

Modeling Extent and Distribution of Zygotic Disequilibrium: Implications for a Multigenerational Canine Pedigree

Tian Liu,* Rory J. Todhunter,[†] Qing Lu,* Lindsay Schoetinger,* Hongying Li,*
Ramon C. Littell,* Nancy Burton-Wurster,[‡] Gregory M. Acland,[‡]
George Lust[‡] and Rongling Wu*¹

*Department of Statistics, University of Florida, Gainesville, Florida 32611 and [†]Department of Clinical Sciences, [‡]James A. Baker Institute for Animal Health, College of Veterinary Medicine, Cornell University, Ithaca, New York 14853

Manuscript received April 28, 2006
Accepted for publication July 1, 2006

ABSTRACT

Unlike gametic linkage disequilibrium defined for a random-mating population, zygotic disequilibrium describes the nonrandom association between different loci in a nonequilibrium population that deviates from Hardy–Weinberg equilibrium. Zygotic disequilibrium specifies five different types of disequilibria simultaneously that are (1) Hardy–Weinberg disequilibria at each locus, (2) gametic disequilibrium (including two alleles in the same gamete, each from a different locus), (3) nongametic disequilibrium (including two alleles in different gametes, each from a different locus), (4) trigenic disequilibrium (including a zygote at one locus and an allele at the other), and (5) quadrigenic disequilibrium (including two zygotes each from a different locus). However, because of the uncertainty on the phase of the double heterozygote, gametic and nongametic disequilibria need to be combined into a composite digenic disequilibrium and further define a composite quadrigenic disequilibrium together with the quadrigenic disequilibrium. To investigate the extent and distribution of zygotic disequilibrium across the canine genome, a total of 148 dogs were genotyped at 247 microsatellite markers located on 39 pairs of chromosomes for an outbred multigenerational pedigree, initiated with a limited number of unrelated founders. A major portion of zygotic disequilibrium was contributed by the composite digenic and quadrigenic disequilibrium whose values and numbers of significant marker pairs are both greater than those of trigenic disequilibrium. All types of disequilibrium are extensive in the canine genome, although their values tend to decrease with extended map distances, but with a greater slope for trigenic disequilibrium than for the other types of disequilibrium. Considerable variation in the pattern of disequilibrium reduction was observed among different chromosomes. The results from this study provide scientific guidance about the determination of the number of markers used for whole-genome association studies.

THE extent and distribution of nonrandom associations between genes at different loci, *i.e.*, linkage disequilibria, throughout the genome have been used often as a criterion to infer demographic and genetic events of a population in the past, such as population history and evolutionary forces governing the loci. Because of its relation with the recombination fraction, the extent of association has provided a foundation for fine-scale mapping of quantitative trait loci (QTL) that control complex diseases in humans (ARDLIE *et al.* 2002) or economical and adaptive traits in livestock (FARNIR *et al.* 2000; MCRÆ *et al.* 2002) and plants (REMINGTON *et al.* 2001). Emerging as an important model system for human health research, canines have recently received a resurgence of interest in unraveling the mysteries of mammalian genomes using linkage

disequilibrium (LD) analysis (HYUN *et al.* 2003; LOU *et al.* 2003; SUTTER and OSTRANDER 2004; SUTTER *et al.* 2004; LINDBLAD-TOH *et al.* 2005). In a study of canine mapping, aimed to detect QTL affecting canine hip dysplasia in a multihierarchic outbred pedigree, we analyzed the extent of pairwise linkage disequilibrium to change over genetic distances with a set of microsatellite markers (240) genotyped from the entire canine genome (LOU *et al.* 2003).

As a common case for many comparable studies, the measure of the extent of linkage disequilibrium between different loci in our canine genetic study was based on multilocus disequilibrium at the gametic level (WEIR 1996). Although such a gametic disequilibrium analysis is mathematically simple, it relies upon a fundamental assumption that the population under study is at Hardy–Weinberg equilibrium (HWE), in which individuals are assumed to be randomly mating to produce the next generations. In such an HWE population, the nonrandom associations of alleles at

¹Corresponding author: Department of Statistics, University of Florida, 533 McCarty Hall C, Gainesville, FL 32611. E-mail: rwu@stat.ufl.edu

different loci occur only within gametes rather than between gametes. The randomly mating assumption may be violated in the canine pedigree used for our earlier study because different offspring are related to each other to a varying degree although multiple dog founders were used.

For a nonequilibrium population at Hardy–Weinberg disequilibrium (HWD), zygotic disequilibria that have power to characterize nonrandom associations at both gametic and zygotic levels (WEIR 1996) may be more relevant. Earlier studies have documented possible genetic and evolutionary causes for zygotic associations in a nonequilibrium population (HALDANE 1949; BENNETT and BINET 1956; CHARLESWORTH 1991; BARTON and GALE 1993). In this article, we revisit our outbred canine pedigree by estimating the extent of zygotic disequilibria throughout the canine genome. Although zygotic disequilibria have been theoretically developed in the literature (see WEIR 1996 for an excellent description), there is no application yet, to our best knowledge, for these measures to extensively study the structure of the genome in a case study. Recently, YANG (2000, 2002) proposed a multilocus zygotic measure for association study in a nonequilibrium population. Yang's two articles present the most thoughtful survey on zygotic disequilibrium analysis. The incorporation of zygotic disequilibrium analysis into genomic research is a necessary first step toward the formulation of an optimal strategy for characterizing genome structure and organization.

ESTIMATION OF ZYGOTIC DISEQUILIBRIUM

Genotype, allele, gamete, and nongamete frequencies: Suppose that there is a natural or experimental population in which there are two codominant markers **A** with two alleles *A* and *a* and **B** with two alleles *B* and *b*, respectively. Let p_A and p_a ($p_A + p_a = 1$) as well as p_B and p_b ($p_B + p_b = 1$) be the corresponding allele frequencies. At each of the two loci, four different formations of zygotic genotypes lead to three distinguishable genotypes, *i.e.*, *AA*, *Aa*, and *aa* for marker **A** and *BB*, *Bb*, and *bb* for marker **B**. The two markers form 10 genotypic configurations, but only 9 can be genetically distinguished from each other. This is because genotypic configurations $\begin{smallmatrix} B & | & b \\ A & | & a \end{smallmatrix}$ and $\begin{smallmatrix} b & | & B \\ a & | & A \end{smallmatrix}$ have the same genotype *AaBb*. Let P , subscripted and superscripted by the genotype notation, be the genotypic configuration frequencies that are individually tabulated in Table 1. It is not difficult to estimate one-marker genotype frequencies from two-marker genotypic configuration frequencies by

$$\begin{aligned} P_{AA} &= P_{AA}^{BB} + P_{AA}^{Bb} + P_{AA}^{bb} \\ P_{Aa} &= P_{Aa}^{BB} + P_{Aa}^{Bb} + P_{Aa}^{bb} + P_{Aa}^{bB} \\ P_{aa} &= P_{aa}^{BB} + P_{aa}^{Bb} + P_{aa}^{bb} \end{aligned} \quad (1)$$

for marker **A** and

$$\begin{aligned} P_{BB} &= P_{AA}^{BB} + P_{Aa}^{BB} + P_{aa}^{BB} \\ P_{Bb} &= P_{AA}^{Bb} + P_{Aa}^{Bb} + P_{Aa}^{bB} + P_{aa}^{Bb} \\ P_{bb} &= P_{AA}^{bb} + P_{Aa}^{bb} + P_{aa}^{bb} \end{aligned} \quad (2)$$

for marker **B** and estimate the allele frequencies from the one-marker genotype frequencies by

$$\begin{aligned} p_A &= P_{AA} + \frac{1}{2}P_{Aa} \\ p_a &= P_{aa} + \frac{1}{2}P_{Aa} \\ p_B &= P_{BB} + \frac{1}{2}P_{Bb} \\ p_b &= P_{bb} + \frac{1}{2}P_{Bb}. \end{aligned} \quad (3)$$

The two markers form four gametes, *AB*, *Ab*, *aB*, and *ab*, whose frequencies can be estimated from genotypic configuration frequencies by

$$\begin{aligned} p_{AB} &= P_{AA}^{BB} + \frac{1}{2}(P_{AA}^{Bb} + P_{Aa}^{BB} + P_{Aa}^{Bb}) \\ p_{Ab} &= P_{AA}^{bb} + \frac{1}{2}(P_{AA}^{bB} + P_{Aa}^{bb} + P_{Aa}^{bB}) \\ p_{aB} &= P_{aa}^{BB} + \frac{1}{2}(P_{Aa}^{BB} + P_{aa}^{Bb} + P_{Aa}^{bB}) \\ p_{ab} &= P_{aa}^{bb} + \frac{1}{2}(P_{Aa}^{bb} + P_{aa}^{Bb} + P_{Aa}^{bB}). \end{aligned} \quad (4)$$

Similarly, the frequencies of nonalleles from different gametes can be estimated by

$$\begin{aligned} p_{A/B} &= P_{AA}^{BB} + \frac{1}{2}(P_{AA}^{Bb} + P_{Aa}^{BB} + P_{Aa}^{bB}) \\ p_{A/b} &= P_{AA}^{bb} + \frac{1}{2}(P_{AA}^{bB} + P_{Aa}^{bb} + P_{Aa}^{bB}) \\ p_{a/B} &= P_{aa}^{BB} + \frac{1}{2}(P_{Aa}^{BB} + P_{aa}^{Bb} + P_{Aa}^{bB}) \\ p_{a/b} &= P_{aa}^{bb} + \frac{1}{2}(P_{Aa}^{bb} + P_{aa}^{Bb} + P_{Aa}^{bB}). \end{aligned} \quad (5)$$

The frequencies of triple alleles from different markers are estimated as

$$\begin{aligned} p_{AA}^B &= P_{AA}^{BB} + \frac{1}{2}P_{AA}^{Bb}, & p_{AA}^b &= P_{AA}^{bb} + \frac{1}{2}P_{AA}^{bB} \\ p_{Aa}^B &= P_{Aa}^{BB} + \frac{1}{2}(P_{Aa}^{Bb} + P_{Aa}^{bB}), & p_{Aa}^b &= P_{Aa}^{bb} + \frac{1}{2}(P_{Aa}^{Bb} + P_{Aa}^{bB}) \\ p_{aa}^B &= P_{aa}^{BB} + \frac{1}{2}P_{aa}^{Bb}, & p_{aa}^b &= P_{aa}^{bb} + \frac{1}{2}P_{aa}^{bB} \\ p_{AA}^{BB} &= P_{AA}^{BB} + \frac{1}{2}P_{Aa}^{BB}, & p_{AA}^{BB} &= P_{AA}^{BB} + \frac{1}{2}P_{Aa}^{BB} \\ p_{Aa}^{Bb} &= P_{Aa}^{Bb} + \frac{1}{2}(P_{Aa}^{Bb} + P_{Aa}^{bB}), & p_{Aa}^{Bb} &= P_{Aa}^{Bb} + \frac{1}{2}(P_{Aa}^{Bb} + P_{Aa}^{bB}) \\ p_{Aa}^{bb} &= P_{Aa}^{bb} + \frac{1}{2}P_{aa}^{bb}, & p_{Aa}^{bb} &= P_{Aa}^{bb} + \frac{1}{2}P_{aa}^{bb}. \end{aligned} \quad (6)$$

Complete disequilibrium parameters: The zygotic disequilibrium is defined as the deviation of two-locus genotype frequencies from products of single-locus

TABLE 1
Frequencies and observations of marker genotypes

Marker A	Marker B			Total
	<i>BB</i> (2)	<i>Bb</i> (1)	<i>bb</i> (0)	
<i>AA</i> (2)	P_{AA}^{BB} n_{22}	P_{AA}^{Bb} n_{21}	P_{AA}^{bb} n_{20}	$P_{AA} = p_A^2 + D_A$ n_2
<i>Aa</i> (1)	P_{Aa}^{BB} n_{12}	$P_{Aa}^{Bb} + P_{Aa}^{bB}$ n_{11}	P_{Aa}^{bb} n_{10}	$P_{Aa} = 2p_A p_a - 2D_A$ n_1
<i>aa</i> (0)	P_{aa}^{BB} n_{02}	P_{aa}^{Bb} n_{01}	P_{aa}^{bb} n_{00}	$P_{aa} = p_a^2 + D_A$ n_0
Total	$P_{BB} = p_B^2 + D_B$ n_2	$P_{Bb} = 2p_B p_b - 2D_B$ n_1	$P_{bb} = p_b^2 + D_B$ n_0	1 n

Genotype *AaBb* contains two different configurations or diplotypes [*AB*][*ab*] and [*Ab*][*aB*].

genotype frequencies and, thus, is composed of all nonallelic genic disequilibria at the two loci (WEIR 1996). Assume that the population considered above is at HWD. This population thus has no desirable property of an equilibrium population, such as independence of different allele frequencies at the same locus (LYNCH and WALSH 1998). The HWD attempts to test for two alleles at the same locus, but on different gametes, whereas (gametic) linkage disequilibrium describes two alleles on the same gametes, but at different loci. For the zygotic disequilibrium, however, there is a third test, *i.e.*, two alleles on different gametes and at different loci.

Since the population is not in HWE, two alleles at each marker are not independent, with the coefficients of Hardy–Weinberg disequilibrium defined as

$$\begin{aligned} D_A &= P_{AA} - p_A^2 \\ &= -\frac{1}{2}P_{Aa} + \frac{1}{2}p_A p_a \\ &= P_{aa} - p_a^2 \end{aligned} \quad (7)$$

for marker **A** and

$$\begin{aligned} D_B &= P_{BB} - p_B^2 \\ &= -\frac{1}{2}P_{Bb} + \frac{1}{2}p_B p_b \\ &= P_{bb} - p_b^2 \end{aligned} \quad (8)$$

for marker **B**, respectively. The coefficient of digenic gametic linkage disequilibrium between the two markers is defined as

$$\begin{aligned} D_{ab} &= p_{AB} - p_A p_B \\ &= -p_{Ab} + p_A p_b \\ &= -p_{aB} + p_a p_B \\ &= p_{ab} - p_a p_b. \end{aligned} \quad (9)$$

For the nonequilibrium population, digenic linkage disequilibrium that occurs between nonalleles at different gametes is defined as

$$\begin{aligned} D_{a/b} &= p_{A/B} - p_A p_B \\ &= -p_{A/b} + p_A p_b \\ &= -p_{a/B} + p_a p_B \\ &= p_{a/b} - p_a p_b. \end{aligned} \quad (10)$$

The trigenic disequilibrium between two alleles from marker **A** and one allele from marker **B** is defined as

$$\begin{aligned} D_{Ab} &= p_{AA}^B - p_A D_{ab} - p_A D_{a/b} - p_B D_A - p_A^2 p_B \\ &= -p_{AA}^b - p_A D_{ab} - p_A D_{a/b} + p_b D_A + p_A^2 p_b \\ &= -\frac{1}{2}p_{Aa}^B - \frac{1}{2}(p_A - p_a)D_{ab} - \frac{1}{2}(p_A - p_a)D_{a/b} - p_B D_A + p_A p_a p_B \\ &= \frac{1}{2}p_{Aa}^b - \frac{1}{2}(p_A - p_a)D_{ab} - \frac{1}{2}(p_A - p_a)D_{a/b} + p_b D_A - p_A p_a p_b \\ &= p_{aa}^B + p_a D_{ab} + p_a D_{a/b} - p_B D_A - p_a^2 p_B \\ &= -p_{aa}^b + p_a D_{ab} + p_a D_{a/b} + p_b D_A + p_a^2 p_b. \end{aligned} \quad (11)$$

The trigenic disequilibrium between two alleles from marker **A** and one allele from marker **B** is defined as

$$\begin{aligned} D_{aB} &= p_A^{BB} - p_B D_{ab} - p_B D_{a/b} - p_A D_A - p_A p_B^2 \\ &= -p_A^{Bb} - p_B D_{ab} - p_B D_{a/b} + p_a D_A + p_A p_B^2 \\ &= -\frac{1}{2}p_A^{Bb} - \frac{1}{2}(p_B - p_b)D_{ab} - \frac{1}{2}(p_B - p_b)D_{a/b} - p_A D_A + p_A p_B p_b \\ &= \frac{1}{2}p_A^{bb} - \frac{1}{2}(p_B - p_b)D_{ab} - \frac{1}{2}(p_B - p_b)D_{a/b} + p_a D_A - p_a p_B p_b \\ &= p_A^{bb} + p_b D_{ab} + p_b D_{a/b} - p_A D_A - p_A p_b^2 \\ &= -p_a^{bb} + p_b D_{ab} + p_b D_{a/b} + p_a D_A + p_a p_b^2. \end{aligned} \quad (12)$$

With genotypic configuration frequencies, allele frequencies, HWD, gametic and nongametic disequilibria, and trigenic disequilibria, we can estimate the quadri-genic disequilibrium (D_{AB}) between two alleles from marker **A** and two alleles from marker **B** using the formulas given in Table 2 (see WEIR 1996). Note that we use lower- and uppercase letters to denote gametic and zygotic disequilibria, respectively. From Table 2, we can see that each of the genotypic configuration frequencies can be expressed in terms of the allele frequencies (p_A ,

TABLE 2

Expressions of quadrigenic disequilibrium D_{AB} in terms of genotypic configuration frequencies, allele frequencies, and lower-order disequilibrium coefficients

Frequency	1	$D_A D_B + D_{ab}^2 + D_{a/b}^2$	D_A	D_B	D_{ab}	$D_{a/b}$	D_{Ab}	D_{aB}
P_{AA}^{BB}	$-p_A^2 p_B^2$	-1	$-p_B^2$	$-p_A^2$	$-2p_A p_B$	$-2p_A p_B$	$-2p_B$	$-2p_A$
$-\frac{1}{2}P_{AA}^{Bb}$	$p_A^2 p_B p_b$	-1	$p_B p_b$	$-p_A^2$	$-p_A p_B + p_A p_b$	$-p_A p_B + p_A p_b$	$-p_B + p_b$	$-2p_A$
P_{AA}^{bb}	$-p_A^2 p_b^2$	-1	$-p_b^2$	$-p_A^2$	$2p_A p_b$	$-2p_A p_b$	$-2p_b$	$-2p_A$
$-\frac{1}{2}P_{Aa}^{BB}$	$p_A p_a p_B^2$	-1	$-p_B^2$	$p_A p_a$	$-p_A p_B + p_a p_B$	$-p_A p_B + p_a p_B$	$-2p_B$	$-p_A + p_a$
$\frac{1}{2}P_{Aa}^{Bb}$	$-p_A p_a p_B p_b$	-1	$p_B p_b$	$p_A p_a$	$-p_A p_B - p_a p_b$	$p_A p_b + p_a p_B$	$-p_B + p_b$	$-p_A + p_a$
$\frac{1}{2}P_{Aa}^{bb}$	$-p_A p_a p_b^2$	-1	$p_B p_b$	$p_A p_a$	$p_A p_b + p_a p_B$	$-p_A p_B - p_a p_b$	$-p_B + p_b$	$-p_A + p_a$
$-\frac{1}{2}P_{Aa}^{bb}$	$p_A p_a p_b^2$	-1	$-p_b^2$	$p_A p_a$	$p_A p_b - p_a p_b$	$p_A p_b - p_a p_b$	$2p_b$	$-p_A + p_a$
P_{aa}^{BB}	$-p_a^2 p_B^2$	-1	$-p_B^2$	$-p_a^2$	$2p_a p_B$	$2p_a p_B$	$-2p_B$	$2p_a$
$-\frac{1}{2}P_{aa}^{Bb}$	$p_a^2 p_B p_b$	-1	$p_B p_b$	$-p_a^2$	$p_a p_B - p_a p_b$	$p_a p_B - p_a p_b$	$-p_B + p_b$	$2p_a$
P_{aa}^{bb}	$-p_a^2 p_b^2$	-1	$-p_b^2$	$-p_a^2$	$-2p_a p_b$	$-2p_a p_b$	$2p_b$	$2p_a$

p_a and p_B, p_b), HWD coefficients (D_A and D_B), and gametic (D_{ab}) and nongametic disequilibria of different orders ($D_{a/b}, D_{Ab}, D_{aB}$, and D_{AB}).

Composite zygotic disequilibria: It can be seen that 10 genotypic configurations have nine independent frequencies that are defined by two allele frequencies for each marker and seven disequilibrium parameters as defined above. But since two configurations of the double heterozygote cannot be separated in practice, it is not possible to estimate all these frequencies and disequilibrium parameters. To solve this problem, WEIR (1996) suggested a set of composite disequilibrium coefficients. These include the digenic disequilibrium measured by the sum of the gametic and nongametic coefficients, *i.e.*,

$$\Delta_{ab} = D_{ab} + D_{a/b}. \quad (13)$$

As shown by Equations 9 and 10, Δ_{ab} will include the summation of gamete (p_{AB}) and nongamete frequencies ($p_{A/B}$). On the basis of the definitions of these two frequencies (Equations 4 and 5), Δ_{ab} will finally need the summation of two configuration frequencies (P_{AA}^{Bb} and P_{Aa}^{BB}) of the double heterozygote. Thus, Δ_{ab} can

be estimated directly on observable genotype frequencies. WEIR (1996) also defined a quadrigenic disequilibrium measured by

$$\Delta_{AB} = D_{AB} - 2D_{ab}D_{a/b}, \quad (14)$$

which can be finally measured from genotype frequencies.

The two composite digenic and quadrigenic disequilibria can make it possible to estimate the parameters on the basis of observable genotype frequencies rather than unobservable configuration frequencies. Table 3 tabulates the compositions of the composite quadrigenic disequilibrium in terms of genotype and allele frequencies and the coefficients of disequilibria with lower orders (see also WEIR and COCKERHAM 1989).

Estimates and tests: Two markers **A** and **B** are observed for a population of size n with nine genotypes listed in Table 1. Let u and v denote the marker genotypes, $u = 2$ for AA , 1 for Aa , and 0 for aa and $v = 2$ for BB , 1 for Bb , and 0 for bb . The multinomial log-likelihood of the genotype frequencies P_u^v given marker observations is written as

TABLE 3

Expressions of composite quadrigenic disequilibrium Δ_{AB} in terms of genotypic and allele frequencies and lower-order disequilibrium coefficients

Frequency	1	$D_A D_B + \Delta_{ab}^2$	D_A	D_B	Δ_{ab}	D_{Ab}	D_{aB}
P_{AA}^{BB}	$-p_A^2 p_B^2$	-1	$-p_B^2$	$-p_A^2$	$-2p_A p_B$	$-2p_B$	$-2p_A$
$-\frac{1}{2}P_{AA}^{Bb}$	$p_A^2 p_B p_b$	-1	$p_B p_b$	$-p_A^2$	$-p_A p_B + p_A p_b$	$-p_B + p_b$	$-2p_A$
P_{AA}^{bb}	$-p_A^2 p_b^2$	-1	$-p_b^2$	$-p_A^2$	$2p_A p_b$	$2p_b$	$-2p_A$
$-\frac{1}{2}P_{Aa}^{BB}$	$p_A p_a p_B^2$	-1	$-p_B^2$	$p_A p_a$	$-p_A p_B + p_a p_B$	$-2p_B$	$-p_A + p_a$
P_{11}^a	$-p_A p_a p_B p_b$	-1	$p_B p_b$	$p_A p_a$	$-\frac{1}{2}p_A p_B - \frac{1}{2}p_a p_b + \frac{1}{2}p_A p_b + \frac{1}{2}p_a p_B$	$-p_B + p_b$	$-p_A + p_a$
$-\frac{1}{2}P_{Aa}^{bb}$	$p_A p_a p_b^2$	-1	$-p_b^2$	$p_A p_a$	$p_A p_b - p_a p_b$	$2p_b$	$-p_A + p_a$
P_{aa}^{BB}	$-p_a^2 p_B^2$	-1	$-p_B^2$	$-p_a^2$	$2p_a p_B$	$-2p_B$	$2p_a$
$-\frac{1}{2}P_{aa}^{Bb}$	$p_a^2 p_B p_b$	-1	$p_B p_b$	$-p_a^2$	$p_a p_B - p_a p_b$	$-p_B + p_b$	$2p_a$
P_{aa}^{bb}	$-p_a^2 p_b^2$	-1	$-p_b^2$	$-p_a^2$	$-2p_a p_b$	$2p_b$	$2p_a$

$$^a P_{11} = P_{Bb}^{Aa} + P_{Bb}^{aA}.$$

$$\log L(P | \mathbf{A}, \mathbf{B}) = \text{Constant} + \sum_{u=0}^2 \sum_{v=0}^2 n_{uv} \log P_u^v, \quad (15)$$

which gives the MLEs of the genotype frequencies as

$$\hat{P}_u^v = \frac{n_{uv}}{n}. \quad (16)$$

On the basis of the estimated genotype frequencies, the allele frequencies for the two markers (p_A and p_B), the HWD coefficients (D_A and D_B), the composite digenic disequilibrium (Δ_{ab}), two trigenic disequilibria (D_{Ab} and D_{aB}), and the composite quadrigenic disequilibrium (Δ_{AB}) can be estimated.

Each of these disequilibria should be tested for its significance. The hypotheses for testing HWD are formulated by

$$H_0: D_A = 0 \text{ vs. } H_1: D_A \neq 0 \quad (17)$$

$$H_0: D_B = 0 \text{ vs. } H_1: D_B \neq 0 \quad (18)$$

for two different markers, respectively. The hypotheses for testing each of the zygotic disequilibria between the two markers are given as

$$H_0: \Delta_{ab} = 0 \text{ vs. } H_1: \Delta_{ab} \neq 0 \quad (19)$$

$$H_0: D_{Ab} = 0 \text{ vs. } H_1: D_{Ab} \neq 0 \quad (20)$$

$$H_0: D_{aB} = 0 \text{ vs. } H_1: D_{aB} \neq 0 \quad (21)$$

$$H_0: \Delta_{AB} = 0 \text{ vs. } H_1: \Delta_{AB} \neq 0. \quad (22)$$

For these hypotheses (17–22), we calculate the likelihoods under H_0 and H_1 , respectively, from which the log-likelihood ratio (LR) is calculated. The LR test statistic calculated follows a χ^2 -distribution with 1 d.f.

The likelihoods for testing HWD on the basis of hypotheses (17) and (18) can be calculated from marginal totals of one-marker genotype frequencies and observations separately for markers **A** and **B**, respectively. For these two hypotheses, allele frequencies under H_0 can be estimated with a closed form and, thus, no EM algorithm is needed for computation. However, for the tests of hypotheses (19–22), parameter estimation under H_0 needs the implementation of numerical algorithms, like the Newton–Raphson method, because the number of unknown parameters to be estimated is less than the number of genotype frequencies. It is also possible to test whether all the disequilibrium coefficients are together equal to zero. The parameters that need to be estimated under $H_0: \Delta_{ab} = D_{Ab} = D_{aB} = \Delta_{AB} = 0$, include allele frequencies and HWD coefficients that can be estimated with a closed form. The LR value for this hypothesis should asymptotically follow the χ^2 -distribution with 4 d.f.

Alternatively, hypotheses (17–22) for a given disequilibrium can be tested by calculating test statistics

$$\chi^2 = \frac{\hat{D}}{\text{var}(\hat{D})},$$

where \hat{D} denotes the estimate of the disequilibrium coefficient and $\text{var}(\hat{D})$ is the sampling variance of the estimate, calculated by formulas given in WEIR (1996). This test statistic is asymptotically χ^2 -distributed with 1 d.f.

Bounds and normalization: To make zygotic disequilibria comparable between different studies, the estimates of disequilibria should be normalized. LEWONTIN (1964) proposed a standardized approach by expressing linkage disequilibrium as a proportion of the most extreme value. Thus, the new measure from this approach will lie between 0 (for linkage equilibrium) and $|\pm 1|$ (for complete linkage disequilibrium). A similar idea was used by WEIR and COCKERHAM (1989) to derive bounds for trigenic and quadrigenic disequilibria for zygotic nonequilibrium analysis. More recently, ZAYKIN (2004) and HAMILTON and COLE (2004) independently proposed algebraically equivalent bounds for a composite measure of gametic linkage disequilibrium. The bound for the composite zygotic disequilibrium has not been provided thus far. In the APPENDIX, we provide bounds and normalized measures for all six disequilibria, D_A , D_B , Δ_{ab} , D_{Ab} , D_{aB} , and Δ_{AB} , for zygotic disequilibrium analysis. These bounds for the first five disequilibria are consistent with those published in WEIR and COCKERHAM (1989), ZAYKIN (2004), and HAMILTON and COLE (2004).

MATERIALS

A canine pedigree was developed to map QTL responsible for canine hip dysplasia (CHD) using molecular markers. Seven founding greyhounds and six founding Labrador retrievers were intercrossed, followed by backcrossing F_1 's to the greyhounds and Labrador retrievers and intercrossing the F_1 's. A series of subsequent intercrosses among the progeny at different generation levels led to a complex network pedigree structure (Figure 1), which maximized phenotypic ranges in CHD-related quantitative traits and the chance to detect substantial linkage disequilibria (TODHUNTER *et al.* 1999, 2003a,b; BLISS *et al.* 2002). A total of 148 dogs from this structured pedigree were chosen for genetic analyses. This set of samples would not be appropriate for traditional gametic linkage disequilibrium analysis because the population is not randomly mating. LOU *et al.* (2003) estimated gametic linkage disequilibria for this pedigree on a critical foundation that the pedigree was originally derived from multiple unrelated founders. But although the resulting conclusions are consistent with the evolutionary history of dogs, Lou *et al.*'s analysis can be improved by estimating and testing the chromosomal distribution of zygotic disequilibria as will be done in this study.

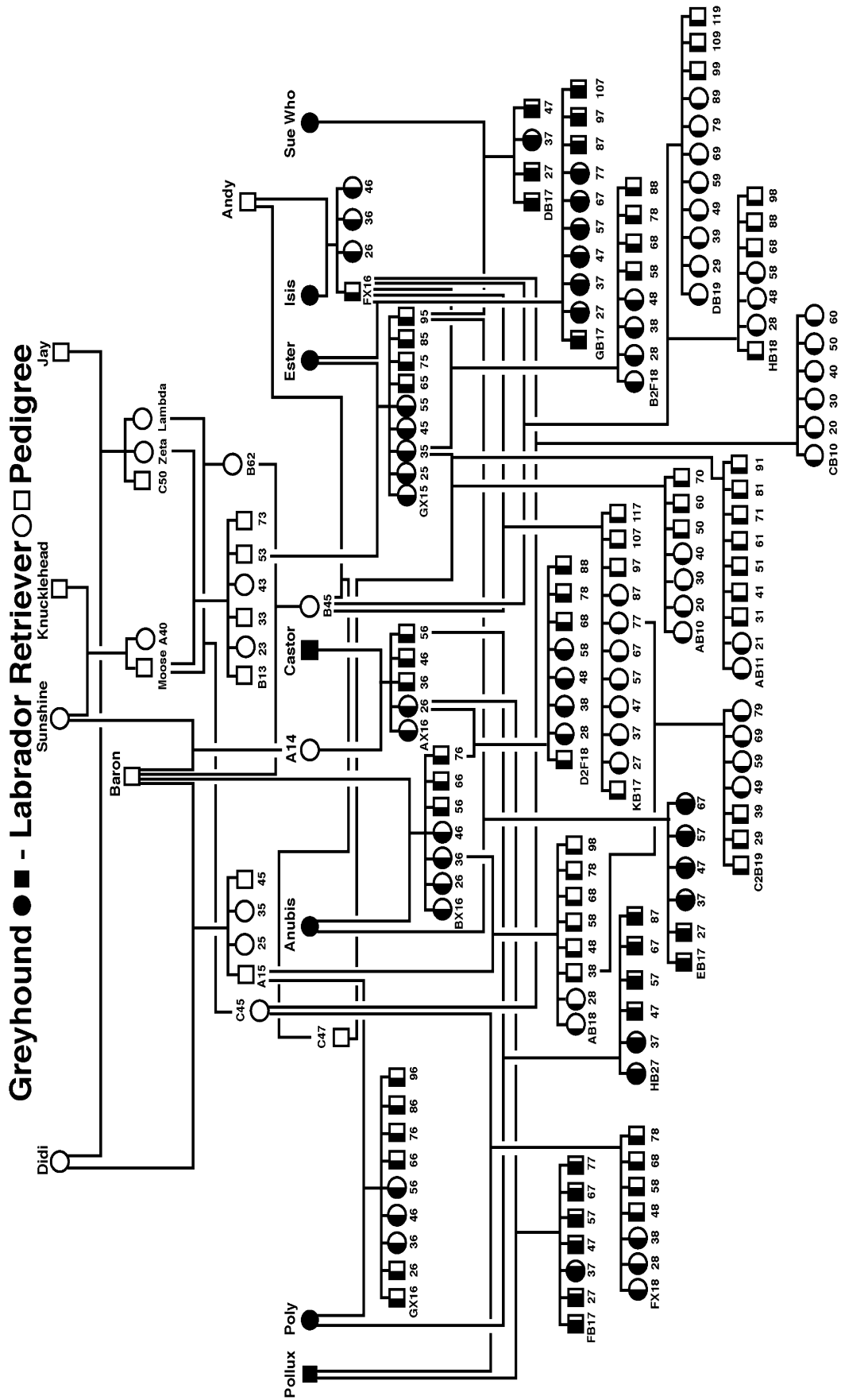


FIGURE 1.—Diagram of an outbred pedigree in dog. Squares and circles represent males and females, respectively. Solid and open portions of each symbol represent the proportion of greyhound and Labrador retriever alleles, respectively, possessed by that dog.

TABLE 4
The percentages and distributions of significant HWD and gametic and zygotic disequilibria through 39 chromosomes in the canine pedigree

Chromosome	No. markers	Averaged genetic distance (cM)	D_A (%)	Δ_{ab} (%)	D_{Ab} (%)	D_{aB} (%)	Δ_{AB} (%)
1	11	11.2	55	62	16	20	24
2	11	5.9	27	56	24	20	18
3	9	10.9	11	61	8	14	14
4	8	10.8	50	64	14	14	18
5	10	8.2	10	60	24	13	11
6	6	9.1	17	67	0	20	20
7	10	8.2	40	78	11	24	24
8	6	9.0	17	73	47	27	20
9	7	8.7	29	57	43	43	19
10	7	13.4	14	57	19	29	24
11	7	12.9	29	38	19	14	5
12	9	8.6	22	47	11	11	14
13	5	15.8	40	30	0	20	30
14	7	7.0	14	48	19	10	19
15	7	5.0	14	48	19	14	29
16	4	3.4	25	67	17	0	0
17	5	15.8	60	40	10	10	0
18	7	9.1	43	57	0	29	5
19	5	11.0	60	50	30	0	10
20	5	10.9	20	20	20	10	10
21	5	12.8	20	70	30	20	10
22	6	9.7	17	40	13	27	20
23	6	10.3	33	53	7	20	7
24	4	12.0	25	83	33	0	50
25	6	8.5	50	47	20	40	20
26	5	6.8	0	50	30	10	40
27	6	8.8	33	47	33	20	33
28	6	10.5	33	87	27	7	33
29	4	9.1	0	67	0	0	17
30	7	6.3	14	57	14	10	14
31	5	6.6	20	70	20	10	0
32	4	11.7	25	83	17	17	17
33	5	4.9	60	90	50	40	10
34	4	13.5	0	83	17	0	33
35	4	6.1	25	67	33	33	0
36	2	7.3	0	100	100	100	100
37	13	7.9	31	73	18	18	22
38	4	4.0	0	50	33	17	17
39	5	11.9	100	70	40	0	90
Overall	247	9.3	28	61	23	19	22

For the sampled dogs from the structured pedigree, 247 microsatellite markers distributed on 38 pairs of autosomes and 1 pair of sex chromosomes were genotyped to construct a linkage map for the canine genome, which displays a good coverage of each chromosome (MELLERSH *et al.* 1997, 2000; BREEN *et al.* 2001; RICHMAN *et al.* 2001). The recombination fractions between different markers were estimated for segregating families, which are converted to genetic distances in centimorgans on the basis of a map function. The average genetic distances between two adjacent markers on each chromosome are listed in Table 4 (BREEN *et al.* 2001).

RESULTS

The microsatellite markers genotyped display high heterozygosity in the dog pedigree, with the number of alleles at a marker ranging from 2 to 11 (TODHUNTER *et al.* 2003b). The multialleles of the microsatellite markers are collapsed into two categories, the most frequent allele *vs.* all the rest pooled alleles. Thus, the simple biallelic model can be directly used to analyze the extent and distribution of zygotic disequilibria throughout the canine genome using the model developed above.

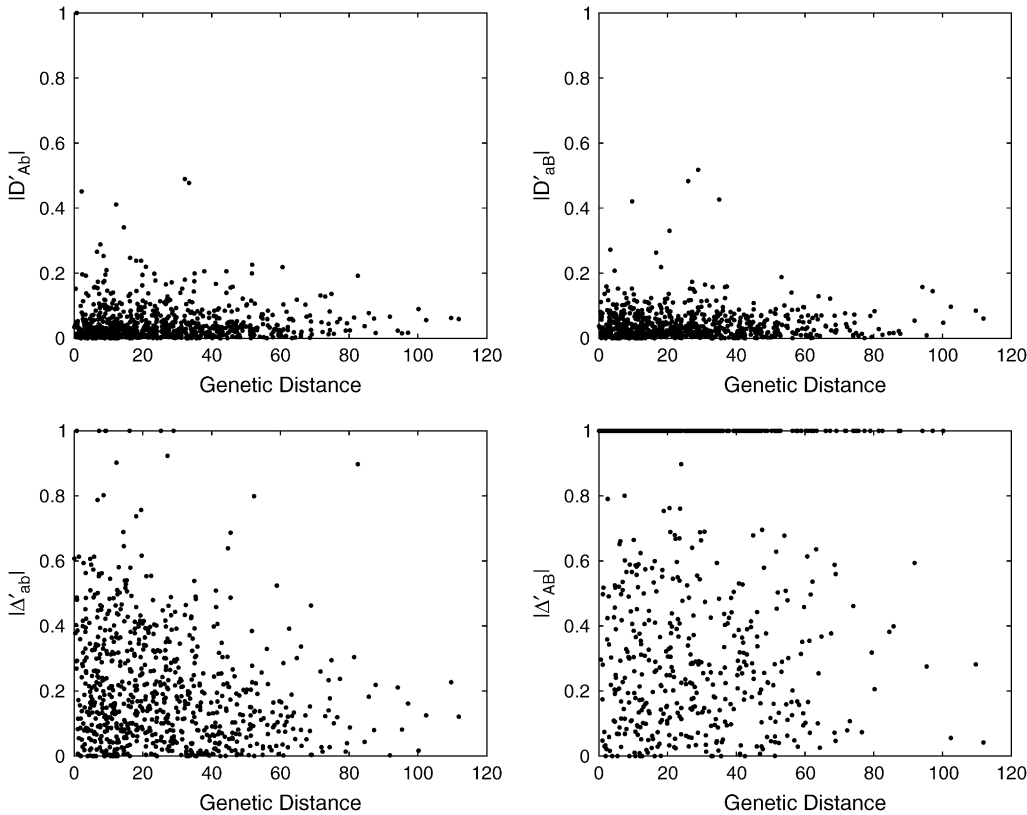


FIGURE 2.—Distributions of gametic and zygotic disequilibria values observed between syntenic marker pairs as a function of genetic distance in centimorgans.

The zygotic disequilibria that describe the association between two different markers in a nonequilibrium population, like the canine pedigree as used in this study, were estimated and tested for each pair of markers located on the same chromosome. The zygotic associations were partitioned into Hardy–Weinberg disequilibria at each locus (D_A), composite gametic disequilibrium including two alleles each from a different locus (Δ_{ab}), trigenic disequilibria including a zygote at one locus and an allele at the other (D_{Ab} or D_{aB}), and composite quadrigenic disequilibrium including two zygotes each from a different locus (Δ_{AB}). All these disequilibrium coefficients were normalized using a procedure described in the APPENDIX. All the comparisons are based on the normalized coefficients.

Overall, 28% of the markers genotyped were observed to deviate from HWE, but showed considerable interchromosomal variation ranging from 0 (chromosomes 26, 29, 34, 36, and 38) to 100% (sex chromosome) (Table 4). Of the four types of dilocus disequilibrium, Δ_{ab} displays the most important impact on zygotic associations because its estimates are generally much larger than those of the other disequilibrium types. Furthermore, this disequilibrium, as well as the composite quadrigenic disequilibrium, has larger normalized values than the other types (Figure 2). Overall, the largest percentage of marker pairs is significant for Δ_{ab} (61%), followed by trigenic disequilibria D_{Ab} (23%) and D_{aB} (19%) and composite quadrigenic disequilibrium Δ_{AB}

(22%). The percentages of marker pairs that exhibit significant associations vary among different chromosomes (Table 4).

Figure 2 illustrates the patterns of the relationship between zygotic disequilibria, Δ_{ab} , D_{Ab} , D_{aB} , and Δ_{AB} , and genetic distances, all exhibiting a trend of decay with increased map distance. All the types of zygotic disequilibria occur more frequently between pairs of markers separated by <40 cM than between those separated by >40 cM. As compared with D_{Ab} and D_{aB} , Δ_{ab} and Δ_{AB} tend to extend within a broader region of the canine genome. Both Δ_{ab} and Δ_{AB} decay with map distance, to a greater extent for the former than for the latter.

Each of the four types of zygotic association was plotted against the map distance separately for individual chromosomes (Figures 3–6). Although the data are sparse, a general trend can be observed for the extent of zygotic disequilibria; *i.e.*, whereas the distributions of D_{Ab} and D_{aB} follow a similar pattern among different chromosomes, there is substantial interchromosomal variation in the extent and distribution of Δ_{ab} and Δ_{AB} over the canine genome.

MONTE CARLO SIMULATION

To our best knowledge, this is the first study of the distribution of zygotic disequilibrium across the

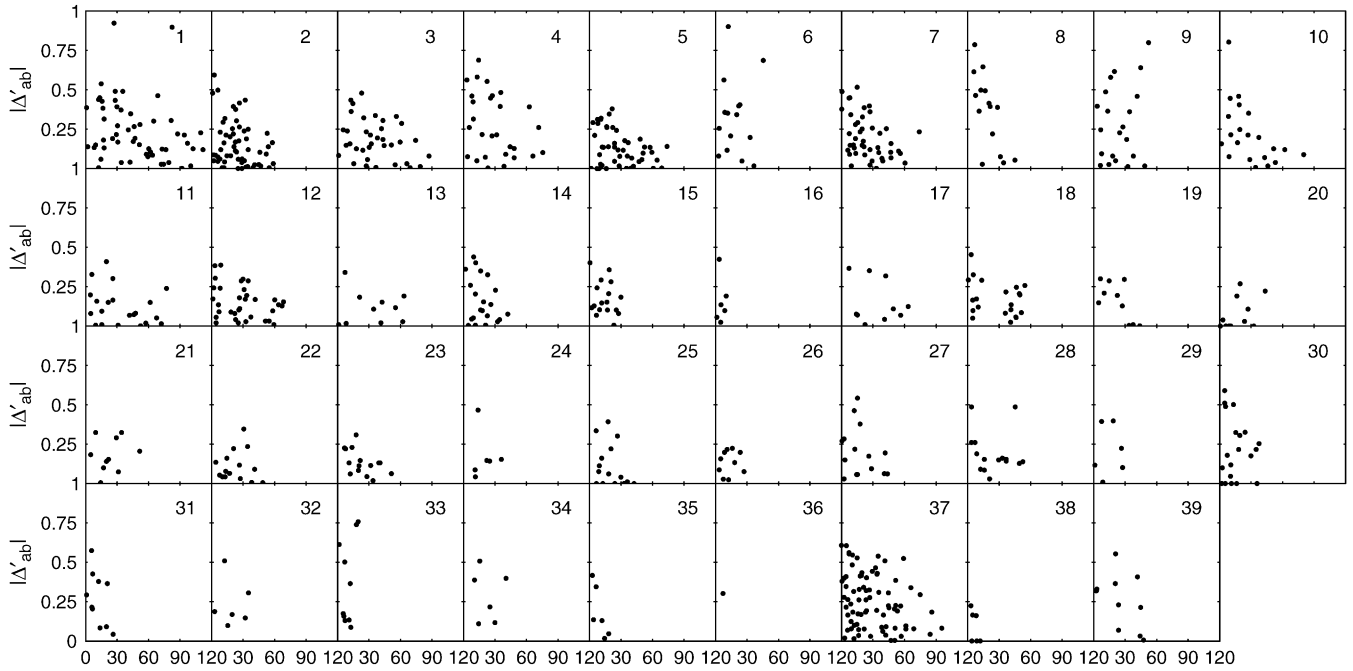


FIGURE 3.—Interchromosomal heterogeneity in the extent and distribution of digenic linkage disequilibrium Δ_{ab} among 39 chromosomes.

genome in a nonequilibrium population. Given the tradition that most current linkage disequilibrium analyses are based on gametic associations without a test for zygotic disequilibria, we perform a reciprocal simulation study to examine the influence of such analyses on the power of the disequilibrium test in a nonequilibrium population. According to this recipro-

cal simulation study, data are simulated, respectively, under zygotic and gametic disequilibrium models, but are subject to separate analyses by each of these two models.

Simulated data by the zygotic model: Table 5 lists four simulation designs in each of which all types of associations occur for an assumed nonequilibrium

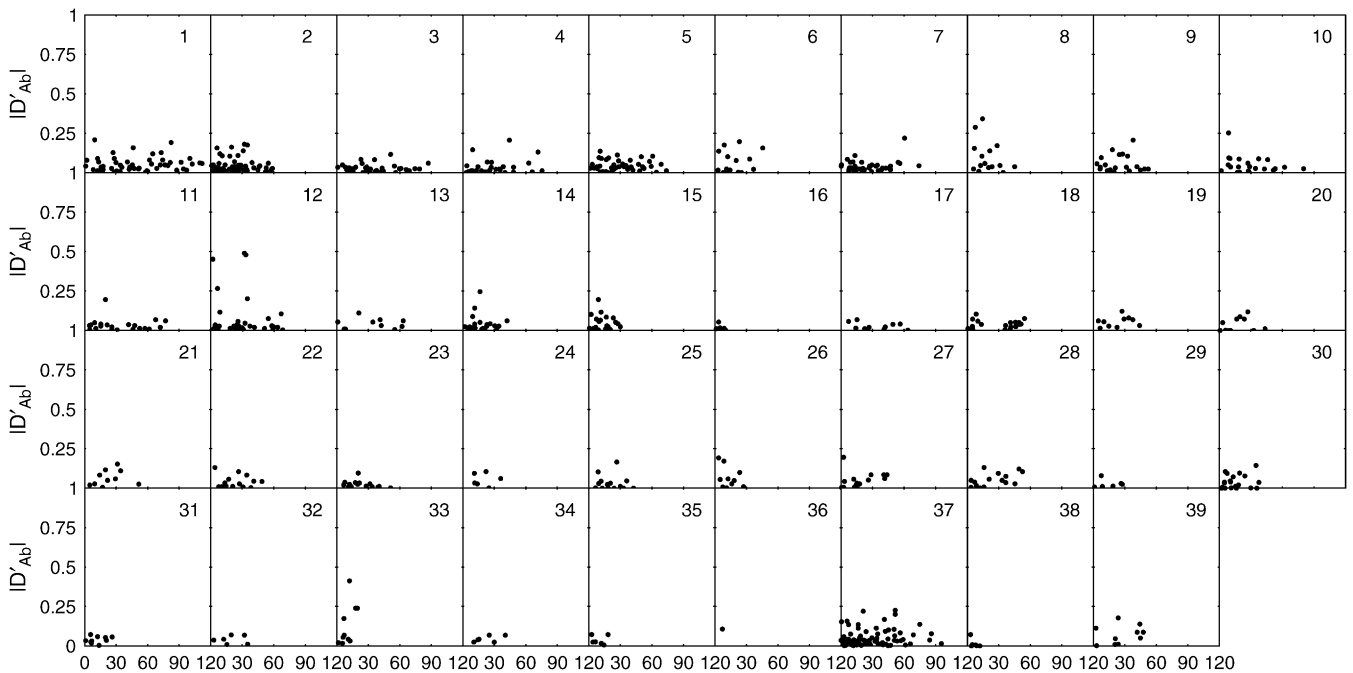


FIGURE 4.—Interchromosomal heterogeneity in the extent and distribution of trigenic linkage disequilibrium D_{Ab} among 39 chromosomes.

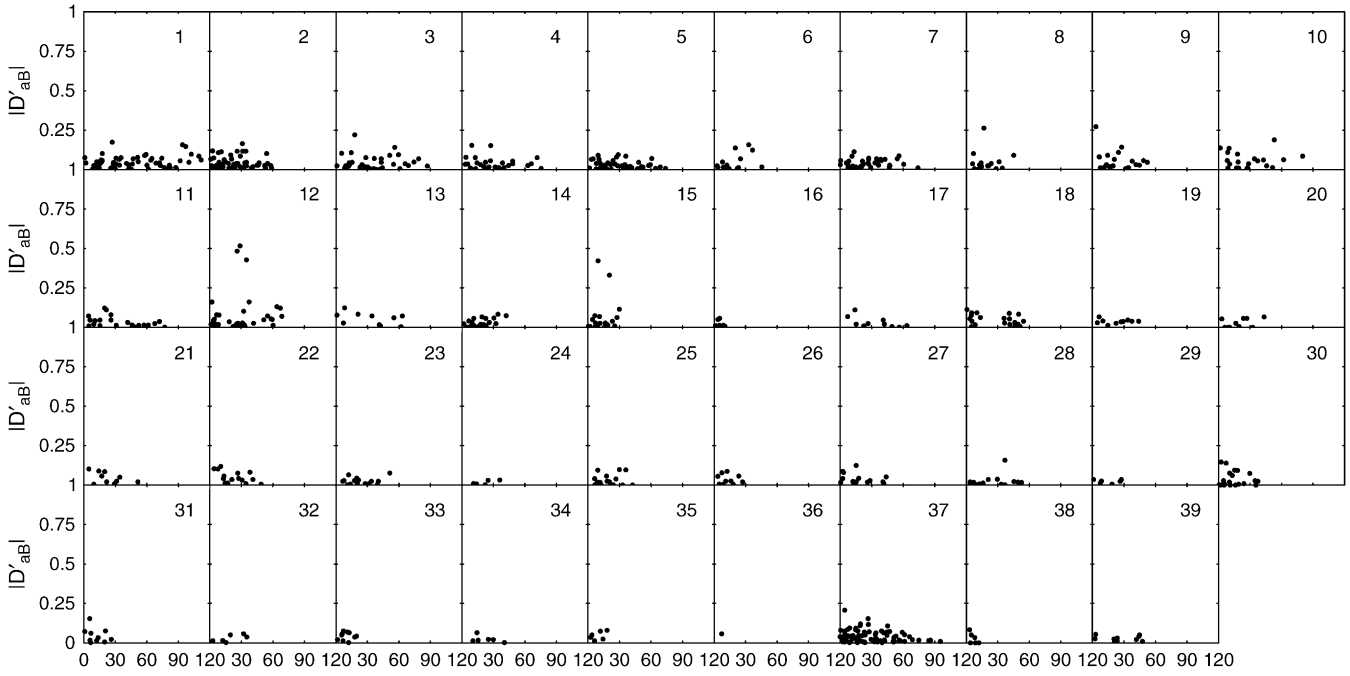


FIGURE 5.—Interchromosomal heterogeneity in the extent and distribution of digenic linkage disequilibrium D_{AB} among 39 chromosomes.

population. But these four designs are different in terms of the allocation pattern of zygotic associations. In designs 1 and 2, a large composite digenic disequilibrium is contributed mainly by gametic or nongametic disequilibrium, respectively. Designs 3 and 4 purport to have a large trigenic and a quadrigenic disequilibrium,

respectively. The sample size is 150, mimicking the canine example used above. The simulated data are analyzed by both the gametic and the zygotic disequilibrium models. The simulation under each design is repeated 200 times to calculate the precision of parameter estimation and statistical power of disequilibrium

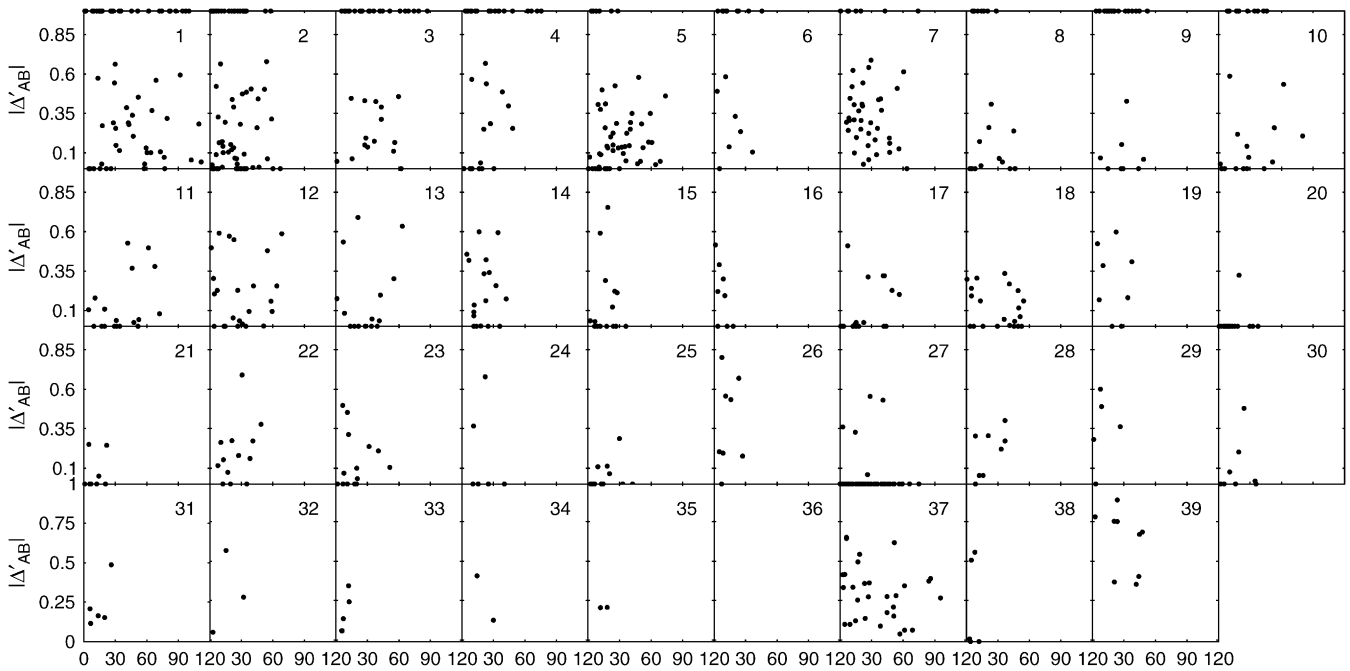


FIGURE 6.—Interchromosomal heterogeneity in the extent and distribution of quadrigenic linkage disequilibrium Δ_{AB} among 39 chromosomes.

TABLE 5
Given parameter values for simulation under different designs

Design	p_A	p_B	D_A	D_B	D_{ab}	$D_{a/b}$	D_{Ab}	D_{aB}	D_{AB}
1	0.5	0.6	0.05	0.05	0.10	0.01	0.01	0.01	0.01
2	0.5	0.6	0.05	0.05	0.01	0.10	0.01	0.01	0.01
3	0.5	0.6	0.05	0.05	0.02	0.02	0.03	0.03	0.01
4	0.5	0.6	0.05	0.05	0.02	0.02	0.01	0.01	0.03

detection. The results from this simulation study (Table 6) are summarized as follows:

1. The zygotic disequilibrium model provides reasonable estimation of any type of disequilibria and shows a great power to detect disequilibria for a nonequilibrium population under simulation.
2. As expected, the gametic linkage disequilibrium model can estimate only gametic linkage disequilibrium, but when used to estimate a nonequilibrium population, its estimation of this parameter is largely biased. Actually, the gametic model tends to estimate the composite gametic and nongametic disequilibrium when both exist, but its estimation precision is very poor. If the composite digenic disequilibrium is mainly due to the nongametic disequilibrium (de-

sign 2), the gametic disequilibrium model cannot be used, given its large estimation error.

3. The gametic disequilibrium model can accurately estimate allele frequencies, but cannot provide precise estimation of these parameters. The second and third findings indicate that gametic disequilibrium analysis should never be used for a nonequilibrium population and that the test for zygotic disequilibrium is always crucial before gametic disequilibrium analysis is used.

Simulated data by gametic model: As a follow-up, we simulated the data for an equilibrium population by a gametic linkage disequilibrium model. The simulated data were analyzed by both the zygotic and the gametic models (Table 7). It can be seen that the zygotic model

TABLE 6

Maximum-likelihood estimates of parameters and the square roots of their mean square errors (in parentheses) estimated by the zygotic- and gametic-LD models for the data simulated under the zygotic-LD model of different designs

Model	p_A	p_B	D_A	D_B	Δ_{ab}	D_{Ab}	D_{aB}	Δ_{AB}	D_{ab}
Design 1									
True	0.5	0.6	0.05	0.05	0.11	0.01	0.1	0.008	
Zygotic	0.5 (0.006)	0.6 (0.006)	0.05 (0.005)	0.05 (0.005)	0.11 (0.005)	0.01 (0.005)	0.1 (0.005)	0.007 (0.005)	
Gametic	0.5 (0.022)	0.6 (0.022)	—	—	—	—	—	—	0.078 (0.022)
Design 2									
True	0.5	0.6	0.05	0.05	0.11	0.01	0.1	0.008	
Zygotic	0.5 (0.004)	0.6 (0.004)	0.05 (0.004)	0.05 (0.004)	0.11 (0.004)	0.01 (0.003)	0.1 (0.003)	0.007 (0.003)	
Gametic	0.5 (0.068)	0.6 (0.067)	—	—	—	—	—	—	0.078 (0.068)
Design 3									
True	0.5	0.6	0.05	0.05	0.04	0.03	0.03	0.0092	
Zygotic	0.5 (0.007)	0.6 (0.007)	0.05 (0.006)	0.05 (0.006)	0.04 (0.006)	0.03 (0.006)	0.03 (0.006)	0.0090 (0.0061)	
Gametic	0.5 (0.022)	0.6 (0.022)	—	—	—	—	—	—	0.041 (0.021)
Design 4									
True	0.5	0.6	0.05	0.05	0.04	0.01	0.01	0.0292	
Zygotic	0.5 (0.004)	0.6 (0.004)	0.05 (0.004)	0.05 (0.004)	0.04 (0.004)	0.01 (0.003)	0.01 (0.003)	0.0280 (0.0034)	
Gametic	0.5 (0.027)	0.6 (0.027)	—	—	—	—	—	—	0.047 (0.028)

TABLE 7

Maximum-likelihood estimates of parameters and the square roots of their mean square errors (in parentheses) estimated by the zygotic- and gametic-LD models for the data simulated under the gametic-LD model

Model	p_A	p_B	D_A	D_B	Δ_{ab}	D_{Ab}	D_{aB}	Δ_{AB}	D_{ab}
True	0.5	0.6	0	0	0.08	0	0	0	0.08
Zygotic	0.5 (0.005)	0.6 (0.005)	0 (0.004)	0 (0.004)	0.08 (0.004)	0 (0.004)	0 (0.004)	0 (0.004)	
Gametic	0.5 (0.003)	0.6 (0.003)	— —	— —	— —	— —	— —	— —	0.08 (0.003)

estimates the coefficient of linkage disequilibrium as precisely as the gametic model. The result from this simulation indicates that the zygotic model is powerful to estimate the degree of linkage disequilibrium for an equilibrium population. In conjunction with the results from the simulation by the zygotic disequilibrium model, it is concluded that the zygotic model is more general than the gametic model.

DISCUSSION

The characterization of the architecture of linkage disequilibrium in the genome is an area of explosive recent growth (FARNIR *et al.* 2000; REMINGTON *et al.* 2001; ARDLIE *et al.* 2002; HYUN *et al.* 2003; LOU *et al.* 2003; SUTTER and OSTRANDER 2004; SUTTER *et al.* 2004; LINDBLAD-TOH *et al.* 2005) because the positional cloning of genes underlying common complex diseases relies on the identification of linkage disequilibrium between genetic markers and disease. Traditional linkage disequilibrium is defined as the nonrandom association between alleles at different loci in gametes or haplotypes. The estimation of such gametic linkage disequilibrium between different loci requires the assumption that the population under consideration is randomly mating, following HWE. However, for many nonequilibrium populations that are founded by a small number of ancestors and/or are frequently under evolutionary pressure, such as mutation, genetic drift, and population admixture and structure, or under artificial selection (LYNCH and WALSH 1998), HWE may be violated and, therefore, a new analysis that relaxes the random-mating assumption should be formulated. WEIR (1996) introduced the concept of zygotic association or zygotic disequilibrium that can characterize the disequilibria between different loci in a nonequilibrium population. Recently, YANG (2000, 2002) proposed a multilocus statistic to examine zygotic associations in nonequilibrium populations. Different disequilibria due to a single locus or multiple loci can be summarized in such a statistic.

In a multigenerational canine pedigree constructed by several founders (TODHUNTER *et al.* 1999), individual dogs are related to each other and, thus, sampled dogs from this pedigree violate the HWE assumption due to

inbreeding. For this reason, zygotic disequilibrium should be more appropriate for this related pedigree to investigate the extent and distribution of associations throughout the canine genome. We found extensive linkage disequilibria in a broad region of chromosomes (≥ 40 cM), as compared with the human genome, even for the most isolated human populations (HALL *et al.* 2002; VARILLO *et al.* 2003; TENESA *et al.* 2004). This finding seems to be comparable with those of earlier linkage disequilibrium studies of purebred dogs (HYUN *et al.* 2003; SUTTER *et al.* 2004). The extent of linkage disequilibrium across the chromosomes was also investigated for the same data set by the gametic linkage disequilibrium model (LOU *et al.* 2003). Although the results of the two models are broadly in agreement, the linkage disequilibrium detected by the zygotic model seems to be distributed more extensively over the genome than that detected previously by the gametic model. Given the finding from the simulation, the gametic model tends to estimate a combined gametic and nongametic linkage disequilibrium, *i.e.*, composite digenic disequilibrium, and, therefore, to provide a biased estimate of gametic linkage disequilibrium especially when a large nongametic linkage disequilibrium exists. The extensive distribution of linkage disequilibrium in the canine genome detected by the zygotic model suggests that a relatively small number of markers will be required for whole-genome association mapping in dogs. However, an optimal number of markers should be determined separately for individual chromosomes, because the extent of linkage disequilibrium shows substantial interchromosomal variation. Historically, different degrees of selection pressure may have been operational on various chromosomes, which causes interchromosomal differentiation in linkage disequilibrium extent (SUTTER and OSTRANDER 2004; OSTRANDER and WAYNE 2005; PARKER and OSTRANDER 2005).

The most significant contribution of this article may lie in the first systematic use of a zygotic disequilibrium analysis to characterize the extent of disequilibrium for a nonequilibrium population of canines although the conclusions obtained from our analysis may be explained only for the specific canine pedigree used, in which individual dogs are related to different extents. On the basis of simulation analyses, the idea and

concept of zygotic disequilibrium can be readily applied to any population genetic studies. Results from simulation analyses indicate that a popular gametic linkage disequilibrium analysis when employed to understand the genetic structure of the population at HWD should be used with caution because the results from this analysis will be misleading. The zygotic disequilibrium model that does not rely on the assumption of random mating has great power to detect various types of disequilibrium at different orders. Therefore, it is safe to say that the zygotic disequilibrium model covers well the gametic disequilibrium model in practical population genetic studies.

In this study, the zygotic disequilibrium model mostly modified from WEIR (1996) was proposed on the basis of biallelic markers although the data from a canine genetic project are multiallelic microsatellites. Given the current modest sample size used, it should be more reasonable to collapse multiple alleles into bialleles than to directly use the multiallelic zygotic model in terms of reducing the number of parameters being estimated. Also, with the development of high-throughput technologies for single-nucleotide polymorphism (SNP) markers, the biallelic model will be useful to analyze the genetic architecture of zygotic disequilibria over the entire genome for any nonequilibrium or isolated populations including humans and other agriculturally important species. However, when a sample size is sufficiently large, the multiallelic model, in which the number of disequilibrium parameters increases exponentially with the number of alleles, will be more informative than the biallelic model based on the collapsing of alleles. Technically, it is straightforward, although tedious, to model zygotic disequilibria with multiallelic markers. For example, consider two triallelic markers that each form six distinguishable genotypes. A total of 35 genotype frequencies for these two markers contain four allele frequencies, six HWD coefficients, four composite digenic disequilibria, 12 trigenic disequilibria, and nine composite quadrigenic disequilibria. Also, our zygotic model can be readily extended to manipulate three biallelic markers at the same time as seen in YANG (2000, 2002). With these extensions and modifications, the zygotic disequilibrium analysis will provide a routine tool for the identification of the overall picture of disequilibria across the genome. The results obtained from the zygotic disequilibrium model, like those for canine genetics in this study, will have important implications for the gene mapping of complex traits.

We thank Dmitri Zaykin and an anonymous reviewer for clarifying the concept of zygotic association and providing other constructive comments. The preparation of this manuscript was supported by a grant from the Morris Animal Foundation, National Institutes of Health (NIH) AR36554, the Consolidated Research Grant Program, the Cornell Advanced Technology Biotechnology Program, Nestle Purina, Marshfield Medical Research Foundation (Marshfield, WI), and Cornell University College of Veterinary Medicine unrestricted

alumni funds, NIH R01 NS041670 and National Science Foundation 0540745.

LITERATURE CITED

- ARDLIE, K. G., L. KRUGLYAK and M. SEIELSTAD, 2002 Patterns of linkage disequilibrium in the human genome. *Nat. Rev. Genet.* **3**: 299–309.
- 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, Oxford.
- BENNETT, J. H., and F. E. BINET, 1956 Association between Mendelian factors with mixed selfing and random mating. *Heredity* **10**: 51–55.
- BLISS, S., R. J. TODHUNTER, R. QUAAS, G. CASELLA, R. L. WU *et al.*, 2002 Quantitative genetics of traits associated with hip dysplasia in a canine pedigree constructed by mating dysplastic Labrador retrievers with unaffected greyhounds. *Am. J. Vet. Res.* **63**: 1029–1035.
- BREEN, M., S. JOUQUAND, C. RENIER, C. S. MELLERSH, C. HITTE *et al.*, 2001 Chromosome-specific single-locus FISH probes allow anchorage of an 1800-marker integrated radiation-hybrid/linkage map of the domestic dog genome to all chromosomes. *Mamm. Genome* **11**: 1784–1795.
- CHARLESWORTH, B., 1991 The evolution of sex chromosomes. *Science* **251**: 1030–1033.
- FARNIR, F., W. COPPIETERS, J. J. ARRANZ, P. BERZI, N. CAMBISANO *et al.*, 2000 Extensive genome-wide linkage disequilibrium in cattle. *Genome Res.* **10**: 220–227.
- HALDANE, J. B. S., 1949 The association of characters as a result of inbreeding and linkage. *Ann. Eugen.* **15**: 15–23.
- HALL, D., E. M. WIJSMAN, J. L. ROOS, J. A. GOGOS and M. KARAYIORGOU, 2002 Extended intermarker linkage disequilibrium in the Afrikaners. *Genome Res.* **12**: 956–961.
- HAMILTON, D. C., and D. E. COLE, 2004 Standardizing a composite measure of linkage disequilibrium. *Ann. Hum. Genet.* **68**: 234–239.
- HYUN, C., L. J. FILIPPICH, R. A. LEA, G. SHEPHERD, I. P. HUGHES *et al.*, 2003 Prospects for whole genome linkage disequilibrium mapping in domestic dog breeds. *Mamm. Genome* **14**: 640–649.
- LEWONTIN, R. C., 1964 The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics* **49**: 49–67.
- LEWONTIN, R. C., 1988 On measures of gametic disequilibrium. *Genetics* **120**: 849–852.
- LINDBLAD-TOH, K., C. M. WADE, T. S. MIKKELSEN, E. K. KARLSSON, D. B. JAFFE *et al.*, 2005 Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* **438**: 803–819.
- LOU, X.-Y., R. J. TODHUNTER, M. LIN, Q. LU, T. LIU *et al.*, 2003 The extent and distribution of linkage disequilibrium in canine. *Mamm. Genome* **14**: 555–564.
- LYNCH, M., and B. WALSH, 1998 *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, Sunderland, MA.
- MCRAE, A. F., J. C. MCEWAN, K. G. DODDS, T. WILSON, A. M. CRAWFORD *et al.*, 2002 Linkage disequilibrium in domestic sheep. *Genetics* **160**: 1113–1122.
- MELLERSH, C. S., A. A. LANGSTON, G. M. ACLAND, M. A. FLEMING, K. RAY *et al.*, 1997 A linkage map of the canine genome. *Genomics* **46**: 326–336.
- MELLERSH, C. S., C. HITTE, M. RICHMAN, F. VIGNAUX, C. PRIAT *et al.*, 2000 An integrated linkage-radiation hybrid map of the canine genome. *Mamm. Genome* **11**: 120–130.
- PARKER, H. G., and E. A. OSTRANDER, 2005 Canine genomics and genetics: running with the pack. *PLoS Genet.* **1**(5): e58.
- OSTRANDER, E. A., and R. K. WAYNE, 2005 The canine genome. *Genome Res.* **15**: 1706–1716.
- REMINGTON, D. L., J. M. THORNSBERRY, Y. MATSUOKADAGGER, L. M. WILSON, S. R. WHITT *et al.*, 2001 Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proc. Natl. Acad. Sci. USA* **98**: 11479–11484.
- RICHMAN, M., C. S. MELLERSH, C. ANDRE, F. GAILBERT and E. A. OSTRANDER, 2001 Characterization of a minimal screening set of 172 microsatellite markers for genome-wide screens of the canine genome. *J. Biochem. Biophys. Methods* **47**: 137–149.

- SUTTER, N. B., and E. A. OSTRANDER, 2004 Dog star rising: the canine genetic system. *Nat. Rev. Genet.* **5**: 900–910.
- SUTTER, N. B., M. A. EBERLE, H. G. PARKER, B. J. PULLAR, E. F. KIRKNESS *et al.*, 2004 Extensive and breed-specific linkage disequilibrium in *Canis familiaris*. *Genome Res.* **14**: 2388–2396.
- TENESA, A., A. F. WRIGHT, S. A. KNOTT, A. D. CAROTHERS, C. HAYWARD *et al.*, 2004 Extent of linkage disequilibrium in a Sardinian sub-isolate: sampling and methodological considerations. *Hum. Mol. Genet.* **13**: 25–33.
- TODHUNTER, R. J., G. M. ACLAND, M. OLIVIER, A. J. WILLIAMS, M. VERNIER-SINGER *et al.*, 1999 An outcrossed canine pedigree for linkage analysis of hip dysplasia. *J. Hered.* **90**: 83–92.
- TODHUNTER, R. J., G. CASELLA, S. P. BLISS, G. LUST, A. J. WILLIAMS *et al.*, 2003a Power of a dysplastic Labrador retriever-greyhound pedigree for linkage analysis of hip dysplasia. *Am. J. Vet. Res.* **64**: 418–424.
- TODHUNTER, R. J., S. R. BLISS, S. R. QUAAAS, G. LUST, G. CASELLA *et al.*, 2003b Genetic structure of susceptibility traits for hip dysplasia and microsatellite informativeness of an outcrossed canine pedigree. *J. Hered.* **94**: 39–48.
- VARILO, T., T. PAUNIO, A. PARKER, M. PEROLA, J. MEYER *et al.*, 2003 The interval of linkage disequilibrium (LD) detected with microsatellite and SNP markers in chromosomes of Finnish populations with different histories. *Hum. Mol. Genet.* **12**: 51–59.
- WEIR, B. S., 1996 *Genetic Data Analysis II*. Sinauer Associates, Sunderland, MA.
- WEIR, B. S., and C. C. COCKERHAM, 1989 Complete characterization of disequilibrium at two loci, pp. 86–110 in *Mathematical Evolutionary Theory*, edited by M. W. FELDMAN. Princeton University Press, Princeton, NJ.
- YANG, R.-C., 2000 Zygotic associations and multilocus statistics in a nonequilibrium diploid population. *Genetics* **155**: 1449–1458.
- YANG, R.-C., 2002 Analysis of multilocus zygotic associations. *Genetics* **161**: 435–445.
- ZAYKIN, D. V., 2004 Bounds and normalization of the composite linkage disequilibrium coefficient. *Genet. Epidemiol.* **27**: 252–257.

Communicating editor: M. K. UYENOYAMA

APPENDIX

In what follows, we derived the ranges of the disequilibrium parameters for a nonequilibrium population and defined the normalized zygotic disequilibrium in a way as for gametic LD (LEWONTIN 1964, 1988). On the basis of Equations 7 and 8, the ranges of the HWD coefficients are expressed as

$$\max(-p_A^2, -p_a^2) \leq D_A \leq p_A p_a,$$

for marker **A**, and

$$\max(-p_B^2, -p_b^2) \leq D_B \leq p_B p_b,$$

for marker **B**.

For the composite gametic disequilibrium, the range is derived, on the basis of Equations 9, 10, and 13, as

$$\max(-E, \dots, -J) \leq \Delta_{\mathbf{ab}} \leq \min(A, \dots, D),$$

where $A = 2p_A p_b$, $B = 2p_a p_b$, $C = p_A^2 p_b + p_a^2 p_B + p_A p_a$, $D = p_a p_B^2 + p_A p_b^2 + p_B p_b$, $E = 2p_A p_B$, $F = 2p_a p_b$, $G = p_A^2 p_B + p_a p_b^2 + p_b^2$, $H = p_A^2 p_B + p_a p_b^2 + p_b^2$, $I = p_A p_B^2 + p_a^2 p_b + p_A^2$, and $J = p_A p_B^2 + p_a^2 p_b + p_a^2$. The normalized $\Delta_{\mathbf{ab}}$ is defined as

$$\Delta'_{\mathbf{ab}} = \frac{\Delta_{\mathbf{ab}}}{\Delta_{\max(\mathbf{ab})}},$$

where

$$\Delta_{\max(\mathbf{ab})} = \begin{cases} \max(-E, \dots, -J) & \text{when } \Delta_{\mathbf{ab}} < 0 \\ \min(A, \dots, D) & \text{when } \Delta_{\mathbf{ab}} > 0. \end{cases}$$

On the basis of Equations 11 and 12, two trigenic disequilibria have the ranges expressed, respectively, as

$$\max(-H, \dots, -CC) \leq D_{\mathbf{Ab}} \leq \min(A, \dots, G),$$

where $A = 2p_A p_a p_b$, $B = p_A^2 p_B + p_A^2 + p_A p_a p_b$, $C = p_a^2 p_B + p_a^2 + p_A p_a p_b$, $D = p_A p_b^2 + p_a p_B^2 + p_A^2 p_B + p_A p_a p_b + p_B p_b$, $E = p_b^2 + p_A^2 p_B + p_A p_a p_b + p_B p_b$, $F = p_b^2 + p_a^2 p_B + p_A p_a p_b + p_B p_b$, $G = p_A p_B^2 + p_a p_B^2 + p_A^2 p_B + p_A p_a p_b + p_B p_b$, $H = p_A p_B$, $I = p_a p_B$, $J = 2p_A^2$, $K = p_A^2 p_b + p_A^2 + p_a^2 p_B$, $L = 2p_a^2$, $M = p_a^2 p_b + p_a^2 + p_A^2 p_B$, $N = p_B^2 + p_a^2 + p_b^2$, $O = p_B^2 + p_a^2 p_b + p_A^2 p_B + p_b^2$, $P = 2p_B^2 + p_a^2$, $Q = 2p_B^2 + p_a^2 p_b + p_A^2 p_B$, $R = 2p_B^2 + p_a^2$, $S = 2p_B^2 + p_a^2 p_b + p_A^2 p_B$, $T = p_B^2 + p_A^2 p_b + p_a^2 p_B + p_b^2$, $U = p_B^2 + p_a^2 + p_b^2$, $V = p_A p_b^2 + p_a p_B^2 + p_A^2 p_B + p_a^2 p_b + p_b^2$, $W = p_A p_b^2 + p_a p_B^2 + p_A^2 p_B + p_a^2 p_b + p_b^2$, $X = p_B^2 + p_A p_b^2 + p_a p_B^2 + p_a^2$, $Y = p_b^2 + p_A p_b^2 + p_a p_B^2 + p_a^2$, $Z = p_B^2 + p_A p_b^2 + p_a p_B^2 + p_a^2$, $AA = p_b^2 + p_A p_b^2 + p_a p_B^2 + p_a^2$, $BB = p_b^2 + p_A p_b^2 + p_a p_B^2 + p_a^2$, and $CC = p_B^2 + p_a^2 p_b + p_A^2 p_B + p_a p_b^2 + p_A p_B^2$, and

$$\max(-H', \dots, -CC') \leq D_{\mathbf{aB}} \leq \min(A', \dots, G'),$$

where $A' = 2p_a p_B p_b$, $B' = p_A p_b^2 + p_b^2 + p_a p_B p_b$, $C' = p_a p_B^2 + p_B^2 + p_A p_B p_b$, $D' = p_A p_b^2 + p_A p_B^2 + p_a^2 p_B + p_A p_B p_b + p_A p_a$, $E' = p_a^2 + p_A p_b^2 + p_a p_B p_b + p_A p_a$, $F' = p_a^2 + p_A p_B^2 + p_a p_B p_b + p_A p_a$, $G' = p_A^2 p_B + p_a^2 p_b + p_A p_b^2 + p_a p_B p_b + p_A p_a$, $H' = p_A p_B$, $I = p_A p_b$, $J = 2p_B^2$, $K' = p_a p_B^2 + p_B^2 + p_A p_b^2$, $L' = 2p_b^2$, $M' = p_a p_b^2 + p_b^2 + p_A p_b^2$, $N' = p_a^2 + p_a^2 + p_b^2$, $O' = p_a^2 + p_a p_b^2 + p_A p_b^2 + p_a^2$, $P' = 2p_A^2 + p_B^2$, $Q' = 2p_A^2 + p_a p_b^2 + p_A p_b^2$, $R' = 2p_A^2 + p_b^2$, $S' = 2p_A^2 + p_a p_b^2 + p_A p_b^2$, $T' = p_a^2 + p_A p_b^2 + p_a p_B^2 + p_a^2$, $U' = p_a^2 + p_a^2 + p_b^2$, $V' = p_A p_b^2 + p_a p_B^2 + p_a p_b^2 + p_a^2$, $W' = p_A^2 p_B + p_a^2 p_b + p_a p_B^2 + p_A p_b^2 + p_a^2$, $X' = p_b^2 + p_A p_B + p_a^2 p_b + p_a^2$, $Y' = p_B^2 + p_A p_B + p_a^2 p_b + p_a^2$, $Z' = p_a^2 + p_a^2 p_B + p_a^2 p_b + p_b^2$, $AA' = p_b^2 + p_A p_b^2 + p_a^2 p_B + p_a^2$, $BB' = p_a^2 + p_a^2 p_B + p_a^2 p_b + p_a p_b^2 + p_A p_B^2$, and $CC' = p_a^2 + p_A p_b^2 + p_a^2 p_B + p_a p_b^2 + p_A p_B^2$. The normalized $D_{\mathbf{Ab}}$ and $D_{\mathbf{aB}}$ are defined, respectively, as

$$D'_{\mathbf{Ab}} = \frac{D_{\mathbf{Ab}}}{D_{\max(\mathbf{Ab})}},$$

where

$$D_{\max(\mathbf{Ab})} = \begin{cases} \max(-H, \dots, -CC) & \text{when } D_{\mathbf{Ab}} < 0 \\ \min(A, \dots, G) & \text{when } D_{\mathbf{Ab}} > 0 \end{cases}$$

and

$$D'_{\mathbf{aB}} = \frac{D_{\mathbf{aB}}}{D_{\max(\mathbf{aB})}},$$

where

$$D_{\max(\mathbf{aB})} = \begin{cases} \max(-H', \dots, -CC') & \text{when } D_{\mathbf{aB}} < 0 \\ \min(A', \dots, G') & \text{when } D_{\mathbf{aB}} > 0. \end{cases}$$

On the basis of Table 3, the range of the composite quadrigenic disequilibrium is expressed as

$$\max(-E, \dots, -I) \leq \Delta_{\mathbf{AB}} \leq \min(A, \dots, D),$$

where $A = p_A^2 p_B p_b - D_A D_B + p_B p_b D_A - p_A^2 D_B + (-p_A p_B + p_A p_b) \Delta_{\mathbf{ab}} - \Delta_{\mathbf{ab}}^2 + (-p_B + p_b) D_{\mathbf{Ab}} - 2p_A D_{\mathbf{aB}}$, $B = p_A p_a p_B^2 - D_A D_B - p_B^2 D_A + p_A p_a D_B + (-p_A p_B + p_a p_B) \Delta_{\mathbf{ab}} - \Delta_{\mathbf{ab}}^2 - 2p_B D_{\mathbf{Ab}} + (-p_A + p_a) D_{\mathbf{aB}}$, $C = p_A p_a p_b^2 - D_A D_B - p_b^2 D_A + p_A p_a D_B + (p_A p_b - p_a p_b) \Delta_{\mathbf{ab}} - \Delta_{\mathbf{ab}}^2 + 2p_b D_{\mathbf{Ab}} + (-p_A + p_a) D_{\mathbf{aB}}$, $D = p_a^2 p_B p_b - D_A D_B + p_B p_b D_A - p_a^2 D_B + (p_a p_B - p_a p_b) \Delta_{\mathbf{ab}} - \Delta_{\mathbf{ab}}^2 + (-p_B + p_b) D_{\mathbf{Ab}} + 2p_a D_{\mathbf{aB}}$, $E = p_A^2 p_B^2 + D_A D_B + p_B^2 D_A + p_A^2 D_B + 2p_A p_B \Delta_{\mathbf{ab}} + \Delta_{\mathbf{ab}}^2 + 2p_B D_{\mathbf{Ab}} + 2p_A D_{\mathbf{aB}}$,

$F = p_A^2 p_b^2 + D_A D_B + p_b^2 D_A + p_A^2 D_B - 2p_A p_b \Delta_{\mathbf{ab}} + \Delta_{\mathbf{ab}}^2 - 2p_b \cdot D_{\mathbf{Ab}} + 2p_A D_{\mathbf{aB}}$, $G = p_A p_a p_B p_b + D_A D_B - p_B p_b D_A - p_A p_a D_B + \frac{1}{2}(p_A p_B + p_a p_b - p_A p_b - p_a p_B) \Delta_{\mathbf{ab}} + \Delta_{\mathbf{ab}}^2 + (p_B - p_b) D_{\mathbf{Ab}} + (p_A - p_a) D_{\mathbf{aB}}$, $H = p_a^2 p_B^2 + D_A D_B + p_B^2 D_A + p_a^2 D_B - 2p_a p_B \Delta_{\mathbf{ab}} + \Delta_{\mathbf{ab}}^2 + 2p_B D_{\mathbf{Ab}} - 2p_a D_{\mathbf{aB}}$, and $I = p_a^2 p_b^2 + D_A D_B + p_b^2 D_A + p_a^2 D_B + 2p_a p_b \Delta_{\mathbf{ab}} + \Delta_{\mathbf{ab}}^2 - 2p_b D_{\mathbf{Ab}} - 2p_a D_{\mathbf{aB}}$. The normalized $\Delta_{\mathbf{ab}}$ is defined as

$$\Delta'_{\mathbf{AB}} = \frac{\Delta_{\mathbf{AB}}}{\Delta_{\max(\mathbf{AB})}},$$

where

$$\Delta_{\max(\mathbf{AB})} = \begin{cases} \max(-E, \dots, -I) & \text{when } \Delta_{\mathbf{AB}} < 0 \\ \min(A, \dots, D) & \text{when } \Delta_{\mathbf{AB}} > 0. \end{cases}$$