}, ω_{13}, ω_{23}, and ω_{123}) at loci 1, 2, and 3.

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 ($\hat{\omega}_{12}$),

$$\chi^2_{12} = \frac{\hat{\omega}^2_{12}}{\mathrm{Var}(\hat{\omega}_{12}|\omega_{12} = 0)} = \frac{n\hat{\omega}^2_{12}}{\hat{H}_1(1 - \hat{H}_1)\,\hat{H}_2(1 - \hat{H}_2)},$$

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, ω_{12}, ω_{13}, ω_{23}, and ω_{123}. The eight frequencies of zygote classes, $f(X_1X_2X_3)$, can be constructed from given values of the four zygotic associations and three heterozygosities, H_1, H_2, 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 (ω_{123}). Thus, there are a total of 54 populations constructed. From each population, 10,000 replicate samples of sizes $n = 30$, 100, and 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 $H_1 = 0.1$, $H_2 = 0.3$, 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 ω_{12}, ω_{13}, ω_{23}, and ω_{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 $\hat{\omega}_{12}$ and $\hat{\omega}_{23}$ are almost identical to their theoretical values, $\hat{\omega}_{13}$ is only less than one-half of its true value and $\hat{\omega}_{23}$ is also a downwardly biased estimate of ω_{123}. However, when ω_{123} is set to zero, the estimates of all three two-locus zygotic associations are unbiased; conversely $\hat{\omega}_{123}$ is also an unbiased estimate of ω_{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_0, $\omega_{123} = 0$, is

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

Thus, the power (the probability of rejecting the false H_0) is given by

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

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 gametic 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 (σ^2_K) is detectable in a sample of moderate size (≥30). However, the magnitude of such association in σ^2_K may be appreciably larger than an individual association examined here

TABLE 4

Powers of detecting three-locus zygotic associations

					$\hat{\omega}_{12}$		$\hat{\omega}_{13}$		$\hat{\omega}_{23}$		$\hat{\omega}_{123}$		Power			
ω_{12}	ω_{13}	ω_{23}	ω_{123}	n	Mean	SD	Mean	SD	Mean	SD	Mean	SD	$P(\chi^2_{12})$	$P(\chi^2_{13})$	$P(\chi^2_{23})$	$P(\chi^2_{123})$
							$H_1 = 0.05$; $H_2 = 0.05$; $H_3 = 0.1$									
0.000	0.000	0.000	0.002	30	0.000	0.009	0.000	0.012	0.000	0.012	0.002	0.007	0.069	0.069	0.069	0.284
				300	0.000	0.003	0.000	0.004	0.000	0.004	0.002	0.002	0.049	0.036	0.040	0.847
0.021	0.021	0.021	−0.005	30	0.021	0.025	0.021	0.025	0.021	0.024	−0.004	0.005	0.680	0.522	0.523	0.001
				300	0.021	0.008	0.021	0.008	0.021	0.008	−0.005	0.002	0.995	0.983	0.982	0.025
0.021	0.021	0.021	0.000	30	0.021	0.025	0.021	0.025	0.021	0.025	0.000	0.010	0.682	0.539	0.532	0.056
				300	0.021	0.008	0.021	0.008	0.021	0.008	0.000	0.003	0.995	0.981	0.981	0.036
0.021	0.021	0.021	0.019	30	0.021	0.025	0.021	0.025	0.021	0.025	0.018	0.019	0.693	0.529	0.545	0.767
				300	0.021	0.008	0.021	0.008	0.021	0.008	0.019	0.007	0.995	0.982	0.982	0.999
							$H_1 = 0.1$; $H_2 = 0.3$; $H_3 = 0.5$									
0.000	0.000	0.000	0.015	30	0.000	0.025	0.000	0.027	−0.001	0.042	0.014	0.012	0.038	0.030	0.057	0.161
				300	0.000	0.008	0.000	0.009	0.000	0.013	0.015	0.004	0.049	0.051	0.052	0.992
0.023	0.026	0.103	−0.032	30	0.023	0.029	0.012	0.027	0.099	0.038	−0.025	0.017	0.188	0.052	0.738	0.552
				300	0.024	0.009	0.012	0.009	0.103	0.012	−0.027	0.006	0.839	0.289	1.000	0.999
0.023	0.026	0.103	0.000	30	0.022	0.029	0.025	0.027	0.099	0.038	0.000	0.013	0.175	0.110	0.723	0.081
				300	0.023	0.009	0.026	0.009	0.102	0.012	0.000	0.004	0.790	0.882	1.000	0.053
0.023	0.026	0.103	0.015	30	0.024	0.028	0.016	0.026	0.096	0.037	0.012	0.011	0.197	0.061	0.695	0.022
				300	0.025	0.009	0.017	0.008	0.099	0.012	0.013	0.003	0.872	0.525	1.000	0.935

Shown are the means and standard deviations (SD) of estimates of zygotic associations (ω_{12}, ω_{13}, ω_{23}, and ω_{123}) from 10,000 samples of sizes $n = 30$ and $n = 300$. The power is estimated as proportion of times that the chi-square statistic exceeds 3.84, the 5% critical value of $\chi^2_{\text{d.f.}=1}$.

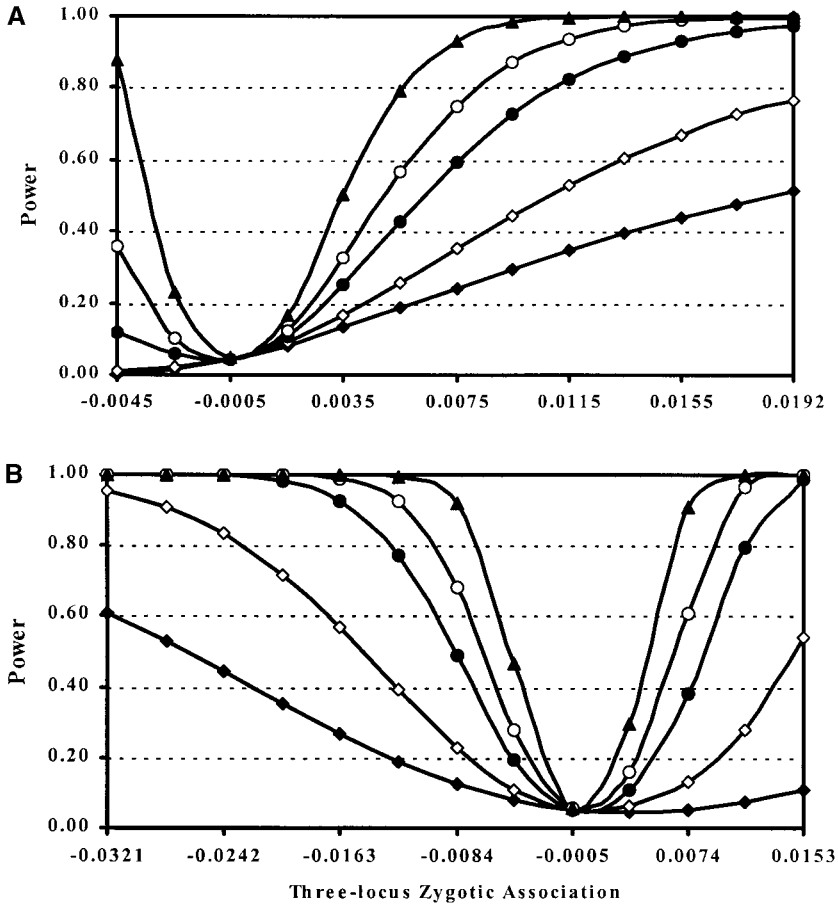

FIGURE 1.—Power to detect three-locus zygotic associations with samples of sizes $n = 30$ (◆), $n = 100$ (◇), $n = 300$ (●), $n = 500$ (○), and $n = 1000$ (▲) for two cases: (A) when heterozygosities at three loci are $H_1 = H_2 = 0.05$ and $H_3 = 0.1$, and three pairwise zygotic associations are $\omega_{12} = \omega_{13} = \omega_{23} = 0.0213$; (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 $\omega_{23} = 0.00.1026$.

because it is the sum of gametic disequilibria or zygotic associations between all pairs of loci.

It is also evident from Figure 1 that low heterozygosities at individual loci cause a very strong asymmetry between positive and negative associations. The unbalanced intensities of associations from both positive and negative sides result in unequal powers unless the sample size is very large. Because of the asymmetry, LEWONTIN's (1964) normalized associations as often used in the literature (*e.g.*, HEDRICK 1987; ZAPATA 2000) may give a false impression about intensities of multilocus associations. For example, BROWN (1975) showed in his Table VII that, with the same amount of normalized three-locus gametic disequilibrium at both sides ($T' = \pm 0.99$), there is a substantial difference in sample size requirements. In the case of gene frequencies equal to 0.2 and two two-locus gametic disequilibria being -0.4 with the third one being zero, BROWN (1975) found that a sample size of $n = 8402$ is required to detect $T' = -0.99$ with the power of 0.9, but only $n = 82$ is needed to detect $T' = +0.99$ with the same amount of power. Had he not given the range of T (-0.0016 to 0.0224), one would be led to believe that the negative disequilibrium is much more difficult to detect than its positive counterpart. The reverse conclusion would be drawn in the cases where the asymmetry is skewed toward the negative side. The truth is that it is the actual, not normalized, association that determines the power and sample size requirement regardless of whether the association is positive or negative. Thus, the use of normalized measures for such purposes should be treated with caution.

DISCUSSION

This article describes measures of zygotic associations at more than two loci and their estimation with samples from diploid populations. These measures are defined as departures of joint zygotic frequencies from the expected values of zero zygotic associations (*cf.* Equations 2 and 5). This is very similar to the definition for gametic disequilibria for two or more loci, which is based on gametic and allelic frequencies (*e.g.*, BENNETT 1954). Thus, it is of little surprise to see that the measures of multilocus zygotic associations share most of the statistical properties by the usual gametic disequilibria. However, the meanings of the two sets of measures are quite different. In fact, a comparative assessment of zygotic associations *vs.* gametic disequilibria may provide some important insights into adaptive significance of genotypes at different loci. For example, if strong zygotic association but little gametic disequilibrium between a pair of loci is observed, then the study population may undergo natural selection favoring highly heterozygous

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 other types of genic disequilibria (Cockerham and Weir 1973; Weir 1979). However, the exceedingly large number of genic disequilibria encountered for multiple alleles at many loci makes such detailed characterization difficult for comparing multilocus organizations among several populations. The multilocus zygotic associations analyzed here summarize different genic disequilibria with no need to consider 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 $\omega_{123} = 0$, given the presence of all three two-locus associations, is given by

$$\Pr\left(\{n_{abc}\}|\{n_{ab+}\},\{n_{a+c}\},\{n_{+bc}\}\right) = \frac{\prod_{a=0}^{1}\prod_{b=0}^{1} n_{ab+}!\prod_{a=0}^{1}\prod_{c=0}^{1} n_{a+c}!\prod_{b=0}^{1}\prod_{c=0}^{1} n_{+bc}!}{n!\prod_{a=0}^{1}\prod_{b=0}^{1}\prod_{c=0}^{1} n_{abc}!},$$

where n_{ab+}, n_{a+c}, and n_{+bc} are marginal total counts of the *ab*th, *ac*th, and *bc*th 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

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

Bartlett, M. S., 1935 Contingency table interactions. J. R. Stat. Soc. Suppl. **2:** 248–252.

Barton, N. H., 2000 Estimating multilocus linkage disequilibria. Heredity **84:** 373–389.

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., 1954 On the theory of random mating. Ann. Eugen. **18:** 311–317.

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., 1975 Sample sizes required to detect linkage disequilibrium between two or three loci. Theor. Popul. Biol. **8:** 184–201.

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

Charlesworth, D., 1991 The apparent selection on neutral marker loci in partially inbreeding populations. Genet. Res. **57:** 159–175.

Cockerham, C. C., and H. Tachida, 1986 Linkage disequilibria in finite populations. Theor. Popul. Biol. **29:** 293–311.

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

Fisher, R. A., 1954 *Statistical Methods for Research Workers*, Ed. 12. Oliver & Boyd, London.

Haldane, J. B. S., 1949 The association of characters as a result of inbreeding and linkage. Ann. Eugen. **15:** 15–23.

Hedrick, P. W., 1987 Gametic disequilibrium measures: proceed with caution. Genetics **117:** 331–341.

Hill, W. G., 1975 Tests for association of gene frequencies at several loci in random mating diploid populations. Biometrics **31:** 881–888.

Lewontin, R. C., 1964 The interaction of selection and linkage. I. General considerations; heterotic models. Genetics **49:** 49–67.

Mitton, J. B., 1997 *Selection in Natural Populations*. Oxford University Press, Oxford.

Ohta, T., and C. C. Cockerham, 1974 Detrimental genes with partial selfing and effects on a neutral locus. Genet. Res. **23:** 191–200.

Pritchard, J. K., and M. Przeworski, 2001 Linkage disequilibrium in humans: model and data. Am. J. Hum. Genet. **69:** 1–14.

Reich, D. E., M. Cargill, S. Bolk, J. Ireland, P. C. Sabeti *et al.*, 2001 Linkage disequilibrium in the human genome. Nature **411:** 199–204.

Thompson, E. A., S. Deeb, D. Walker and A. G. Motulsky, 1988 The detection of linkage disequilibrium between closely linked markers: RFLPs at the AI-CIII apolipoprotein genes. Am. J. Hum. Genet. **42:** 113–124.

Thomson, G., and M. P. Baur, 1984 Third order linkage disequilibrium. Tissue Antigens **24:** 250–255.

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.

Yang, R.-C., 2000 Zygotic associations and multilocus statistics in a nonequilibrium diploid population. Genetics **155:** 1449–1458.

Zapata, C., 2000 The D' measure of overall gametic disequilibrium between pairs of multiallelic loci. Evolution **54:** 1809–1812.

Zaykin, D., L. Zhivotovsky and B. S. Weir, 1995 Exact tests for association between alleles at arbitrary numbers of loci. Genetica **96:** 169–178.

Communicating editor: Y.-X. Fu