

Modeling Epistasis of Quantitative Trait Loci Using Cockerham's Model

Chen-Hung Kao* and Zhao-Bang Zeng†

*Institute of Statistical Science, Academia Sinica, Taipei 11529, Taiwan, Republic of China and †Bioinformatics Research Center, Departments of Statistics and Genetics, North Carolina State University, Raleigh, North Carolina 27695-7566

Manuscript received May 10, 2001
Accepted for publication January 9, 2002

ABSTRACT

We use the orthogonal contrast scales proposed by Cockerham to construct a genetic model, called Cockerham's model, for studying epistasis between genes. The properties of Cockerham's model in modeling and mapping epistatic genes under linkage equilibrium and disequilibrium are investigated and discussed. Because of its orthogonal property, Cockerham's model has several advantages in partitioning genetic variance into components, interpreting and estimating gene effects, and application to quantitative trait loci (QTL) mapping when compared to other models, and thus it can facilitate the study of epistasis between genes and be readily used in QTL mapping. The issues of QTL mapping with epistasis are also addressed. Real and simulated examples are used to illustrate Cockerham's model, compare different models, and map for epistatic QTL. Finally, we extend Cockerham's model to multiple loci and discuss its applications to QTL mapping.

GENES interact when they express their effects; *i.e.*, the effects of genotypes at one locus depend on what genotypes are present at other loci. Interaction (epistasis) between genes affecting qualitative trait variation has been demonstrated for a long time since Gregor Mendel in 1865. Although the evidence of epistasis between genes controlling quantitative traits [quantitative trait loci (QTL)] has been reported by traditional techniques, such as variance component analyses (BRIM and COCKERHAM 1961; LEE *et al.* 1968; STUBER and MOLL 1971), epistasis between individual QTL generally has been difficult to discern by traditional techniques. The recent advances in molecular biology have allowed fine-scale genetic marker maps of various organisms to be constructed for the study of individual QTL. Using such maps, statistical methods for estimating the positions and effects of individual QTL (QTL mapping) have been proposed (LANDER and BOTSTEIN 1989; JANSEN 1993; ZENG 1994; KAO *et al.* 1999; SEN and CHURCHILL 2001). The problem of epistasis has been considered in some QTL mapping studies (*e.g.*, STUBER *et al.* 1992; CHEVERUD and ROUTMAN 1995; DOEBLEY *et al.* 1995; COCKERHAM and ZENG 1996; KAO *et al.* 1999; GOODNIGHT 2000; ZENG *et al.* 2000), but not sufficiently, and many theoretical and statistical issues involved with epistasis have not been discussed. Here, we discuss a genetic model, called Cockerham's model, in relation to QTL mapping with epistasis. We also investigate the model properties under linkage disequilibrium.

FISHER (1918) first partitioned genetic variance into

components corresponding to additive, dominance, and epistatic variances using the least-squares principle. COCKERHAM (1954) further partitioned the epistatic variance into components using orthogonal contrasts. KEMPTHONE (1957) and HAYMAN and MATHER (1955) adopted the same epistasis model. HAYMAN and MATHER (1955) and MATHER (1967) proposed other epistasis models for modeling epistasis. VAN DER VEEN (1959) reviewed the genetic models of digenic epistasis published by then and summarized them into three categories:

- The pure-line-metric or F_{∞} -metric model (F_{∞} denotes the population derived from selfing F_2 individuals for t generations as $t \rightarrow \infty$): The parameters in the F_{∞} -metric model show orthogonality with respect to genotypic frequencies of an F_{∞} population under linkage equilibrium.
- The F_2 -metric model (corresponding to Cockerham's model): The parameters in the F_2 -metric model are mutually orthogonal with respect to genotypic frequencies of an F_2 population under linkage equilibrium.
- The mixed-metric model (corresponding to Hayman and Mather's model): The mixed-metric model is a mixture of the Cockerham's model and F_{∞} -metric model, and it can be transformed to the F_2 -metric model by subtraction of the mean.

Later, CROW and KIMURA (1970), MATHER and JINKS (1982), HALEY and KNOTT (1992), and KEARSEY and POONI (1996) applied the F_{∞} -metric model to the study of epistasis between genes, and GOODNIGHT (2000) adopted an alternative model modified from COCKERHAM (1954) to study gene interaction. Although these

Corresponding author: Institute of Statistical Science, Academia Sinica, Taipei 11529, Taiwan, Republic of China.
E-mail: chkao@stat.sinica.edu.tw

three models can be translated to each other by addition or subtraction of a constant (see Table 1 of VAN DER VEEN's 1959 article), they have different meanings in interpreting gene effects, show different structures of variance components, and possess different properties in statistical estimation, which may affect the study of QTL as shown in this article.

In this article, we start from the traditional partition of genetic variance into variance components using COCKERHAM's (1954) orthogonal contrasts, then lead up to a definition of the genetic parameters for genetic effects, and present Cockerham's epistasis model. The properties of Cockerham's model in modeling and mapping epistatic genes are investigated when genes are in linkage equilibrium and disequilibrium. The differences between Cockerham's model and the other models are compared, and the advantages of Cockerham's model are discussed. It shows that Cockerham's model is a more appropriate model than the other models for the study of epistasis between genes and QTL mapping in the populations, such as F₂ and backcross. Real and simulated examples are used to illustrate Cockerham's model, compare different genetic models in the analysis of epistasis between genes, and map for epistatic QTL. Finally, we generalize Cockerham's model to multiple loci and discuss its applications to QTL mapping.

COCKERHAM'S GENETIC MODEL

COCKERHAM (1954) used eight orthogonal contrast scales to partition the genetic variance contributed by two genes into eight components and to define the genotypic value of a genotype to find the correlation between relatives in a population. His definition of genotypic value using the orthogonal scales leads the way to construct a genetic model, which is called Cockerham's model, for modeling epistasis and defining gene effects in a population. In this section, the orthogonal contrast scales are introduced to present Cockerham's model, and the genetic parameters of Cockerham's model are defined. The similarities and differences between Cockerham's model and alternative models are compared, and their variance component structures are presented.

Orthogonal contrasts: Assuming that allele frequencies at one locus are uncorrelated with frequencies at another locus (two loci are in linkage equilibrium), COCKERHAM (1954) partitioned the genetic variance caused by two loci, A and B, each with two alleles (A, a, and B, b), of a diploid organism using the orthogonal contrast scales in Table 2 of his article. The scales W'_i 's, which are functions of genotypic frequencies p_{ij} 's, have to satisfy two requirements

$$\sum_{ij} p_{ij} W_{ij} = 0 \quad \text{and} \quad \sum_{ij} p_{ij} W_{ij} W_{i'j'} = 0,$$

where i (j) indexed by 2, 1, or 0 refers to the genotype AA (BB), Aa (Bb), or aa (bb) at locus A (B), and W_{ij} is

the scale component of genotype ij for the t th contrast. The first requirement ensures that deviations around the mean are compared (the scales W_{ij} 's are contrasts). The second requirement ensures that the contrasts are orthogonal. W_1 and W_2 (W_3 and W_4) are the linear and quadratic orthogonal contrasts for locus A (locus B). W_5 is the linear \times linear contrast. W_6 is the linear \times quadratic contrast. W_7 is the quadratic \times linear contrast. W_8 is the quadratic \times quadratic contrast. Cockerham's orthogonal contrast scales serve the same purpose as the orthogonal contrasts for partitioning the sum of squares due to treatment into independent single-degree-of-freedom components in experimental design (STEEL and TORRIE 1981). The statistical linear and quadratic terms correspond to the genetical additive and dominance terms, respectively. Cockerham used these orthogonal scales to partition the genetic variance and find the partition of variance σ_i^2 due to orthogonal scale W_i by

$$\sigma_i^2 = \frac{(\sum_{ij} p_{ij} G_{ij} W_{ij})^2}{(\sum_{ij} p_{ij} W_{ij}^2)},$$

where G_{ij} denotes the genotypic value of the genotype ij . He also defined G in terms of the scales as

$$G_{ij} = E_0 + \sum_{t=1}^8 E_t W_{ij}, \tag{1}$$

where E_t 's are the corresponding coefficients, by solving the equations themselves, and used it to find the correlation between relatives in a population. His idea of defining the genotypic value by the orthogonal contrast scales leads up to Cockerham's genetic model for modeling epistasis between genes.

Cockerham's genetic model: We now apply Cockerham's orthogonal contrast scales to the F₂ population to derive Cockerham's model for the F₂ population. For an F₂ population, the genotypic frequencies of the nine genotypes *AABB*, *AABb*, *AAbb*, *AaBB*, *AaBb*, *Aabb*, *aaBB*, *aaBb*, and *aabb* are 1/16, 1/8, 1/16, 1/8, 1/4, 1/8, 1/16, 1/8, and 1/16, respectively, and Cockerham's orthogonal contrasts can be modified as shown in Table 1 (see also COCKERHAM and ZENG 1996). By solving Equation 1 with the scales in Table 1, the unique solutions of the coefficients in terms of the genotypic values are

$$E_0 = \frac{G_{22}}{16} + \frac{G_{21}}{8} + \frac{G_{20}}{16} + \frac{G_{12}}{8} + \frac{G_{11}}{4} + \frac{G_{10}}{8} + \frac{G_{02}}{16} + \frac{G_{01}}{8} + \frac{G_{00}}{16}, \tag{2}$$

$$E_1 = \frac{G_{22}}{8} + \frac{G_{21}}{4} + \frac{G_{20}}{8} - \frac{G_{02}}{8} - \frac{G_{01}}{4} - \frac{G_{00}}{8}, \tag{3}$$

$$E_2 = \frac{G_{12}}{8} + \frac{G_{11}}{4} + \frac{G_{10}}{8} - \frac{G_{22}}{16} - \frac{G_{21}}{8} - \frac{G_{20}}{16} - \frac{G_{02}}{16} - \frac{G_{01}}{8} - \frac{G_{00}}{16}, \tag{4}$$

TABLE 1
The eight orthogonal contrast scales (W 's) for the F_2 population

Genotype	<i>AABB</i>	<i>AABb</i>	<i>AAbb</i>	<i>AaBB</i>	<i>AaBb</i>	<i>Aabb</i>	<i>aaBB</i>	<i>aaBb</i>	<i>aabb</i>
<i>G</i>	G_{22}	G_{21}	G_{20}	G_{12}	G_{11}	G_{10}	G_{02}	G_{01}	G_{00}
<i>P</i>	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{16}$
W_1	1	1	1	0	0	0	-1	-1	-1
W_2	$-\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$
W_3	1	0	-1	1	0	-1	1	0	-1
W_4	$-\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{2}$
W_5	1	0	-1	0	0	0	-1	0	1
W_6	$-\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{2}$	0	0	0	$\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{2}$
W_7	$-\frac{1}{2}$	0	$\frac{1}{2}$	$\frac{1}{2}$	0	$-\frac{1}{2}$	$-\frac{1}{2}$	0	$\frac{1}{2}$
W_8	$\frac{1}{4}$	$-\frac{1}{4}$	$\frac{1}{4}$	$-\frac{1}{4}$	$\frac{1}{4}$	$-\frac{1}{4}$	$\frac{1}{4}$	$-\frac{1}{4}$	$\frac{1}{4}$

G's and *P*'s denote the genotypic values and expected genotypic frequencies for the nine genotypes of two unlinked genes, A and B.

$$E_3 = \frac{G_{22}}{8} + \frac{G_{12}}{4} + \frac{G_{02}}{8} - \frac{G_{20}}{8} - \frac{G_{10}}{4} - \frac{G_{00}}{8}, \quad (5)$$

$$E_4 = \frac{G_{21}}{8} + \frac{G_{11}}{4} + \frac{G_{01}}{8} - \frac{G_{22}}{16} - \frac{G_{12}}{8} - \frac{G_{02}}{16} - \frac{G_{20}}{16} - \frac{G_{10}}{8} - \frac{G_{00}}{16}, \quad (6)$$

$$E_5 = \frac{(G_{22} - G_{02}) - (G_{20} - G_{00})}{4} = \frac{(G_{22} - G_{20}) - (G_{02} - G_{00})}{4}, \quad (7)$$

$$E_6 = \frac{(2G_{21} - G_{22} - G_{20}) - (2G_{01} - G_{02} - G_{00})}{4}, \quad (8)$$

$$E_7 = \frac{(2G_{12} - G_{22} - G_{02}) - (2G_{10} - G_{20} - G_{00})}{4}, \quad (9)$$

$$E_8 = \frac{2(2G_{11} - G_{21} - G_{01}) - (2G_{12} - G_{22} - G_{02}) - (2G_{10} - G_{20} - G_{00})}{4} = \frac{2(2G_{11} - G_{12} - G_{10}) - (2G_{21} - G_{22} - G_{20}) - (2G_{01} - G_{02} - G_{00})}{4}. \quad (10)$$

If the two genes are in linkage equilibrium, E_0 is the mean of the genotypic values, $\bar{G}_{..}$, and therefore can be denoted as μ . Coefficient E_1 is equivalent to $(\bar{G}_2 - \bar{G}_0)/2$, which is one-half of the difference in genotypic value between the two homozygote means of AA and aa and thus is defined as the genetic parameter of additive effects of gene A, a_1 . Coefficient E_2 is equivalent to $(2\bar{G}_1 - \bar{G}_2 - \bar{G}_0)/2$, which represents the departure in genotypic value of the heterozygote mean of Aa from the midpoint between the two homozygote means of

AA and aa and thus is defined as the genetic parameter of dominance effect of gene A, d_1 . The same argument leads us to define coefficients E_3 and E_4 as the genetic parameters of additive and dominance effects of gene B, a_2 and d_2 . If the substitution effects at one locus depend on genotypes at the other locus, there is an interaction between the two genes in the usual sense. Coefficient E_5 quantifies the difference between additive effects of gene A (gene B), $(G_{2*} - G_{0*})/2 [(G_{*2} - G_{*0})/2]$, in the background of two different homozygotes of gene B (gene A), BB and bb (AA and aa), and this difference is defined as the genetic parameter of additive \times additive epistatic effect, i_{aa} . The larger the difference is, the stronger the interaction is. The same argument leads to the definitions of E_6 , E_7 , and E_8 as the genetic parameters of additive \times dominance, i_{ad} ; dominance \times additive, i_{da} ; and dominance \times dominance, i_{dd} ; epistatic effects between genes A and B. The definitions of these nine genetic parameters are summarized in Table 2. After defining the genetic parameters of genetic effects, Equation 1 can be expressed more succinctly as

$$G_{ij} = \mu + a_1x_1 + d_1z_1 + a_2x_2 + d_2z_2 + i_{aa}w_{aa} + i_{ad}w_{ad} + i_{da}w_{da} + i_{dd}w_{dd}, \quad (11)$$

by defining the coded variables as

$$x_1 = \begin{cases} 1 & \text{if A is AA} \\ 0 & \text{if A is Aa} \\ -1 & \text{if A is aa,} \end{cases} \quad x_2 = \begin{cases} 1 & \text{if B is BB} \\ 0 & \text{if B is Bb} \\ -1 & \text{if B is bb,} \end{cases}$$

$$z_1 = \begin{cases} \frac{1}{2} & \text{if A is Aa} \\ -\frac{1}{2} & \text{otherwise,} \end{cases} \quad z_2 = \begin{cases} \frac{1}{2} & \text{if B is Bb} \\ -\frac{1}{2} & \text{otherwise,} \end{cases}$$

$$w_{aa} = x_1 \times x_2, \quad w_{ad} = x_1 \times z_2, \quad w_{da} = z_1 \times x_2, \quad w_{dd} = z_1 \times z_2.$$

The coded variables of this model are mutually independent to each other due to orthogonality. The model

TABLE 2
Definition of genetic parameters

Solution	Parameter definition	Notation
E_0	Mean	μ
E_1	Additive effect of locus A	a_1
E_2	Dominance effect of locus A	d_1
E_3	Additive effect of locus B	a_2
E_4	Dominance effect of locus B	d_2
E_5	Additive \times additive effect of loci A and B	i_{aa}
E_6	Additive \times dominance effect of loci A and B	i_{ad}
E_7	Dominance \times additive effect of loci A and B	i_{da}
E_8	Dominance \times dominance effect of loci A and B	i_{dd}

E_i 's are the solutions of Equation 1 with the orthogonal contrast scales in Table 2. The exact expressions of E_i 's are shown in Equations 2–10.

can also be represented in a different form as Table 3. Note that the marginal means of the three genotypes, \bar{G}_2 , \bar{G}_1 , and \bar{G}_0 , for locus A are $\mu + a_1 - d_1/2$, $\mu + d_1/2$, and $\mu - a_1 - d_1/2$, respectively, as the segregation ratio is 1:2:1. There are similar forms for locus B. The grand mean \bar{G} is equivalent to μ .

Genetic variance structure: When applying Cockerham's model to modeling genotypic values in a popula-

tion, the structure of variance components for the total genetic variance, V_G , contributed by the two genes, each with two alleles, is shown in APPENDIX C. From APPENDIX C, we can see that the total genetic variance is composed of genetic variance of individual effects and covariances between different effects, and it will change with gene frequencies (p 's) and linkage disequilibrium (D). Certainly, the relative strengths of genetic effects will vary according to the change in gene frequency and linkage disequilibrium. For an F_2 population ($p_A = p_B = 0.5$), the total genetic variance reduces to Equation 34 and contains covariances between different genetic effects through linkage. If genes are unlinked in the F_2 population ($p_A = p_B = 0.5$ and $D = 0$), the total genetic variance can be partitioned into eight independent components without covariance as

$$V_G = \frac{1}{2}a_1^2 + \frac{1}{4}d_1^2 + \frac{1}{2}a_2^2 + \frac{1}{4}d_2^2 + \frac{1}{4}i_{aa}^2 + \frac{1}{8}i_{ad}^2 + \frac{1}{8}i_{da}^2 + \frac{1}{16}i_{dd}^2. \tag{12}$$

Each variance component is contributed by its own genetic parameter. For example, the additive variance component of gene A, $a_1^2/2$, is contributed by its additive effect, a_1 , and it has no genetic covariance with other effects. This property greatly facilitates the evaluation of the contribution of an effect to the genetic variance. The other models, such as F_∞ -metric and mixed-metric models, do not have such a property (see Equation 18).

Linkage disequilibrium: The coded variables in Cocker-

TABLE 3
Cockerham's model (the F_2 -metric model)

	AA	Aa	aa	
	G_{22}	G_{12}	G_{02}	\bar{G}_2
<i>BB</i>	$\mu + a_1 - \frac{d_1}{2} + a_2 - \frac{d_2}{2} + i_{aa} - \frac{i_{ad}}{2} - \frac{i_{da}}{2} + \frac{i_{dd}}{4}$	$\mu + \frac{d_1}{2} + a_2 - \frac{d_2}{2} + \frac{i_{da}}{2} - \frac{i_{dd}}{4}$	$\mu - a_1 - \frac{d_1}{2} + a_2 - \frac{d_2}{2} - i_{aa} + \frac{i_{ad}}{2} - \frac{i_{da}}{2} + \frac{i_{dd}}{4}$	$\mu + a_2 - \frac{d_2}{2}$
	G_{21}	G_{11}	G_{01}	\bar{G}_1
<i>Bb</i>	$\mu + a_1 - \frac{d_1}{2} + \frac{d_2}{2} + \frac{i_{ad}}{2} - \frac{i_{dd}}{4}$	$\mu + \frac{d_1}{2} + \frac{d_2}{2} + \frac{i_{dd}}{4}$	$\mu - a_1 - \frac{d_1}{2} + \frac{d_2}{2} - \frac{i_{ad}}{2} - \frac{i_{dd}}{4}$	$\mu + \frac{d_2}{2}$
	G_{20}	G_{10}	G_{00}	\bar{G}_0
<i>bb</i>	$\mu + a_1 - \frac{d_1}{2} - a_2 - \frac{d_2}{2} - i_{aa} - \frac{i_{ad}}{2} + \frac{i_{da}}{2} + \frac{i_{dd}}{4}$	$\mu + \frac{d_1}{2} - a_2 - \frac{d_2}{2} - \frac{i_{da}}{2} - \frac{i_{dd}}{4}$	$\mu - a_1 - \frac{d_1}{2} - a_2 - \frac{d_2}{2} + i_{aa} + \frac{i_{ad}}{2} + \frac{i_{da}}{2} + \frac{i_{dd}}{4}$	$\mu - a_2 - \frac{d_2}{2}$
	\bar{G}_2	\bar{G}_1	\bar{G}_0	\bar{G}
	$\mu + a_1 - \frac{d_1}{2}$	$\mu + \frac{d_1}{2}$	$\mu - a_1 - \frac{d_1}{2}$	μ

The marginal means \bar{G}_i (\bar{G}_j) for locus A (B) are calculated under segregation ratio 1:2:1 for AA (*BB*), Aa (*Bb*), and aa (*bb*) in the F_2 population. The genetic parameters a_1 , d_1 , a_2 , d_2 , i_{aa} , i_{ad} , i_{da} , and i_{dd} are defined in Table 2.

TABLE 4
The F_{∞} -metric model

	<i>AA</i>	<i>Aa</i>	<i>aa</i>	
	G_{22}	G_{12}	G_{02}	$\bar{G}_{.2}$
<i>BB</i>	$\mu + a_1 + a_2 + i_{aa}$	$\mu + d_1 + a_2 + i_{da}$	$\mu - a_1 + a_2 - i_{aa}$	$\mu + a_2 + \frac{d_1}{2} + \frac{i_{da}}{2}$
	G_{21}	G_{11}	G_{01}	$\bar{G}_{.1}$
<i>Bb</i>	$\mu + a_1 + d_2 + i_{ad}$	$\mu + d_1 + d_2 + i_{dd}$	$\mu - a_1 + d_2 - i_{ad}$	$\mu + \frac{d_1}{2} + d_2 + \frac{i_{dd}}{2}$
	G_{20}	G_{10}	G_{00}	$\bar{G}_{.0}$
<i>bb</i>	$\mu + a_1 - a_2 - i_{aa}$	$\mu + d_1 - a_2 - i_{da}$	$\mu - a_1 - a_2 + i_{aa}$	$\mu + \frac{d_1}{2} - a_2 - \frac{i_{da}}{2}$
	$\bar{G}_{.2}$	$\bar{G}_{.1}$	$\bar{G}_{.0}$	$\bar{G}_{.}$
	$\mu + a_1 + \frac{d_2}{2} + \frac{i_{ad}}{2}$	$\mu + d_1 + \frac{d_2}{2} + \frac{i_{dd}}{2}$	$\mu - a_1 + \frac{d_2}{2} - \frac{i_{ad}}{2}$	$\mu + \frac{d_1}{2} + \frac{d_2}{2} + \frac{i_{dd}}{4}$

The marginal means \bar{G}_i (\bar{G}_j) for locus A (B) are calculated under segregation ratio 1:2:1 for *AA* (*BB*), *Aa* (*Bb*), and *aa* (*bb*). a_1 (a_2) and d_1 (d_2) are the additive and dominance effects of locus A (B). i_{aa} , i_{ad} , i_{da} , and i_{dd} are the additive-by-additive, additive-by-dominance, dominance-by-additive, and dominance-by-dominance epistatic effects.

ham's model (the scales in Table 1) are orthogonal and contrast to each other when the ratio of genotypic frequencies is 1:2:1:2:4:2:1:2:1 (genes are unlinked) in an F_2 population. Therefore, the definition of the genetic parameters in Table 2 is appropriate for interpreting the gene effects and the genetic variance can be partitioned (Equation 12) as if genes are unlinked. If there is segregation distortion and/or linkage, the ratio will deviate from 1:2:1:2:4:2:1:2:1 (Table 6) and there will be covariances between some genetic effects (Equation 34). To take linkage disequilibrium into account in using Cockerham's model, we introduce statistical parameters to contrast with genetic parameters in interpreting gene effects when genes are in linkage disequilibrium (see next section).

F_{∞} -metric and mixed-metric models: The F_{∞} -metric model can be expressed in equation form as Equation 11 by coding

$$x_1 = \begin{cases} 1 & \text{if A is AA} \\ 0 & \text{if A is Aa} \\ -1 & \text{if A is aa,} \end{cases} \quad x_2 = \begin{cases} 1 & \text{if B is BB} \\ 0 & \text{if B is Bb} \\ -1 & \text{if B is bb,} \end{cases}$$

$$z_1 = \begin{cases} 1 & \text{if A is Aa} \\ 0 & \text{otherwise,} \end{cases} \quad z_2 = \begin{cases} 1 & \text{if B is Bb} \\ 0 & \text{otherwise,} \end{cases}$$

and $w_{aa} = x_1 \times x_2$, $w_{ad} = x_1 \times z_2$, $w_{da} = z_1 \times x_2$, and $w_{dd} = z_1 \times z_2$, where the coded variables for epistasis are just the products of marginal variables. Equivalently, the F_{∞} -metric model can be illustrated by Table 4. It is easy to check that the coded variables of the F_{∞} -metric model do not have the property of orthogonal contrast. Also, the marginal means of one locus are involved in the genetic parameter of another locus and their epistasis parameters, and the difference in genotypic values

between the two homozygotes is not equal to the genetic parameter of additive effect a_1 (a_2). For example, $\bar{G}_{.2} = (\mu + a_1 + d_2/2 + i_{ad}/2)$, $(\bar{G}_{.2} - \bar{G}_{.0}) = a_1 + i_{ad}/2$, and $\bar{G}_{.} = \mu + d_1/2 + d_2/2 + i_{dd}/4$ (Table 4). This result deviates from the usual definition in the one-locus analysis. The solutions of the marginal genetic parameters, a_1 , d_1 , a_2 , d_2 , in terms of the genotypic values for the F_{∞} -metric model are

$$\mu = \frac{G_{22} + G_{20} + G_{02} + G_{00}}{4}, \tag{13}$$

$$a_1 = \frac{(G_{22} + G_{20}) - (G_{02} + G_{00})}{4}, \tag{14}$$

$$d_1 = \frac{2(G_{12} - G_{10}) - (G_{22} + G_{20}) - (G_{02} + G_{00})}{4}, \tag{15}$$

$$a_2 = \frac{(G_{22} + G_{02}) - (G_{20} + G_{00})}{4}, \tag{16}$$

$$d_2 = \frac{2(G_{21} - G_{01}) - (G_{22} + G_{02}) - (G_{20} + G_{00})}{4}, \tag{17}$$

and the solutions of epistasis genetic parameters, i_{aa} , i_{ad} , i_{da} , and i_{dd} , are the same as those in Cockerham's model. Apparently, most of the heterozygotes are excluded in the estimation of μ and marginal parameters, making the F_{∞} -metric model difficult in interpreting the gene action for the F_2 population.

The equation form for the mixed-metric model, which is a mixture of Cockerham's model and the F_{∞} -metric model with the first part of marginal effects from the F_{∞} -metric model and the latter part of epistatic effects from Cockerham's model, is trivial (not shown), and it is tabulated in Table 5. The coded variables of the mixed-metric model are orthogonal, but not contrasts. Except

TABLE 5
The mixed-metric model

	AA	Aa	aa	
	G_{22}	G_{12}	G_{02}	\bar{G}_2
<i>BB</i>	$\mu + a_1 + a_2 + i_{aa} - \frac{i_{ad}}{2} - \frac{i_{da}}{2} + \frac{i_{dd}}{4}$	$\mu + d_1 + a_2 + \frac{i_{da}}{2} - \frac{i_{dd}}{4}$	$\mu - a_1 + a_2 - i_{aa} + \frac{i_{ad}}{2} - \frac{i_{da}}{2} + \frac{i_{dd}}{4}$	$\mu + a_2 + \frac{d_1}{2}$
	G_{21}	G_{11}	G_{01}	\bar{G}_1
<i>Bb</i>	$\mu + a_1 + d_2 + \frac{i_{ad}}{2} - \frac{i_{dd}}{4}$	$\mu + d_1 + d_2 + \frac{i_{dd}}{4}$	$\mu + a_1 + d_2 - \frac{i_{ad}}{2} - \frac{i_{dd}}{4}$	$\mu + d_2 + \frac{d_1}{2}$
	G_{20}	G_{10}	G_{00}	\bar{G}_0
<i>bb</i>	$\mu + a_1 - a_2 - i_{aa} - \frac{i_{ad}}{2} + \frac{i_{da}}{2} + \frac{i_{dd}}{4}$	$\mu + d_1 - a_2 - \frac{i_{da}}{2} - \frac{i_{dd}}{4}$	$\mu - a_1 - a_2 + i_{aa} + \frac{i_{ad}}{2} + \frac{i_{da}}{2} + \frac{i_{dd}}{4}$	$\mu - a_2 + \frac{d_1}{2}$
	\bar{G}_2	\bar{G}_1	\bar{G}_0	$\bar{G}_.$
	$\mu + a_1 + \frac{d_2}{2}$	$\mu + d_1 + \frac{d_2}{2}$	$\mu - a_1 + \frac{d_2}{2}$	$\mu + \frac{d_1}{2} + \frac{d_2}{2}$

The marginal means \bar{G}_i (\bar{G}_j) for loci A (B) are calculated under segregation ratio 1:2:1 for AA (*BB*), Aa (*Bb*), and aa (*bb*). a_1 (a_2) and d_1 (d_2) are the additive and dominance effects of locus A (B). i_{aa} , i_{ad} , i_{da} , and i_{dd} are the additive-by-additive, additive-by-dominance, dominance-by-additive, and dominance-by-dominance epistatic effects.

for μ , the solutions of the genetic parameters of the mixed-metric model are the same as those of Cockerham's model. The solution of μ is not equal to $\bar{G}_.$. By subtracting $d_1/2 + d_2/2$, the mixed-metric model will become Cockerham's model. In Table 5, the marginal means of one locus involve the dominance effect of another locus, which deviates from the one-locus analysis. For example, the marginal mean of genotype AA, \bar{G}_2 , is $\mu + a_1 + d_2/2$. Except for μ , the solutions of the genetic parameters of the mixed-metric model are the same as those of Cockerham's model.

As the F_∞ -metric model is not an orthogonal model, the total genetic variance contributed by two genes in linkage equilibrium is

$$\begin{aligned}
 V_G = & \frac{1}{2}a_1^2 + \frac{1}{4}d_1^2 + \frac{1}{2}a_2^2 + \frac{1}{4}d_2^2 + \frac{1}{4}i_{aa}^2 + \frac{1}{4}i_{ad}^2 + \frac{1}{4}i_{da}^2 \\
 & + \frac{3}{16}i_{dd}^2 + \frac{1}{2}a_1i_{ad} + \frac{1}{2}a_2i_{da} - \frac{1}{4}d_1i_{aa} \\
 & - \frac{1}{4}d_2i_{aa} + \frac{1}{4}d_1i_{dd} + \frac{1}{4}d_2i_{dd}, \tag{18}
 \end{aligned}$$

which consists of the covariances between marginal and epistatic gene effects. These covariances make the evaluation of the contribution of an individual effect to the total genetic variance difficult. The genetic variance structure of the mixed-metric model is the same as that of Cockerham's model. Note that the genetic variance structures of Cockerham's model and the F_∞ -metric model cannot be translated to each other by adding or subtracting a constant value, and therefore they are different models from this point.

MODELING QUANTITATIVE TRAITS

When applying Cockerham's model to analyze a quantitative trait, controlled by two epistatic genes A and B, from a sample of size n of an F_2 population, the trait value of the k th individual with genotype ij can be modeled as

$$\begin{aligned}
 y_{ijk} = & G_{ij} + \varepsilon_{ijk} \\
 = & \mu + a_1x_1 + d_1z_1 + a_2x_2 + d_2z_2 + i_{aa}w_{aa} \\
 & + i_{ad}w_{ad} + i_{da}w_{da} + i_{dd}w_{dd} + \varepsilon_{ijk}, \tag{19}
 \end{aligned}$$

where ε_{ijk} is a residual. Let \bar{P}_{ij} and n_{ij} denote the observed frequency and sample size of genotype ij where $n_{ij} = n \times \bar{P}_{ij}$. In expectation, $E(n_{ij}) = n \times P_{ij}$ and $E(y_{ij}) = n \times P_{ij} \times G_{ij}$, where P_{ij} is the population frequency of genotype ij and depends on the linkage strength between genes (Table 6). Note that the ratio of P_{ij} 's reduces to 1:2:1:2:4:2:1:2:1 if genes are unlinked ($D = 0$).

Least-squares estimates of genetic parameters: The least-squares estimates (LSE) of the genetic parameters in Equation (19) have similar formulations as those of Equations 2–10 except that G_{ij} is replaced with \bar{y}_{ij} . For example, the LSE of a_1 is

$$\hat{a}_1 = \frac{\bar{y}_{22}}{8} + \frac{\bar{y}_{21}}{4} + \frac{\bar{y}_{20}}{8} - \frac{\bar{y}_{02}}{8} - \frac{\bar{y}_{01}}{4} - \frac{\bar{y}_{00}}{8}. \tag{20}$$

When genes are unlinked (the segregation ratio is 1:2:1:2:4:2:1:2:1), the expectation of \hat{a}_1 is

$$E(\hat{a}_1) = \frac{\bar{G}_2 - \bar{G}_0}{2}, \tag{21}$$

which corresponds to the additive effect of gene A.

TABLE 6

Genotypic frequencies (P 's) in terms of allele frequencies (p 's) and the linkage disequilibrium coefficient (D)

	AA	Aa	aa	Total
	P_{22}	P_{12}	P_{02}	P_2
BB	$(p_A p_B + D)^2$	$2(p_A p_B + D) \times (p_a p_b - D)$	$(p_a p_b - D)^2$	p_B^2
	P_{21}	P_{11}	P_{01}	P_1
Bb	$2(p_A p_b + D) \times (p_A p_b - D)$	$2(p_A p_b - D)(p_a p_b - D) + 2(p_A p_b + D) \times (p_a p_b + D)$	$2(p_a p_b - D) \times (p_a p_b + D)$	$2p_b p_b$
	P_{20}	P_{10}	P_{00}	P_0
bb	$(p_A p_b - D)^2$	$2(p_A p_b - D)(p_a p_b + D)$	$(p_a p_b + D)^2$	p_b^2
Total	$P_2 \cdot p_A^2$	$P_1 \cdot 2p_A p_a$	$P_0 \cdot p_a^2$	1

$p_A, p_a, p_B,$ and p_b denote the frequencies of alleles A, a, B, and b of genes A and B. P_{ij} denotes the genotypic frequencies. D is the linkage disequilibrium between genes A and B.

However, when genes are linked, \hat{a}_1 is not a measure of the difference between the two homozygote means as the ratio is no longer 1:2:1:2:4:2:1:2:1. Likewise, the LSE of the genetic parameters are appropriate estimates of the nine gene effects when genes are unlinked, but they are not appropriate estimates when genes are linked. To remedy this problem, statistical parameters of gene effects are introduced for interpretation in contrast to genetic parameters of gene effects.

Statistical parameters of gene effects: When the derivatives of the error sum of squares in Equation 19 with respect to every genetic parameter in turn are set equal to zero, it gives the nine normal equations. For example, the normal equation with respect to a_1 is

$$\begin{aligned}
 y_{2..} - y_{0..} &= (n_{22} + n_{21} + n_{20} - n_{02} - n_{01} - n_{00})\hat{\mu} \\
 &+ (n_{22} + n_{21} + n_{20} + n_{02} + n_{01} + n_{00})\hat{a}_1 \\
 &+ (-n_{22} - n_{21} - n_{20} + n_{02} + n_{01} + n_{00})\frac{\hat{d}_1}{2} \\
 &+ (n_{22} - n_{20} - n_{02} + n_{00})\hat{a}_2 \\
 &+ (-n_{22} + n_{21} - n_{20} + n_{02} - n_{01} + n_{00})\frac{\hat{d}_2}{2} \\
 &+ (n_{22} - n_{20} + n_{02} - n_{00})\hat{i}_{aa} \\
 &+ (-n_{22} + n_{21} - n_{20} - n_{02} + n_{01} - n_{00})\frac{\hat{i}_{ad}}{2} \\
 &+ (-n_{22} + n_{20} + n_{02} - n_{00})\frac{\hat{i}_{da}}{2} \\
 &+ (n_{22} - n_{21} + n_{20} - n_{02} + n_{01} - n_{00})\frac{\hat{i}_{dd}}{4}. \quad (22)
 \end{aligned}$$

By taking expectation, the expected normal equations can be obtained and expressed in terms of genotypic values G_{ij} 's, population genotypic frequencies P_{ij} 's, and genetic parameters E 's, as shown from Equations A1–A9 (APPENDIX A). For simplicity, the left-hand sides of these

nine expected normal equations are denoted as $\beta_0, \beta_1, \dots, \beta_8$. Then, Equation A1 can be written as

$$\begin{aligned}
 \beta_0 &= P_{22}G_{22} + P_{21}G_{21} + P_{20}G_{20} + P_{12}G_{12} + P_{11}G_{11} \\
 &+ P_{10}G_{10} + P_{02}G_{02} + P_{01}G_{01} + P_{00}G_{00},
 \end{aligned}$$

which is the mean genotypic value in the population. Equation A2 can be written as

$$\beta_1 = \frac{P_{22}G_{22} + P_{21}G_{21} + P_{20}G_{20} - P_{02}G_{02} - P_{01}G_{01} - P_{00}G_{00}}{P_{22} + P_{21} + P_{20} + P_{02} + P_{01} + P_{00}},$$

and it can be reformulated as

$$\begin{aligned}
 \beta_1 &= \frac{(P_{22}G_{22} + P_{21}G_{21} + P_{20}G_{20}) / (P_{22} + P_{21} + P_{20})}{2} \\
 &- \frac{(P_{02}G_{02} + P_{01}G_{01} + P_{00}G_{00}) / (P_{02} + P_{01} + P_{00})}{2} \\
 &= \frac{\bar{G}_2 - \bar{G}_0}{2}
 \end{aligned}$$

since $P_{22} + P_{21} + P_{20} = P_{02} + P_{01} + P_{00} = \frac{1}{4}$ in the F_2 population. That is, β_1 quantifies one-half of the difference in genotypic value between the two homozygote means of gene A; *i.e.*, β_1 is a quantity to measure the additive effect of gene A, no matter whether genes are in linkage equilibrium or not. Further, as the genotypic frequencies of gene A (B) have relationship $2P_2 = 2P_0 = P_1 = \frac{1}{2}(2P_2 = 2P_0 = P_1 = \frac{1}{2})$ in the F_2 population despite linkage,

$$\begin{aligned}
 \beta_2 &= 2(P_{12}G_{12} + P_{11}G_{11} + P_{10}G_{10} - P_{22}G_{22} - P_{21}G_{21} \\
 &- P_{20}G_{20} - P_{02}G_{02} - P_{01}G_{01} - P_{00}G_{00}) \\
 &= \frac{2(P_{12}G_{12} + P_{11}G_{11} + P_{10}G_{10}) / (P_{12} + P_{11} + P_{10})}{2} \\
 &- \frac{(P_{22}G_{22} + P_{21}G_{21} + P_{20}G_{20}) / (P_{22} + P_{21} + P_{20})}{2}
 \end{aligned}$$

$$\begin{aligned} & - \frac{(P_{02}G_{02} + P_{01}G_{01} + P_{00}G_{00})/(P_{02} + P_{01} + P_{00})}{2} \\ & = \frac{2\bar{G}_1 - \bar{G}_2 - \bar{G}_0}{2}, \end{aligned} \quad (23)$$

which measures the dominance effect of gene A. For epistasis parameters,

$$\begin{aligned} \beta_5 & = \frac{(P_{22}G_{22} - P_{20}G_{20}) - (P_{02}G_{02} - P_{00}G_{00})}{P_{22} + P_{20} + P_{02} + P_{00}} \\ & = \frac{(P_{22}G_{22} - P_{02}G_{02}) - (P_{20}G_{20} - P_{00}G_{00})}{P_{22} + P_{20} + P_{02} + P_{00}}, \end{aligned}$$

which is a weighted version of the additive \times additive epistasis. When genes are unlinked, β_5 reduces to the genetic parameter i_{aa} . When genes are linked, the genetic parameter i_{aa} is still valid for the additive \times additive effect since marginal means are not involved in it. Similarly, the genetic parameters i_{ad} , i_{da} , and i_{dd} are still appropriate to measure the additive \times dominance, dominance \times additive, and dominance \times dominance effects under linkage disequilibrium, and β_6 , β_7 , and β_8 are weighted versions of the epistatic effects, and they all reduce to i_{ad} , i_{da} , and i_{dd} if genes are in linkage equilibrium.

Given genotypic values G 's, the quantities, β 's, will have different values according to different strengths of linkage (ratios of genotypic frequencies). On the contrary, the genetic parameters, E 's, will not change according to different strengths of linkage. Therefore, we define β 's as statistical parameters to contrast with the genetic parameters of gene effects. The genetic parameters can be obtained directly from Cockerham's model, but the statistical parameters cannot. However, there exists a one-to-one relationship between the two kinds of parameters as shown below. It allows that once the genetic parameters are obtained from the model the statistical parameters can be obtained by transformation.

Relationship between genetic and statistical parameters: In a population, the frequency of the gamete AB , P_{AB} , can be expressed in terms of allele frequencies (p 's) and the linkage disequilibrium coefficient D (WEIR 1996) as

$$P_{AB} = p_A p_B + D,$$

where D is equivalent to $(1 - 2r)/4$ (r is the recombination fraction between loci A and B). If the union of gametes is random, the genotypic frequencies P_{ij} 's are products of gametic frequencies (Table 6). The expected normal equations from Equations A1–A9 can be further expressed in terms of the genetic parameters (E 's), the statistical parameters (β 's), the population allele frequencies (p 's), and the linkage disequilibrium coefficient D as shown in Equations B1–B9 in APPENDIX B. In the F_2 population, the allele frequencies p_A , p_a , p_B , and p_b are one-half, and the nine expected normal equations in APPENDIX B reduce to the following:

$$\beta_0 = \mu + \frac{\lambda}{2}i_{aa} + \lambda^2\frac{i_{dd}}{4}, \quad (24)$$

$$\beta_1 = a_1 + \lambda a_2 - \lambda^2\frac{i_{ad}}{2} - \lambda\frac{i_{da}}{2}, \quad (25)$$

$$\beta_3 = \lambda a_1 + a_2 - \lambda\frac{i_{ad}}{2} - \lambda^2\frac{i_{da}}{2} \quad (26)$$

$$\frac{\beta_2}{2} = \frac{d_1}{2} + \lambda^2\frac{d_2}{2} - \frac{\lambda}{2}i_{aa}, \quad (27)$$

$$\frac{\beta_4}{2} = \lambda^2\frac{d_1}{2} + \frac{d_2}{2} - \frac{\lambda}{2}i_{aa}, \quad (28)$$

$$\beta_5 = \frac{4}{1 + \lambda^2} \left[\frac{\lambda}{2}\mu - \frac{\lambda}{2}\frac{d_1}{2} - \frac{\lambda}{2}\frac{d_2}{2} + \left(\frac{1 + \lambda^2}{4}\right)i_{aa} + \frac{\lambda}{2}\frac{i_{dd}}{4} \right], \quad (29)$$

$$\frac{\beta_6}{2} = 2 \left(-\frac{\lambda^2}{2}a_1 - \frac{\lambda}{2}a_2 + \frac{1}{2}\frac{i_{ad}}{2} + \frac{\lambda}{2}\frac{i_{da}}{2} \right), \quad (30)$$

$$\frac{\beta_7}{2} = 2 \left(-\frac{\lambda}{2}a_1 - \frac{\lambda^2}{2}a_2 + \frac{\lambda}{2}\frac{i_{ad}}{2} + \frac{1}{2}\frac{i_{da}}{2} \right), \quad (31)$$

$$\frac{\beta_8}{4} = \lambda^2\mu + \frac{\lambda}{2}i_{aa} + \frac{i_{dd}}{4}, \quad (32)$$

where $\lambda = 1 - 2r$. The statistical parameters (β 's) are functions of the genetic parameters (E 's) and a linkage parameter (λ), and vice versa. The approximation of the genetic parameter to its corresponding statistical parameter depends on the strength of linkage and the sizes of other genetic parameters. In matrix equation, the above equations can be also expressed as

$$B = T E, \quad (33)$$

where

$$B' = \left[\beta_0 \quad \beta_1 \quad \frac{\beta_2}{2} \quad \beta_3 \quad \frac{\beta_4}{2} \quad \frac{1 + \lambda^2}{4} \beta_5 \quad \frac{\beta_6}{2} \quad \frac{\beta_7}{2} \quad \frac{\beta_8}{4} \right]$$

contains the statistical parameters,

$$E' = \left[\mu \quad a_1 \quad \frac{d_1}{2} \quad a_2 \quad \frac{d_2}{2} \quad i_{aa} \quad \frac{i_{ad}}{2} \quad \frac{i_{da}}{2} \quad \frac{i_{dd}}{4} \right]$$

contains the genetic parameters, and

$$T = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & \frac{\lambda}{2} & 0 & 0 & \lambda^2 \\ 0 & 1 & 0 & \lambda & 0 & 0 & -\lambda^2 & -\lambda & 0 \\ 0 & 0 & 1 & 0 & \lambda^2 & -\frac{\lambda}{2} & 0 & 0 & 0 \\ 0 & \lambda & 0 & 1 & 0 & 0 & -\lambda & -\lambda^2 & 0 \\ 0 & 0 & \lambda^2 & 0 & 1 & -\frac{\lambda}{2} & 0 & 0 & 0 \\ \frac{\lambda}{2} & 0 & -\frac{\lambda}{2} & 0 & -\frac{\lambda}{2} & \frac{1 + \lambda^2}{4} & 0 & 0 & \frac{\lambda}{2} \\ 0 & -\lambda^2 & 0 & -\lambda & 0 & 0 & 1 & \lambda & 0 \\ 0 & -\lambda & 0 & -\lambda^2 & 0 & 0 & \lambda & 1 & 0 \\ \lambda^2 & 0 & 0 & 0 & 0 & \frac{\lambda}{2} & 0 & 0 & 1 \end{bmatrix}$$

is a symmetric and nonsingular matrix with components associated with the linkage parameter λ . The inverse of T is

$$T^{-1} = \frac{1}{(1 - \lambda^2)^2}$$

$$\times \begin{bmatrix} \frac{1}{1 + \lambda^2} & 0 & \frac{-\lambda^2}{1 + \lambda^2} & 0 & \frac{-\lambda^2}{1 + \lambda^2} & -2\lambda & 0 & 0 & \frac{\lambda^4}{1 + \lambda^2} \\ 0 & 1 & 0 & -\lambda & 0 & 0 & -\lambda^2 & \lambda & 0 \\ \frac{-\lambda^2}{1 + \lambda^2} & 0 & \frac{1}{1 + \lambda^2} & 0 & \frac{\lambda^4}{1 + \lambda^2} & 2\lambda & 0 & 0 & \frac{-\lambda^2}{1 + \lambda^2} \\ 0 & -\lambda & 0 & 1 & 0 & 0 & \lambda & -\lambda^2 & 0 \\ \frac{-\lambda^2}{1 + \lambda^2} & 0 & \frac{\lambda^4}{1 + \lambda^2} & 0 & \frac{1}{1 + \lambda^2} & 2\lambda & 0 & 0 & \frac{-\lambda^2}{1 + \lambda^2} \\ -2\lambda & 0 & 2\lambda & 0 & 2\lambda & 4(1 + \lambda^2) & 0 & 0 & -2\lambda \\ 0 & -\lambda^2 & 0 & \lambda & 0 & 0 & 1 & -\lambda & 0 \\ 0 & \lambda & 0 & -\lambda^2 & 0 & 0 & -\lambda & 1 & 0 \\ \frac{\lambda^4}{1 + \lambda^2} & 0 & \frac{-\lambda^2}{1 + \lambda^2} & 0 & \frac{-\lambda^2}{1 + \lambda^2} & -2\lambda & 0 & 0 & \frac{1}{1 + \lambda^2} \end{bmatrix}$$

(WOLFRAM 1992). The two kinds of parameters have a one-to-one relationship. When genes are in linkage equilibrium ($\lambda = 0$), T is diagonal, and β 's are equal to E 's. When genes are not in linkage equilibrium ($\lambda \neq 0$), they are different, but transferable.

Random mating: Linkage disequilibrium decays after random mating. If the F_2 progeny are further randomly mated, linkage disequilibrium is mitigated by a factor $1 - r$, $0 < r < 0.5$, gradually in each generation. The general formula of the linkage disequilibrium coefficient in generation F_t under random mating is

$$\lambda_t = (1 - r)^{t-2} \lambda,$$

where $t \geq 2$ is the number of generations. As t gets larger, λ_t approaches zero; *i.e.*, linkage equilibrium will be gradually attained in later generations by random mating. After random mating, λ_t changes (becomes smaller), as do the genotypic frequencies (P_{ij} 's), and accordingly the statistical parameters (β 's) change and become closer to the genetic parameters (E 's). Therefore, the statistical parameters (β 's) depend on the population frequencies (P_{ij} 's) and will have different values in different generations. When λ_t approaches 0, the ratio of the genotypic frequencies approaches 1:2:1:2:1:2:1, and the statistical parameters (β 's) will approach the genetic parameters (E 's). Hence, the genetic parameters of genes in linkage disequilibrium estimated in the F_2 population can be regarded as the gene effects in later generations when linkage equilibrium is attained.

Variance components: The genetic variance contributed by two genes in the F_2 population is

$$\begin{aligned} \text{Var}(G) &= \frac{1}{2}a_1^2 + \frac{1}{4}d_1^2 + \frac{1}{2}a_2^2 + \frac{1}{4}d_2^2 + \frac{1}{4}i_{aa}^2 + \frac{1}{8}i_{ad}^2 + \frac{1}{8}i_{da}^2 \\ &+ \frac{1}{16}[1 - \lambda^4]i_{dd}^2 + \lambda a_1 a_2 - \frac{\lambda^2}{2}a_1 i_{ad} - \frac{\lambda}{2}a_1 i_{da} \\ &+ \frac{\lambda^2}{2}d_1 d_2 - \frac{\lambda}{2}d_1 i_{aa} - \frac{\lambda^2}{2}a_2 i_{ad} - \frac{\lambda^2}{2}a_2 i_{da} - \frac{\lambda}{2}d_2 i_{aa} \end{aligned}$$

$$+ \frac{\lambda}{4}[1 - \lambda^2]i_{aa}i_{dd} + \frac{\lambda}{4}i_{ad}i_{da} \tag{34}$$

(APPENDIX C). The genetic variance is composed of the variances and covariances of genetic parameters. If genes are in linkage equilibrium or attain equilibrium in later generations by random mating ($\lambda = 0$), the covariances disappear and the genetic variance will be partitioned into eight independent components (Equation 12).

QTL MAPPING USING COCKERHAM'S MODEL

In this section, we apply Cockerham's model to construct a statistical epistasis model to map for epistatic QTL and analyze epistasis between QTL. The problems when epistasis is present and ignored in QTL mapping are also investigated. By taking epistasis into account in QTL mapping, the accuracy of estimation and power of detection can be improved.

Mapping epistatic QTL: Assume that a quantitative trait y is controlled by two interacting QTL, Q_A and Q_B , located at positions p_1 and p_2 , in two different intervals, I_1 and I_2 . The statistical QTL mapping model can be written as

$$\begin{aligned} y_i &= \mu + a_1 x_1^* + d_1 z_1^* + a_2 x_2^* + d_2 z_2^* + i_{aa} x_1^* x_2^* \\ &+ i_{ad} x_1^* z_2^* + i_{da} z_1^* x_2^* + i_{dd} z_1^* z_2^* + \varepsilon_i, \\ i &= 1, 2, \dots, n, \end{aligned} \tag{35}$$

where ε_i follows $N(0, \sigma^2)$, and the codes for variables x_1^* (x_2^*) and z_1^* (z_2^*) follow the codes of x_1 (x_2) and z_1 (z_2) in Cockerham's model (Equation 11). As Q_A (Q_B) is not located at the marker, its genotypes, *i.e.*, the value of x_1^* and z_1^* (x_2^* and z_2^*), are not observable. However, through its flanking markers, the conditional genotypic distribution of Q_A (Q_B) can be inferred on the basis of Haldane's mapping function (HALDANE 1919) as listed in Table 2 of KAO and ZENG (1997). The joint conditional genotypic distribution of Q_A and Q_B in intervals I_1 and I_2 can be obtained using the property of conditional independence between them (KAO and ZENG 1997). Let p_{ij} 's, $j = 1, 2, \dots, 9$, denote the conditional probabilities of the nine possible QTL genotypes for individual i . The likelihood of the statistical model is a mixture of nine normals as

$$L(\mu, a_1, d_1, a_2, d_2, i_{aa}, i_{ad}, i_{da}, i_{dd}, \sigma^2 | Y, X, Z) = \prod_{i=1}^n [\sum_{j=1}^9 p_{ij} N(\mu_{ij}, \sigma^2)], \tag{36}$$

where p_{ij} 's and μ_{ij} 's are the mixing proportions and genotypic values of the nine genotypes for individual i . To obtain the maximum-likelihood estimates (MLE) of the genetic parameters and their asymptotic variance-covariance matrix for the normal mixture model, the general formulas by KAO and ZENG (1997) based on the expectation-maximization (EM) algorithm (DEMPSTER

et al. 1977) can be used. The general formulas are based on two matrices, the genetic design matrix D and the conditional probability matrix Q . Here, the genetic design matrix is a matrix with dimension 9×8 as

$$D = [W_1 \ W_2 \ W_3 \ W_4 \ W_5 \ W_6 \ W_7 \ W_8], \quad (37)$$

where W_i 's, $i = 1, 2, \dots, 9$, are the orthogonal contrast scales of Cockerham's model in Table 1, and the conditional probability matrix Q is a $9^2 \times 3^2$ matrix with elements associated with the mixing proportions. By applying the matrices D and Q to the general formulas, the MLE and the asymptotic variance-covariance matrix can be obtained.

The proposed statistical QTL mapping model in Equation 35 can be used to search for epistatic QTL as well as to analyze epistasis between QTL by taking epistasis into account. In QTL mapping, we usually first search for QTL by ignoring epistasis. When epistasis is ignored, the accuracy in estimation and power of detection could be affected (see below). Thus, it is very likely that the detected epistatic QTL are those with relatively large marginal effects and the undetected epistatic QTL are those with relatively minor marginal effects. By taking epistasis into account, Equation 35 can be used to search for the undetected minor epistatic QTL by testing hypotheses

$$\begin{aligned} H_0: a_2 = d_2 = i_{aa} = i_{ad} = i_{da} = i_{dd} = 0; \\ H_1: \text{at least one of them is not } 0; \end{aligned} \quad (38)$$

given the detected QTL with marginal effects a_1 and d_1 in the model. Note that hypotheses in (38) can consider only additive effect and a part of the four epistatic effects in testing. Alternatively, Equation 35 can be used to test for the existence of epistasis between two detected QTL by setting hypotheses

$$\begin{aligned} H_0: i_{aa} = i_{ad} = i_{da} = i_{dd} = 0; \\ H_1: \text{at least one of them is not } 0; \end{aligned} \quad (39)$$

given their marginal effects in the model. Certainly, the hypotheses in (39) can contain individual epistasis parameters in the analysis. The hypotheses in (38) and (39) can be tested using the likelihood-ratio test (LRT) statistic,

$$\text{LRT} = -2 \log \frac{L_0}{L_1},$$

where L_0 and L_1 are the likelihoods under H_0 and H_1 . The critical value of the LRT statistic for rejecting H_0 can be chosen from χ^2 distribution on the basis of the Bonferroni argument.

What are the problems if epistasis is present and ignored? Although epistasis is an ubiquitous phenomenon (WRIGHT 1980), many QTL mapping methods ignore epistasis in the analysis for simplicity. It is important to investigate the problems if epistasis is present and ignored and further to solve the problems and analyze epi-

stasis in QTL mapping. When the quantitative trait affected by the two epistatic QTL, Q_A and Q_B , is regressed on a marker M along the genome to infer QTL, under Cockerham's model, the regression coefficient for the additive effect of M is

$$\begin{aligned} a_M = (1 - 2r_{AM})a_1 + (1 - 2r_{BM})a_2 \\ - \frac{1}{2}(1 - 2r_{AB})(1 - 2r_{BM})i_{ad} \\ - \frac{1}{2}(1 - 2r_{AB})(1 - 2r_{AM})i_{da}, \end{aligned} \quad (40)$$

where r_{AM} , r_{BM} , and r_{AB} are the recombination fractions between Q_A and M , Q_B and M , and Q_A and Q_B , respectively, and the regression coefficient for the dominance effect is

$$d_M = (1 - 2r_{AM})^2d_1 + (1 - 2r_{BM})^2d_2 - (1 - 2r_{AM})(1 - 2r_{BM})i_{aa}. \quad (41)$$

If the marker M is coincident with Q_A , the coefficient a_M , which reduces to the estimate of additive effect of Q_A , is confounded by the additive effect of Q_B and their epistatic effects, i_{ad} and i_{da} , via linkage, and the coefficient d_M , which is the estimate of dominance effect of Q_A , is confounded by the dominance effect of Q_B and their epistatic effects, i_{aa} , via linkage. When the quantitative trait is regressed on both Q_A and Q_B , the partial regression coefficient for the additive effect of Q_A , given the additive effect of Q_B is

$$a_{A.B_a} = a_1 - \frac{1}{2}(1 - 2r_{AB})i_{da}, \quad (42)$$

and the partial regression coefficient for the dominance effect of Q_A given the dominance effect of Q_B is

$$d_{A.B_d} = d_1 - \frac{(1 - 2r_{AB})}{1 + (1 - 2r_{AB})^2}i_{aa}. \quad (43)$$

Again, the partial regression coefficients $a_{A.B_a}$ and $d_{A.B_d}$ are confounded by their epistasis, i_{da} and i_{aa} , respectively, via linkage. If Q_A and Q_B are unlinked ($r_{AB} = 0.5$), the confounding of epistasis disappears and the coefficients (Equations 40–43) are all unbiased for a_1 and d_1 . It implies that if epistasis between QTL is present and ignored in QTL mapping, the estimation of the marginal effects and positions of QTL are asymptotic unbiased if the epistatic QTL are unlinked. But, if the epistatic QTL are linked, the estimates of QTL positions and marginal effects are biased and confounded by epistatic effects via linkage. This unbiasedness property for unlinked QTL attributes to the orthogonal property of Cockerham's model. The approaches of interval mapping (LANDER and BOTSTEIN 1989; JANSEN 1993; ZENG 1994; KAO *et al.* 1999), which test every position within marker intervals along the entire genome for QTL detection, share the same problems and properties under Cockerham's model.

The similar investigation on the problems if epistasis

is present and ignored in QTL mapping can also be done for the F_{∞} -metric and mixed-metric models. Under the F_{∞} -metric model, the regression coefficient for the additive effect of a marker M is

$$\begin{aligned} a_M &= (1 - 2r_{AM})a_1 + (1 - 2r_{BM})a_2 \\ &+ \frac{1}{2}(1 - 2r_{AM})[1 - (1 - 2r_{BM})^2]i_{ad} \\ &+ \frac{1}{2}(1 - 2r_{BM})[1 - (1 - 2r_{AM})^2]i_{da}, \quad (44) \end{aligned}$$

and the regression coefficient for the dominance effect of a marker M is

$$\begin{aligned} d_M &= (1 - 2r_{AM})^2d_1 + (1 - 2r_{BM})^2d_2 \\ &- (1 - 2r_{AM})(1 - 2r_{BM})i_{aa} \\ &+ \frac{1}{2}[(1 - 2r_{AM})^2 + (1 - 2r_{BM})^2]i_{dd}. \quad (45) \end{aligned}$$

If the marker M is coincident with Q_A , the coefficients, a_M and d_M , reduce to the estimates of the additive and dominance effects of Q_A . The estimate of the additive effect of Q_A is confounded by the additive effect of Q_B , a_2 , and their epistatic effects, i_{ad} and i_{da} , and the estimate of dominance effect of Q_A is confounded by the dominance effect of Q_B and their epistatic effects, i_{aa} and i_{dd} . When the quantitative trait is regressed on both Q_A and Q_B , the partial regression coefficient for the additive effect of Q_A given the additive effect of Q_B is

$$a_{A.B_a} = a_1 + \frac{1}{2}i_{ad} - \frac{1}{2}(1 - 2r_{AB})i_{da}, \quad (46)$$

and the partial regression coefficient for the dominance effect of Q_A given the dominance effect of Q_B is

$$d_{A.B_d} = d_1 - \frac{1 - 2r_{AB}}{1 + (1 - 2r_{AB})^2}i_{aa} + \frac{1}{2}i_{dd}. \quad (47)$$

Again, the partial regression coefficients of the additive and dominance effects are confounded by epistatic effects, and they are biased estimates of the additive and dominance effects. If $r_{AB} = 0.5$, the four coefficients in Equations 44–47 are still biased. For example, when $r_{AB} = 0.5$, $a_A = a_1 + i_{ad}/2$ and $a_{A.B_a} = a_1 + i_{ad}/2$, which are all biased estimates of a_1 . Therefore, the F_{∞} -metric model always has the problems of confounding and is biased in estimation if epistasis is present and ignored whether the QTL are linked or not. This implies that QTL mapping could be problematic for the F_{∞} -metric model if epistasis is ignored. As the mixed-metric model is also orthogonal, it possesses the same properties as those of Cockerham's model in the QTL analysis.

When epistasis is present and ignored in QTL mapping, the genetic variance contributed by epistasis is not controlled in the model and becomes a part of the genetic residual. Thus, the sampling variances of the effects are inflated, and the power of detecting QTL

decreases. If epistasis is taken into account, the epistatic variance can be controlled, and the power will increase. The increase in power depends on the size of the epistatic effect. The larger the epistatic effect that can be controlled in mapping, the larger the increase in power that can be gained. In conclusion, by taking epistasis into account in QTL mapping, the chance of finding more QTL and the accuracy of estimating QTL positions and effects can be improved.

ADVANTAGES OF COCKERHAM'S MODEL

Cockerham's model has several advantages in the study of epistasis as compared to the F_{∞} -metric and mixed-metric models. When genes are in linkage equilibrium, the advantages include the following:

1. The genetic variance can be partitioned into eight independent components (Equation 12), and there is no genetic covariance. Each component is contributed by its corresponding genetic parameter. This is a desirable property in modeling. On the contrary, the F_{∞} -metric does not have such a property (Equation 18).
2. The marginal means of one locus do not involve the parameters of another locus and the epistasis parameters, which would make Cockerham's model readily interpretable (Table 3). The marginal means of locus A are $(a_1 - d_1/2)$, $d_1/2$, and $(-a_1 - d_1/2)$, which correspond to the one-locus analysis (differing by a constant $d_1/2$) despite epistasis. For the F_{∞} -metric model, the marginal means of locus A are $(a_1 + d_2/2 + i_{ad}/2)$, $(d_1 + d_2/2 + i_{ad}/2)$, and $(-a_1 + d_2/2 - i_{ad}/2)$, which are confounded by the genetic parameter of dominance effect of locus B (d_2) and their epistasis parameters, i_{ad} and i_{dd} . In the mixed-metric model, the marginal means of locus A are $a_1 + d_2/2$, $d_1 + d_2/2$, and $-a_1 + d_2/2$, which are confounded by the genetic parameters of dominance effect of locus B (d_2). Both the F_{∞} -metric and mixed-metric models do not follow the definition in the one-locus analysis.
3. The difference between the two homozygote means, $(G_2 - G_0)/2[(G_2 - G_0)/2]$, estimates the genetic parameter a_1 (a_2) of locus A (B), and the departure of the heterozygote mean to the midpoint between the two homozygote means, $(2G_1 - G_2 - G_0)/2[(2G_1 - G_2 - G_0)/2]$, estimates the genetic parameter d_1 (d_2) of locus A (B). They follow the same definition of additive and dominance effects in the one-locus analysis. In the F_{∞} -metric model, they estimate $a_1 + i_{ad}/2$ ($a_2 + i_{da}/2$) and $d_1 + i_{dd}/2$ ($d_2 + i_{dd}/2$) and violate the definition in the one-locus analysis.
4. With the orthogonal property, the estimation of one genetic (marginal or epistatic) effect will not be affected by the presence or absence of other genetic effects in the model. Essentially, when epistasis is

TABLE 7
The means of trait LBIL in the F₂ population
from DOEBLEY *et al.* (1995)

	Q _A			Mean
	AA	Aa	aa	
Q _B	101.6 ^a	83.62	47.80	77.21
BB	8 ^b	20	11	39
	66.50	47.55	54.57	54.19
Bb	22	42	21	85
	61.11	40.94	17.98	36.37
bb	3	24	10	37
	74.52	54.09	44.08	55.67
Mean	33	86	42	161

LBIL, average length of vegetative internodes in the primary lateral branch. Q_A and Q_B represent unlinked genes UMC107 and BV302, respectively.

^a Trait mean.

^b Sample size.

present and ignored, the estimation of the marginal effects and the location of epistatic QTL is still asymptotically unbiased and not affected by epistasis. This advantage ensures that QTL mapping can be first performed without taking epistasis into account without causing a problem under Cockerham's model. The F_{∞} -metric model does not have such property (see QTL MAPPING USING COCKERHAM'S MODEL).

EXAMPLES

In this section, real and simulated data were used to illustrate Cockerham's model, compare the differences between Cockerham's model and other models, verify the properties in statistical estimation, and map for epistatic QTL.

Real data: DOEBLEY *et al.* (1995) crossed two corn inbred lines, *Teosinte-MIL* × *Teosinte-M3L*, to generate 183 F₂ progeny, and they concluded that two unlinked markers UMC107 (Q_A) and BV302 (Q_B) are the candidate QTL for trait LBIL (average length of vegetative internodes in the primary lateral branch) in QTL analysis. Among the 183 progeny, 21 individuals have a missing trait and one individual has a missing genotype. Therefore, only the 161 individuals with complete trait and genotype information were used in the analysis. The observed allele frequencies are $\hat{p}_A = 0.4720$, $\hat{p}_a = 0.5280$, $\hat{p}_B = 0.5062$, and $\hat{p}_b = 0.4938$. The genotypic frequencies are 0.050, 0.137, 0.019, 0.124, 0.261, 0.149, 0.068, 0.130, and 0.062 for genotypes *AABB*, *AABb*, *AAbb*, *AaBB*, *AaBb*, *Aabb*, *aaBB*, *aaBb*, and *aabb*, respectively, which significantly deviate from the expected frequencies for two unlinked genes. The small sample size of *Aabb* in 3 individuals is responsible for the deviation.

The observed genotypic means (\bar{y}_{ij} 's) and sample sizes (n_{ij} 's) of the data are listed in Table 7. If all eight genetic

parameters (full model) are considered, the estimated genetic parameters by Cockerham's model, the F_{∞} -metric model, and the mixed-metric model are listed in Table 9. In Table 9, except for μ , the estimates of the eight genetic parameters by Cockerham's model and the mixed-metric model are the same. Cockerham's model and the F_{∞} -metric model have different estimates of marginal effects, but the same estimates of epistatic effects (see Cockerham's genetic model for the reasons). The estimates of a_1 and d_1 are 15.11 (P value 0.0008) and -3.92 (P value 0.5035), respectively, for Cockerham's model, and they are 24.25 (P value 0.0008) and 5.15 (P value 0.5617), respectively, for the F_{∞} -metric model. The estimates of a_2 and d_2 are 19.46 (P value 0.0001) and -5.66 (P value 0.3336), respectively, for Cockerham's model, and they are 17.59 (P value 0.0001) and 3.40 (P value 0.3336), respectively, for the F_{∞} -metric model. Very likely, the marginal effects of Q_A and Q_B are mostly additive, and their dominance effects are not significant. The estimate of i_{aa} is 2.68 (P value 0.7054). Analytically, it means that the additive effects of Q_B (Q_A) in the background of AA (BB) and aa (bb), which are $(\bar{Y}_{22} - \bar{Y}_{20})/2 = 20.27$ [$(\bar{Y}_{22} - \bar{Y}_{02})/2 = 26.93$] and $(\bar{Y}_{02} - \bar{Y}_{00})/2 = 14.91$ [$(\bar{Y}_{20} - \bar{Y}_{00})/2 = 21.57$], differ by 2.68, and this difference is not statistically significant at the 5% level (Figure 1a). The estimate of i_{ad} is -18.28 (P value 0.0411). Analytically, it means that the dominance effects of Q_B in the background of AA and aa, which are 21.68 [$(2\bar{Y}_{01} - \bar{Y}_{02} - \bar{Y}_{00})/2$] and -14.88 [$(2\bar{Y}_{21} - \bar{Y}_{22} - \bar{Y}_{20})/2$], are significantly different at the 5% level. The significance of additive-by-dominance interaction can be illustrated by Figure 1b. In Figure 1b, the cross between the two lines tells that genotype *Bb* performs better than *BB* in the background of *aa*, but it does worse in the background of *AA*. The estimate of i_{da} , 3.75, is not significant (P value 0.6725) as illustrated by the three nearly parallel lines in Figure 1c. The estimate of i_{dd} , -18.13 , is not statistically significant at the 5% level (P value 0.1227), although it shows that there is a cross between lines in Figure 1d. The proportion of the genetic variance in the total variance (model R^2) is 23.66% (Table 8).

The estimates of the statistical parameters are $\hat{\beta}_0 = 55.67$, $\hat{\beta}_1 = 8.10$, $\hat{\beta}_2 = 4.24$, $\hat{\beta}_3 = 21.91$, $\hat{\beta}_4 = 3.10$, $\hat{\beta}_5 = 8.59$, $\hat{\beta}_6 = 0.71$, $\hat{\beta}_7 = -7.52$, and $\beta_8 = -38.88$ following the definitions, or they can be obtained by using Equations A1–A9 by plugging in observed genotypic frequencies in Table 7 and the nine estimated genetic parameters in Table 9. Although the values of the statistical and genetic parameters are expected to be very close for unlinked genes, they are very different based on this data set. The difference occurs because the observed segregation ratio deviates from the expected segregation ratio.

If only the significant effects, a_1 , a_2 , and i_{ad} (reduced model), are considered for Cockerham's model, the estimates of a_1 , a_2 , and i_{ad} are 15.27 (SD 4.13, P value

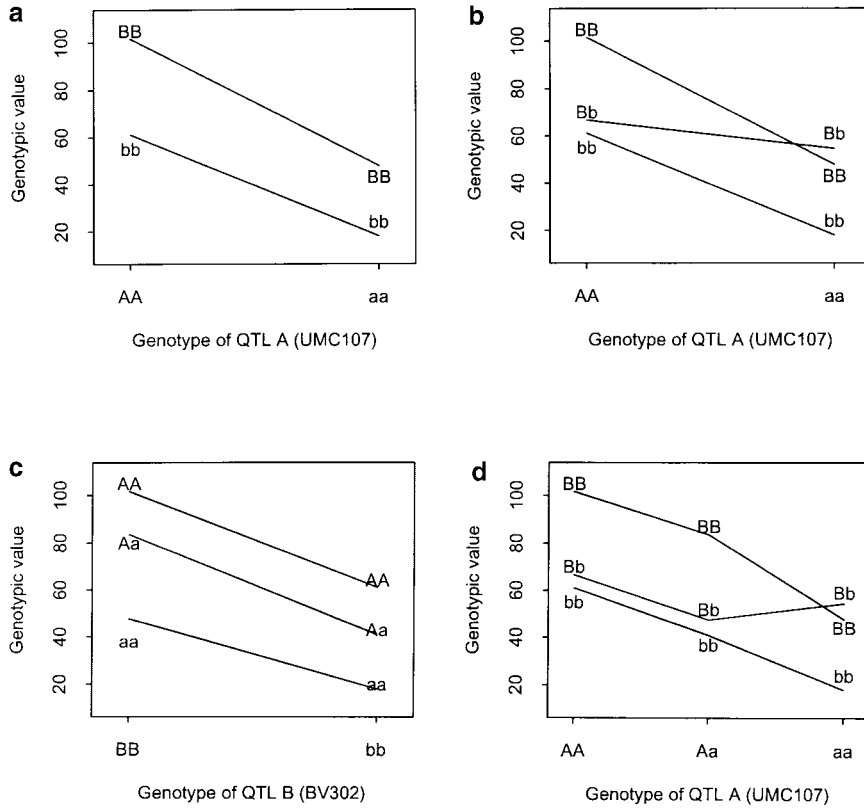


FIGURE 1.—Epistasis plot of the four types of epistasis from the data of DOEBLEY *et al.* (1995) in Table 7. (a) Additive-by-additive epistasis. (b) Additive-by-dominance epistasis. (c) Dominance-by-additive epistasis. (d) Dominance-by-dominance epistasis.

0.0003), 19.13 (SD 4.04, P value < 0.0001), and -18.44 (SD 8.27, P value < 0.0271), respectively, which are very close to the estimates in the full model. This shows that the estimation of one genetic parameter in Cockerham's model will not be affected by the presence or absence of other genetic parameters due to its orthogonal property. However, the F_{∞} -metric model does not have such a property. If the reduced model is considered for the F_{∞} -metric model, the estimates of a_1 , a_2 , and i_{ad} become 24.48 (SD 6.28, P value 0.0001), 19.12 (SD 4.04, P value < 0.0001), and -18.44 (SD 8.27, P value 0.0271), respectively. The estimate of a_2 changes from 17.59 in the full model to 19.12 in the reduced model due to the confounding of $i_{da}/2 = 3.75/2$ (estimated in the full model) by Equation 46. Both models have the same model $R^2 = 0.2121$.

If only the significant effects are taken into account

in calculating the variance components, the additive effect of Q_A , a_1 , contributes $\sim 34.05\%$ to the total genetic variance (Equation 12), the additive effect of Q_B , a_2 , contributes $\sim 52.04\%$ to the total genetic variance, and the epistatic effect, i_{ad} , contributes $\sim 13.90\%$ to the total genetic variance under Cockerham's model. There is no genetic covariance between effects for unlinked loci. The mixed-metric model has the same genetic variance structure as Cockerham's model. The genetic variance and covariance components under the F_{∞} -metric model can be obtained using Equation 18.

Simulation: Assume that a quantitative trait is affected by two unlinked epistatic QTL. The first QTL, Q_A , is located at 52 cM on the first chromosome, and the second QTL, Q_B , is located at 93 cM on the second chromosome. There are 11 15-cM equally spaced markers on each chromosome. The additive and dominance effects

TABLE 8
Two-way ANOVA of DOEBLEY *et al.*'s (1995) data in Table 7

Source	d.f.	Sum of square	Mean square	F value	P value
Q_A	2	16995.08	8497.54	6.89	0.0014
Q_B	2	19227.48	9613.74	7.80	0.0006
$Q_A \times Q_B$	4	10921.70	2730.42	2.21	0.0701
Error	152	187440.72	1233.16		
Total	160	245527.72			

R -square is 0.2366.

TABLE 9
Results from analysis of DOEBLEY *et al.*'s (1995) data in Table 7 by Cockerham's model, the F_{∞} -metric model, and the mixed-metric model

	Parameter estimate		Standard error		Test for parameter = 0		P value	
	Cockerham	F_{∞}	Cockerham	F_{∞}	Cockerham	F_{∞}	Cockerham	F_{∞}
μ	56.87	57.13	2.92	7.07	19.48	8.08	0.0001	0.0001
a_1	15.11	24.25	4.47	7.07	3.41	3.30	0.0008	0.0008
d_1	-3.92	5.15	5.84	8.05	-0.67	0.58	0.5035	0.5617
a_2	19.46	17.59	4.42	7.07	4.40	2.49	0.0001	0.0140
d_2	-5.66	3.40	5.84	8.87	-0.97	0.38	0.3336	0.7022
i_{aa}	2.68	2.68	7.07	7.07	0.38	0.38	0.7054	0.7054
i_{ad}	-18.28	-18.28	8.87	8.87	-2.06	-2.06	0.0411	0.0411
i_{da}	3.75	3.75	8.85	8.85	0.42	0.42	0.6725	0.6725
i_{dd}	-18.13	18.13	11.68	11.68	-1.55	-1.55	0.1227	0.1227

Except for μ , the mixed-metric model gives the same estimates of the parameters a_1 , d_1 , a_2 , d_2 , i_{aa} , i_{ad} , i_{da} , and i_{dd} as those of Cockerham's model. The estimate of μ by the mixed-metric model is 61.66 with SD 5.79.

of Q_A are $a_1 = 3$ and $d_1 = 1$. Q_B has additive effect $a_2 = 1$ and no dominance effect. The additive-by-additive epistatic effect is $i_{aa} = 2$, and the other three epistatic effects are assumed to be zero. With these parameter settings, the marginal effects of Q_A and Q_B contribute 76 and 8% to the total genetic variance, and epistasis contributes 16% to the total genetic variance. The heritability of the quantitative trait is assumed to be 0.2, or equivalently the environmental variance is 25. The sample size is 200, and the number of simulated replicates is 500. When using the statistical model in Equation 35 for QTL mapping, a stepwise selection procedure (KAO *et al.* 1999) was adopted to detect QTL and analyze epistasis, and the critical value for claiming significance was chosen as $\chi^2_{k,0.05/20}$, where k is the number of parameters in testing.

The simulation results are shown in Table 10. When epistasis is ignored in QTL mapping, the powers to detect Q_A and Q_B are 1.0 and 0.238 ($\chi^2_{1,0.05/20} = 9.14$; $\chi^2_{2,0.05/20} = 11.98$), respectively. The mean estimates of positions of Q_A and Q_B are 51.25 with standard deviation (SD) 7.73 and 89.63 with SD 24.19, respectively. The means of the estimated additive and dominance effects of Q_A are 2.9941 (SD 0.5969) and 0.9816 (SD 0.9018). The mean estimate of the additive effect of Q_B (from significant replicates) is 1.8567 (SD 0.4196), which is poorly estimated. If the mean of the estimated effect of Q_B is calculated on the basis of all 500 replicates, it is 1.1214 (SD 0.7956), which is much closer to the true value. This corresponds to the theoretical proof of asymptotical unbiasedness for marginal effect in estimation if epistasis is present and ignored under Cock-

TABLE 10
Simulation results of mapping epistatic QTL in the F_2 population

	Without epistasis ^a		With epistasis ^b	
	Q_A	Q_B	Q_A	Q_B
Power	1.000	0.238	1.000	0.500
Position	51.25 (7.73)	89.63 (24.19)	50.99 (7.95)	90.46 (18.67)
		0.0125 (0.3701)		0.0091 (0.3635)
a	2.9941 (0.5696)	1.8567 (0.4196)	2.9658 (0.5700)	1.3314 (0.6368)
d	0.9816 (0.9018)		1.0024 (0.8697)	
i_{aa}				1.9897 (0.9948)
σ^2	24.26 (3.29)		24.02 (3.50)	
h^2	0.2158 (0.0445)		0.2257 (0.0494)	

Numbers in parentheses are standard deviations. The critical values for claiming significance are $\chi^2_{1,0.05/20} = 9.14$ and $\chi^2_{2,0.05/20} = 11.98$. Two unlinked QTL, Q_A and Q_B , are simulated on the chromosomes with 11 15-cM equally spaced markers. Q_A is placed at 52 cM with additive effect 3 and dominance effect 1. Q_B is placed at 93 cM with additive effect 1 ($a_1 = 3$, $d_1 = 1$, $a_2 = 1$; $V_A/V_G = 76\%$, $V_B/V_G = 16\%$). The mean is 0, and the additive-by-additive epistatic effect is 2 ($\mu = 0$, $i_{aa} = 2$; $V_i/V_G = 16\%$). The heritability is 0.2 ($h^2 = 0.2$, $\sigma^2 = 25$). The sample size is 200 and the simulated replicates are 500.

^a QTL mapping with epistasis ignored.

^b QTL mapping with epistasis taken into account.

TABLE 11

The three orthogonal contrast scales (W 's) for the backcross population

Genotype	$AABB$	$AABb$	$AaBB$	$AaBb$
G	G_{22}	G_{21}	G_{12}	G_{11}
P	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$
W_1	$\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$
W_2	$\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{2}$
W_3	$\frac{1}{4}$	$-\frac{1}{4}$	$-\frac{1}{4}$	$\frac{1}{4}$

G 's and P 's denote the genotypic values and expected genotypic frequencies for the four genotypes of two unlinked genes, A and B.

erham's model (Equation 42). The estimates of environmental variance and heritability are 24.26 (SD 3.29) and 0.2158 (SD 0.0445), respectively. If epistasis is considered, the powers to detect Q_A and Q_B are 1.0 and 0.5, respectively. The power of detecting Q_B improves from 0.238 without epistasis to 0.5 with epistasis. The mean estimates of positions of Q_A and Q_B are 50.99 (SD 7.95) and 90.46 (SD 18.67), respectively. The estimated additive and dominance effects of Q_A have means 2.9658 (SD 0.5700) and 1.0024 (SD 0.8697), respectively. The mean of the estimated additive effect for Q_B is 1.3314 (SD 0.6368) from significant replicates, and it is 1.0447 (SD 0.6941) from all replicates. The mean of the estimated epistatic effects is 1.9897 (SD 0.948). The estimates of environmental variance and heritability are 24.02 (SD 3.50) and 0.225 (SD 0.0494).

CONCLUSION AND DISCUSSION

We use the orthogonal contrast scales proposed by COCKERHAM (1954) to define gene effects and to construct a genetic model, called Cockerham's model, for the study of epistasis between genes. The properties of Cockerham's model in modeling and mapping epistatic genes are investigated, and its variance component structure is also derived when genes are in linkage equilibrium and disequilibrium. The differences between Cockerham's model and other models in analyzing epistasis and mapping epistatic QTL are also compared. There are several advantages of using Cockerham's model in modeling epistasis of genes because of its orthogonal property. The advantages can benefit the study of QTL mapping. The issues of QTL mapping when epistasis is involved are also discussed. Real and simulated examples are used to illustrate Cockerham's model, verify its statistical properties, and map for epistatic QTL.

Parameterization of epistasis: Different types and degrees of epistasis can be also quantified by Cockerham's model. For example, if $a_1 = a_2 = d_1 = d_2 = 3i_{aa}/2 = 3i_{ad}/2 = 3i_{dd}/2$, the genes show classical complementary interaction with a 9:7 ratio among different genotypic values, and the marginal effects of A and B

TABLE 12

Cockerham's epistasis model for the backcross population

	AA	Aa	Mean
	G_{22}	G_{12}	\bar{G}_2
BB	$\frac{a_1}{2} + \frac{a_2}{2} + \frac{w_{12}}{4}$	$-\frac{a_1}{2} + \frac{a_2}{2} - \frac{w_{12}}{4}$	$\frac{a_2}{2}$
	G_{21}	G_{11}	\bar{G}_1
Bb	$\frac{a_1}{2} - \frac{a_2}{2} - \frac{w_{12}}{4}$	$-\frac{a_1}{2} - \frac{a_2}{2} + \frac{w_{12}}{4}$	$-\frac{a_2}{2}$
Mean	\bar{G}_2	\bar{G}_1	
	$\frac{a_1}{2}$	$-\frac{a_1}{2}$	

The marginal means \bar{G}_i and \bar{G}_j are calculated for genes in linkage equilibrium.

contribute 42.86% each and the epistatic effects contribute 14.28% to the total genetic variance. If $a_1 = a_2 = d_1 = d_2 = -i_{aa}/2 = -i_{ad}/2 = -i_{dd}/2$, the genotypes show classical duplicate interaction with a 15:1 ratio. The contributions of marginal effects of A and B and epistatic effects to the total genetic variance are 20, 20, and 60%, respectively. Other classical epistasis, such as recessive, dominant, and suppression, can also be quantified by Cockerham's model. The parameterization of epistasis can facilitate the study of epistasis in quantitative trait analyses.

Backcross populations: Cockerham's model for a backcross population can be also obtained on the basis of the same orthogonal contrast principle. When two loci A and B are considered, the F_1 progeny, which produce four different digenic gametes, AB , Ab , aB , and ab , with expected frequencies $(1 - r)/2$, $r/2$, $r/2$, and $(1 - r)/2$, respectively, are backcrossed to one of the parents. The allele frequencies p_A , p_a , p_B , and p_b in the gametes of the F_1 progeny are one-half. The scales for partitioning the genetic variance of two unlinked genes in the backcross population are listed in Table 11, and Cockerham's model for the backcross population is shown in Table 12 or can be expressed succinctly as

$$G_{ij} = \mu + a_1x_1 + a_2x_2 + w_{12}(x_1x_2), \quad (48)$$

where i (j) is 2 or 1 to denote the genotype AA (BB) or Aa (Bb) or QTL A (B), and x_1 (x_2) is the orthogonal contrast scale $\frac{1}{2}$ or $-\frac{1}{2}$ if genotype of A (B) is AA (BB) or Aa (Bb). The unique solutions of μ , a_1 , a_2 , and w_{12} are

$$\begin{aligned} \mu &= \frac{G_{22} + G_{21} + G_{12} + G_{11}}{4}, \\ a_1 &= \frac{G_{22} + G_{21} - G_{12} - G_{11}}{2}, \\ a_2 &= \frac{G_{22} - G_{21} + G_{12} - G_{11}}{2}, \end{aligned}$$

$$\begin{aligned} w_{12} &= (G_{22} - G_{21}) - (G_{12} - G_{11}) \\ &= (G_{22} - G_{12}) - (G_{21} - G_{11}). \end{aligned}$$

When genes are unlinked, the parameters $\mu = \bar{G}_.$, $a_1 = \bar{G}_2 - \bar{G}_1$, $a_1 = \bar{G}_2 - \bar{G}_1$ are the mean and marginal effects, and w_{12} measures the difference in genotypic value between the two genotypes, AA (BB) and Aa (Bb), of the A (B) gene in the background of two different genotypes, BB (AA) and Bb (Aa), of the B (A) gene. Therefore, the parameters μ , a_1 , a_2 , and w_{12} are defined as the genetic parameters of the mean, the marginal effects of genes A and B, and their epistatic effect. The advantages and properties of Cockerham's model shown in the F_2 population are also true in the backcross population. Likewise, the statistical parameters can also be derived to contrast with the genetic parameters for interpreting gene effects in linkage disequilibrium.

Cockerham's model for multiple loci: It is straightforward to extend Cockerham's model to consider multiple loci without further setting the orthogonal contrast scales for partitioning the genetic variance. The extension is based on a regression principle. In regression analysis, interaction among independent variables can generally be described in terms of a product term. Following the same principle, the model considering m loci for a backcross population can be written as

$$G = \mu + \sum_{j=1}^m a_j x_j + \sum_{j < k}^m w_{jk} (x_j x_k) + \sum_{j < k < l}^m w_{jkl} (x_j x_k x_l) + \dots,$$

where w is the genetic parameter of epistatic effect between genes. The coded variables for the epistasis are just the product of x_j 's. For m loci, there are $2^m - 1$ components consisting of m components for marginal effects and $2^m - m - 1$ epistatic components. Of the $2^m - m - 1$ epistatic components there are $m(m-1)/2$ second-order epistasis components, $m(m-1)(m-2)/3$ third-order components, etc. Generally, higher order epistasis (order higher than two) contributes very little to the total genetic variation and can be ignored. The extension of Cockerham's model to a multiple-locus model in the F_2 population is also straightforward, but trivial. Statistical parameters for the multiple-locus model can also be derived.

Cockerham's model distinguishes itself from other models by the property of orthogonal contrast and will show the advantages in partitioning variance, genetic interpretation, statistical estimation, and QTL mapping over the other models as described in this article. When Cockerham's multiple-locus model is used to form the base of a multiple interval mapping (MIM) model (KAO *et al.* 1999) for QTL mapping, the likelihood of the MIM model is a 3^m (2^m) normal mixture for the F_2 (backcross) population, and the general formulas by KAO and ZENG (1997) can be applied to obtain the MLE and the asymptotic variance-covariance matrix by using the coded variables to set up the genetic design matrix D and the condi-

tional independence property to construct the conditional probability matrix Q . The orthogonalized MIM model can facilitate the search of epistatic QTL, enhance the resolution of QTL mapping, and help outline a better QTL mapping strategy.

This article is dedicated to the memory of Dr. C. Clark Cockerham. The authors are grateful to Dr. Christopher Basten for his suggestions and two anonymous reviewers for helpful comments. C.-H.K. was supported by grants NSC90-2313-B-001-019 from the National Science Council, Taiwan, Republic of China. Z.-B.Z. was funded by GM45344 from the National Institutes of Health and no. 9801300 from the U.S. Department of Agriculture, Plant Genome Program.

LITERATURE CITED

- BRIM, C. A., and C. C. COCKERHAM, 1961 Inheritance of quantitative characters in soybeans. *Crop Sci.* **1**: 187-190.
- CHEVERUD, J. M., and E. J. ROUTMAN, 1995 Epistasis and its contribution to genetic variance components. *Genetics* **139**: 1455-1461.
- COCKERHAM, C. C., 1954 An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present. *Genetics* **39**: 859-882.
- COCKERHAM, C. C., and Z.-B. ZENG, 1996 Design III with marker loci. *Genetics* **143**: 1437-1456.
- CROW, J. F., and M. KIMURA, 1970 *An Introduction to Population Genetic Theory*. Burgess Publishing Company, Minneapolis, MN.
- DEMPSTER, A. P., N. M. LARID and D. B. RUBIN, 1977 Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc.* **39**: 1-38.
- DOEBLEY, J., A. STEC and C. GUSTUS, 1995 *teosinte branched1* and origin of maize: evidence for epistasis and the evolution of dominance. *Genetics* **141**: 333-346.
- FISHER, R. A., 1918 The correlation between relatives on the supposition of Mendelian inheritance. *Proc. R. Soc. Edinb.* **52**: 399-433.
- GOODNIGHT, C. J., 2000 Quantitative trait loci and gene interaction: the quantitative genetics of metapopulation. *Heredity* **84**: 587-598.
- HALDANE, J. B. S., 1919 The combination of linkage values and the calculation of distances between the loci of linked factors. *J. Genet.* **8**: 299-309.
- HALEY, C. S., and S. A. KNOTT, 1992 A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* **69**: 315-324.
- HAYMAN, B. I., and K. MATHER, 1955 The description of genetic interactions in continuous variation. *Biometrics* **11**(1): 69-82.
- JANSEN, R. C., 1993 Interval mapping of multiple quantitative trait loci. *Genetics* **135**: 205-211.
- KAO, C.-H., and Z.-B. ZENG, 1997 General formulas for obtaining the MLE and the asymptotic variance-covariance matrix in mapping quantitative trait loci when using the EM algorithm. *Biometrics* **53**: 359-371.
- KAO, C.-H., Z.-B. ZENG and R. D. TEASDALE, 1999 Multiple interval mapping for quantitative trait loci. *Genetics* **152**: 1203-1216.
- KEARSEY, M. J., and H. S. POONI, 1996 *The Genetical Analysis of Quantitative Traits*. Chapman & Hall, London/New York.
- KEMPTHORNE, O., 1957 *An Introduction to Genetic Statistics*. John Wiley & Sons, New York.
- LANDER, E. S., and D. BOTSTEIN, 1989 Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**: 185-199.
- LEE, J. A., C. C. COCKERHAM and F. H. SMITH, 1968 The inheritance of gossypol level in gossypium. I: additive, dominance, epistatic, and maternal effects associated with seed gossypol in two varieties of *Gossypium hirsutum* L. *Genetics* **59**: 285-298.
- MATHER, K., 1967 Complementary and duplicate gene interactions in biometrical genetics. *Heredity* **22**: 97-103.
- MATHER, K., and J. L. JINKS, 1982 *Biometrical Genetics*, Ed. 3. Chapman & Hall, London.
- SEN, S., and G. A. CHURCHILL, 2001 A statistical framework for quantitative trait mapping. *Genetics* **159**: 371-387.

STEELE, R. G. D., and J. H. TORRIE, 1981 *Principles and Procedures of Statistics: A Biometrical Approach*. McGraw-Hill, New York.

STUBER, C. W., and R. H. MOLL, 1971 Epistasis in maize (*Zea mays* L.) II. Comparison of selected with unselected populations. *Genetics* **67**: 137-149.

STUBER, C. W., S. E. LINCOLN, D. W. WOFF, T. HELENTJARIS and E. S. LANDER, 1992 Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular marker. *Genetics* **132**: 832-839.

VAN DER VEEN, J. H., 1959 Tests of non-allelic interaction and linkage for quantitative characters in generations derived from two diploid pure lines. *Genetica* **30**: 201-232.

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

WEIR, B. S., and C. C. COCKERHAM, 1977 Two-locus theory in quantitative genetics, pp. 247-269 in *Proceedings of the International Conference on Quantitative Genetics*, edited by E. POLLAK, O. KEMPTHORNE and T. B. BAILEY, JR. Iowa State University Press, Ames, IA.

WOLFRAM, S., 1992 *Mathematica*. Wolfram Research, Champaign, IL.

WRIGHT, S., 1980 Genic and organismic selection. *Evolution* **34**: 825-843.

ZENG, Z-B., 1994 Precision mapping of quantitative trait loci. *Genetics* **136**: 1457-1468.

ZENG, Z-B., J. LIU, L. F. STAM, C.-H. KAO, J. M. MERCER *et al.*, 2000 Genetic architecture of a morphological shape difference between two drosophila species. *Genetics* **154**: 299-310.

Communicating editor: M. A. F. NOOR

APPENDIX A

By taking expectation of each normal equation of the model in Equation 19, the nine expected normal equations are as follows using $E(n_{ij}) = nP_{ij}$ and $E(y_{ij}) = nP_{ij}G_{ij}$:

$$\begin{aligned}
 & P_{22}G_{22} + P_{21}G_{21} + P_{30}G_{30} + P_{12}G_{12} + P_{11}G_{11} + P_{10}G_{10} + P_{02}G_{02} + P_{01}G_{01} \\
 & + P_{00}G_{00} \\
 = & \mu - (-P_{22} - P_{21} - P_{20} + P_{02} + P_{01} + P_{00})a_1 \\
 & - (P_{22} + P_{21} + P_{20} - P_{12} - P_{11} - P_{10} + P_{02} + P_{01} + P_{00})\frac{d_1}{2} \\
 & - (-P_{22} + P_{20} - P_{12} + P_{10} - P_{02} + P_{00})a_2 \\
 & - (P_{22} - P_{21} + P_{20} + P_{12} - P_{11} + P_{10} + P_{02} - P_{01} + P_{00})\frac{d_2}{2} \\
 & - (P_{22} - P_{20} - P_{02} + P_{00})i_{aa} \\
 & - (P_{22} - P_{21} + P_{20} - P_{02} + P_{01} - P_{00})\frac{i_{ad}}{2} \\
 & - (P_{22} - P_{20} - P_{12} + P_{10} + P_{02} - P_{00})\frac{i_{da}}{2} \\
 & - (-P_{22} + P_{21} - P_{20} + P_{12} - P_{11} + P_{10} - P_{02} + P_{01} - P_{00})\frac{i_{dd}}{4},
 \end{aligned} \tag{A1}$$

$$\begin{aligned}
 & \frac{(P_{22}G_{22} + P_{21}G_{21} + P_{30}G_{30}) - (P_{02}G_{02} + P_{01}G_{01} + P_{00}G_{00})}{(P_{22} + P_{21} + P_{20} + P_{02} + P_{01} + P_{00})} \\
 = & \frac{1}{(P_{22} + P_{21} + P_{20} + P_{02} + P_{01} + P_{00})} \\
 \times & \{(P_{22} + P_{21} + P_{20} - P_{02} - P_{01} - P_{00})\mu \\
 & + P_{22} + P_{21} + P_{20} + P_{02} + P_{01} + P_{00}a_1
 \end{aligned}$$

$$\begin{aligned}
 & + (-P_{22} - P_{21} - P_{20} + P_{02} + P_{01} + P_{00})\frac{d_1}{2} + (P_{22} - P_{20} - P_{02} + P_{00})a_2 \\
 & + (-P_{22} + P_{21} - P_{20} + P_{02} - P_{01} + P_{00})\frac{d_2}{2} + (P_{22} - P_{20} + P_{02} - P_{00})i_{aa} \\
 & + (-P_{22} + P_{21} - P_{20} - P_{02} + P_{01} - P_{00})\frac{i_{ad}}{2} + (-P_{22} + P_{20} + P_{02} - P_{00})\frac{i_{da}}{2} \\
 & + (P_{22} - P_{21} + P_{20} - P_{02} + P_{01} - P_{00})\frac{i_{dd}}{4},
 \end{aligned} \tag{A2}$$

$$\begin{aligned}
 & \frac{(P_{22}G_{22} + P_{12}G_{12} + P_{02}G_{02}) - (P_{30}G_{30} + P_{10}G_{10} + P_{00}G_{00})}{(P_{22} + P_{20} + P_{12} + P_{10} + P_{02} + P_{00})} \\
 = & \frac{1}{(P_{22} + P_{20} + P_{12} + P_{10} + P_{02} + P_{00})} \\
 \times & \{(P_{22} - P_{20} + P_{12} - P_{10} + P_{02} - P_{00})\mu \\
 & + (P_{22} - P_{20} - P_{02} + P_{00})a_1 \\
 & + (-P_{22} + P_{20} + P_{12} - P_{10} - P_{02} + P_{00})\frac{d_1}{2} \\
 & + (P_{22} + P_{20} + P_{12} + P_{10} + P_{02} + P_{00})a_2 \\
 & + (-P_{22} + P_{20} - P_{12} + P_{10} - P_{02} + P_{00})\frac{d_2}{2} \\
 & + (P_{22} + P_{20} - P_{02} - P_{00})i_{aa} \\
 & + (-P_{22} + P_{20} + P_{02} - P_{00})\frac{i_{ad}}{2} \\
 & + (-P_{22} - P_{20} + P_{12} + P_{10} - P_{02} - P_{00})\frac{i_{da}}{2} \\
 & + (P_{22} - P_{20} - P_{12} + P_{10} + P_{02} - P_{00})\frac{i_{dd}}{4},
 \end{aligned} \tag{A3}$$

$$\begin{aligned}
 & (P_{12}G_{12} + P_{11}G_{11} + P_{10}G_{10}) - (P_{22}G_{22} + P_{21}G_{21} + P_{30}G_{30}) \\
 & - (P_{02}G_{02} + P_{01}G_{01} + P_{00}G_{00}) \\
 = & (-P_{22} - P_{21} - P_{20} + P_{12} + P_{11} + P_{10} - P_{02} - P_{01} - P_{00})\mu \\
 & + (-P_{22} - P_{21} - P_{20} + P_{02} + P_{01} + P_{00})a_1 + \frac{d_1}{2} \\
 & + (-P_{22} + P_{20} + P_{12} - P_{10} - P_{02} + P_{00})a_2 \\
 & + (P_{22} - P_{21} + P_{20} - P_{12} + P_{11} - P_{10} + P_{02} - P_{01} + P_{00})\frac{d_2}{2} \\
 & + (-P_{22} + P_{20} + P_{02} - P_{00})i_{aa} + (P_{22} - P_{21} + P_{20} - P_{02} + P_{01} - P_{00})\frac{i_{ad}}{2} \\
 & + (P_{22} - P_{20} + P_{12} - P_{10} + P_{02} - P_{00})\frac{i_{da}}{2} \\
 & + (-P_{22} + P_{21} - P_{20} - P_{12} + P_{11} - P_{10} - P_{02} + P_{01} - P_{00})\frac{i_{dd}}{4},
 \end{aligned} \tag{A4}$$

$$\begin{aligned}
 & (P_{21}G_{21} + P_{11}G_{11} + P_{01}G_{01}) - (P_{22}G_{22} + P_{12}G_{12} + P_{02}G_{02}) \\
 & - (P_{30}G_{30} + P_{10}G_{10} + P_{00}G_{00}) \\
 = & (-P_{22} + P_{21} - P_{20} - P_{12} + P_{11} - P_{10} - P_{02} + P_{01} - P_{00})\mu \\
 & + (-P_{22} + P_{21} - P_{20} + P_{02} - P_{01} + P_{00})a_1 \\
 & + (P_{22} - P_{21} + P_{20} - P_{12} + P_{11} - P_{10} + P_{02} - P_{01} + P_{00})\frac{d_1}{2}
 \end{aligned}$$

$$\begin{aligned}
& + (-P_{22} + P_{20} - P_{12} + P_{10} - P_{02} + P_{00})a_2 + \frac{d_2}{2} \\
& + (-P_{22} + P_{20} + P_{02} - P_{00})i_{aa} \\
& + (P_{22} + P_{21} + P_{20} - P_{02} - P_{01} - P_{00})\frac{i_{ad}}{2} \\
& + (P_{22} - P_{20} - P_{12} + P_{10} + P_{02} - P_{00})\frac{i_{da}}{2} \\
& + (-P_{22} - P_{21} - P_{20} + P_{12} + P_{11} + P_{10} - P_{02} - P_{01} - P_{00})\frac{i_{dd}}{4}, \tag{A5}
\end{aligned}$$

$$\begin{aligned}
& \frac{P_{22}G_{22} - P_{20}G_{20} + P_{02}G_{02} - P_{00}G_{00}}{P_{22} + P_{20} + P_{02} + P_{00}} = \frac{1}{(P_{22} + P_{20} - P_{02} + P_{00})} \\
& \times \{(P_{22} - P_{20} - P_{02} + P_{00})\mu + (P_{22} - P_{20} + P_{02} - P_{00})a_1 \\
& + (-P_{22} + P_{20} + P_{02} - P_{00})\frac{d_1}{2} \\
& + (P_{22} + P_{20} - P_{02} - P_{00})a_2 \\
& + (-P_{22} + P_{20} + P_{02} - P_{00})\frac{d_2}{2} \\
& + (P_{22} + P_{20} + P_{02} + P_{00})i_{aa} \\
& + (-P_{22} + P_{20} - P_{02} + P_{00})\frac{i_{ad}}{2} \\
& + (-P_{22} - P_{20} + P_{02} + P_{00})\frac{i_{da}}{2} \\
& + (P_{22} - P_{20} - P_{02} + P_{00})\frac{i_{dd}}{4}\}, \tag{A6}
\end{aligned}$$

$$\begin{aligned}
& \frac{P_{21}G_{21} - P_{22}G_{22} - P_{20}G_{20} - P_{01}G_{01} + P_{02}G_{02} + P_{00}G_{00}}{P_{22} + P_{21} + P_{20} + P_{02} + P_{01} + P_{00}} \\
& = \frac{1}{(P_{22} + P_{21} + P_{20} + P_{02} + P_{01} + P_{00})} \\
& \times \{(-P_{22} + P_{21} - P_{20} + P_{02} - P_{01} + P_{00})\mu \\
& + (-P_{22} + P_{21} - P_{20} - P_{02} + P_{01} - P_{00})a_1 \\
& + (P_{22} - P_{21} + P_{20} - P_{02} + P_{01} - P_{00})\frac{d_1}{2} \\
& + (-P_{22} + P_{20} + P_{02} + P_{00})a_2 + (P_{22} + P_{21} + P_{20} - P_{02} - P_{01} - P_{00})\frac{d_2}{2} \\
& + (-P_{22} + P_{20} + P_{02} + P_{00})i_{aa} \\
& + (P_{22} + P_{21} + P_{20} + P_{02} + P_{01} + P_{00})\frac{i_{ad}}{2} \\
& + (P_{22} - P_{20} - P_{02} + P_{00})\frac{i_{da}}{2} \\
& + (-P_{22} - P_{21} - P_{20} + P_{02} + P_{01} + P_{00})\frac{i_{dd}}{4}\}, \tag{A7}
\end{aligned}$$

$$\begin{aligned}
& \frac{P_{12}G_{12} - P_{22}G_{22} - P_{02}G_{02} - P_{10}G_{10} + P_{20}G_{20} + P_{00}G_{00}}{P_{22} + P_{20} + P_{12} + P_{10} + P_{02} + P_{00}} \\
& = \frac{1}{(P_{22} + P_{20} + P_{12} + P_{10} + P_{02} + P_{00})} \\
& \times \{(-P_{22} + P_{20} + P_{12} - P_{10} - P_{02} + P_{00})\mu \\
& + (-P_{22} + P_{20} + P_{02} - P_{00})a_1 \\
& + (P_{22} - P_{20} + P_{12} - P_{10} + P_{02} - P_{00})\frac{d_1}{2} \\
& + (-P_{22} - P_{20} + P_{12} + P_{10} - P_{02} - P_{00})a_2
\end{aligned}$$

$$\begin{aligned}
& + (P_{22} - P_{20} - P_{12} + P_{10} + P_{02} - P_{00})\frac{d_2}{2} \\
& + (-P_{22} - P_{20} + P_{02} + P_{00})i_{aa} + (P_{22} - P_{20} - P_{02} + P_{00})\frac{i_{ad}}{2} \\
& + (P_{22} + P_{20} + P_{12} + P_{10} + P_{02} + P_{00})\frac{i_{da}}{2} \\
& + (-P_{22} + P_{20} - P_{12} + P_{10} - P_{02} + P_{00})\frac{i_{dd}}{4}, \tag{A8}
\end{aligned}$$

$$\begin{aligned}
& 4(P_{22}G_{22} - P_{21}G_{21} + P_{20}G_{20} - P_{12}G_{12} + P_{11}G_{11} - P_{10}G_{10} + P_{02}G_{02} \\
& - P_{01}G_{01} + P_{00}G_{00}) \\
& = 4\{(P_{22} - P_{21} + P_{20} - P_{12} + P_{11} - P_{10} + P_{02} - P_{01} + P_{00})\mu \\
& + (P_{22} - P_{21} + P_{20} - P_{02} + P_{01} - P_{00})a_1 \\
& + (-P_{22} + P_{21} - P_{20} - P_{12} + P_{11} - P_{10} - P_{02} + P_{01} - P_{00})\frac{d_1}{2} \\
& + (P_{22} - P_{20} - P_{12} + P_{10} + P_{02} - P_{00})a_2 \\
& + (-P_{22} - P_{21} - P_{20} + P_{12} + P_{11} + P_{10} - P_{02} - P_{01} - P_{00})\frac{d_2}{2} \\
& + (P_{22} - P_{20} - P_{02} + P_{00})i_{aa} \\
& + (-P_{22} - P_{21} - P_{20} + P_{02} + P_{01} + P_{00})\frac{i_{ad}}{2} \\
& + (-P_{22} + P_{20} - P_{12} + P_{10} - P_{02} + P_{00})\frac{i_{da}}{2} \\
& + (P_{22} + P_{21} + P_{20} + P_{12} + P_{11} + P_{10} + P_{02} + P_{01} + P_{00})\frac{i_{dd}}{4}\}. \tag{A9}
\end{aligned}$$

The expected normal equations can be expressed as a function of genotypic values, genotypic frequencies, and genetic parameters.

APPENDIX B

When the genotypic frequencies (P_{ij} 's) are expressed in terms of allele frequencies (p 's) and the linkage disequilibrium coefficient (D) as shown in Table 6, and the left-hand sides of the expected normal equations in APPENDIX A are replaced with statistical parameters, they can be expressed as the following equations.

$$\begin{aligned}
\beta_0 = & \mu - \tau_A a_1 - \tau_A^2 \frac{d_1}{2} - \tau_B a_2 - \tau_B^2 \frac{d_2}{2} + (\tau_A \tau_B + 2D)i_{aa} \\
& + (\tau_A \tau_B^2 + 4\tau_B D)\frac{i_{ad}}{2} + (\tau_A^2 \tau_B + 4\tau_A D)\frac{i_{da}}{2} + (\tau_A \tau_B + 4D)\frac{i_{dd}}{4}, \tag{B1}
\end{aligned}$$

where $\tau_A = p_a - p_A$, $\tau_B = p_b - p_B$, $\pi_A = p_a p_A$, $\pi_B = p_b p_B$ (WEIR 1996),

$$\begin{aligned}
\beta_1 = & \frac{1}{1 - 2\pi_A} \{-\tau_A \mu + (1 - 2\pi_A)a_1 + \tau_A \frac{d_1}{2} + (\tau_A \tau_B + 2D)a_2 \\
& + (\tau_A \tau_B^2 + 4D\tau_B)\frac{d_2}{2} - [(1 - 2\pi_A)\tau_B + 2D\tau_A]i_{aa} \\
& - [(1 - 2\pi_A)\tau_B^2 + (4\tau_A \tau_B D + 8D^2)]i_{ad} - (\tau_A \tau_B + 2D)i_{da} \\
& - (\tau_A \tau_B^2 + 4D\tau_B)\frac{i_{dd}}{4}\}, \tag{B2}
\end{aligned}$$

APPENDIX C

The variance components contributed by two genes each with two alleles in a population are shown (see APPENDIX B for the definitions of π_A , π_B , τ_A , τ_B):

$$\begin{aligned} \beta_3 = & \frac{1}{1-2\pi_B} \{ -\tau_B \mu + (\tau_A \tau_B + 2D) a_1 + (\tau_B \tau_A^2 + 4D\tau_A) \frac{d_1}{2} + (1-2\pi_B) a_2 \\ & + \tau_B \frac{d_2}{2} - (\tau_A(1-2\tau_B) + 2D\tau_B) i_{aa} - (\tau_A \tau_B + 2D) \frac{i_{ad}}{2} \\ & - [(1-2\pi_B)\tau_A^2 + 4D\tau_A \tau_B + 8D^2] \frac{i_{dd}}{2} - (\tau_B \tau_A^2 + 4D\tau_A) \frac{i_{d1}}{4} \}, \end{aligned} \quad (B3)$$

$$\begin{aligned} \frac{\beta_2}{2} = & -\tau_A^2 \mu + \tau_A a_1 + \frac{d_1}{2} + (\tau_A^2 \tau_B + 4\tau_A D) a_2 + (\tau_A \tau_B + 4D) \frac{d_2}{2} \\ & - (\tau_A \tau_B + 2D) i_{aa} - (\tau_A \tau_B^2 + 4D\tau_B) \frac{i_{ad}}{2} - \tau_B \frac{i_{dd}}{2} - \tau_B \frac{i_{d1}}{4}, \end{aligned} \quad (B4)$$

$$\begin{aligned} \frac{\beta_1}{2} = & -\tau_B \mu + (\tau_B^2 \tau_A + 4\tau_B D) a_1 + (\tau_A \tau_B + 4D) \frac{d_1}{2} + \tau_B a_2 + \frac{d_2}{2} \\ & - (\tau_A \tau_B + 2D) i_{aa} - \tau_A \frac{i_{ad}}{2} - (\tau_B \tau_A^2 + 4D\tau_A) \frac{i_{dd}}{2} - \tau_A \frac{i_{d1}}{4}, \end{aligned} \quad (B5)$$

$$\begin{aligned} \beta_5 = & \frac{1}{(1-2\pi_A)(1-2\pi_B) + 2\tau_A \tau_B D + 4D^2} \\ & \times \{ (\tau_A \tau_B + 2D) \mu - [(1-2\pi_A)\tau_B + 2\tau_A D] a_1 \\ & - (\tau_A \tau_B + 2D) \frac{d_1}{2} - [\tau_A(1-2\pi_B) + 2D\tau_B] a_2 \\ & - (\tau_A \tau_B + 2D) \frac{d_2}{2} + [(1-2\pi_A)(1-2\pi_B) + 2\tau_A \tau_B D + 4D^2] i_{aa} \\ & + [(1-2\pi_A)\tau_B + 2\tau_A D] \frac{i_{ad}}{2} + [(1-2\pi_B)\tau_A + 2\tau_B D] \frac{i_{dd}}{2} \\ & + (\tau_A \tau_B + 2D) \frac{i_{d1}}{4} \}, \end{aligned} \quad (B6)$$

$$\begin{aligned} \frac{\beta_4}{2} = & \frac{1}{1-2\pi_A} \{ (\tau_A \tau_B^2 + 4D\tau_B) \mu - [(1-2\pi_A)\tau_B^2 + 4\tau_A \tau_B D + 8D^2] a_1 - (\tau_A \tau_B^2 + 4D\tau_B) \frac{d_1}{2} \\ & - (\tau_A \tau_B + 2D) a_2 - \tau_A \frac{d_2}{2} + [(1-2\pi_A)\tau_B + 2\tau_A D] i_{aa} \\ & + (1-2\pi_A) \frac{i_{ad}}{2} + (\tau_A \tau_B + 2D) \frac{i_{dd}}{2} + \tau_A \frac{i_{d1}}{4} \}, \end{aligned} \quad (B7)$$

$$\begin{aligned} \frac{\beta_2}{2} = & \frac{1}{1-2\pi_B} \{ (\tau_B \tau_A^2 + 4D\tau_A) \mu - (\tau_A \tau_B + 2D) a_1 \\ & - \tau_B \frac{d_1}{2} - [(1-2\pi_A)\tau_B^2 + 4\tau_A \tau_B D + 8D^2] a_2 \\ & - (\tau_B \tau_A^2 + 4D\tau_A) \frac{d_2}{2} + [(1-2\pi_B)\tau_A + 2\tau_B D] i_{aa} \\ & + (\tau_A \tau_B + 2D) \frac{i_{ad}}{2} + (1-2\pi_B) \frac{i_{dd}}{2} + \tau_B \frac{i_{d1}}{4} \}, \end{aligned} \quad (B8)$$

$$\begin{aligned} \frac{\beta_8}{4} = & \{ (\tau_A \tau_B^2 + 4D) \mu - (\tau_A \tau_B^2 + 2D\tau_B) a_1 - \tau_A \frac{d_1}{2} - [\tau_A^2 \tau_B + 4\tau_A D] a_2 - \tau_A \frac{d_2}{2} \\ & + (\tau_A \tau_B + 2D) i_{aa} + \tau_A \frac{i_{ad}}{2} + \tau_B \frac{i_{dd}}{2} + \frac{i_{d1}}{4} \}, \end{aligned} \quad (B9)$$

$$\begin{aligned} V_G = & \sum_{ij=1,2} P_{ij} (G_{ij} - \mu)^2 \\ = & 2\pi_A a_1^2 + \pi_A(1 + \tau_A^2) d_1^2 + 2\pi_B a_2^2 + \pi_B(1 + \tau_B^2) d_2^2 \\ & + 2(\pi_A \tau_B^2 + \pi_B \tau_A^2 + 2\pi_A \pi_B - \tau_A \tau_B D) i_{aa}^2 \\ & + \frac{1}{4} [(1-2\pi_A) - \tau_B^2 (\tau_A \tau_B + 4D)]^2 i_{ad}^2 \\ & + \frac{1}{4} [(1-2\pi_B) - \tau_A^2 (\tau_A \tau_B + 4D)]^2 i_{dd}^2 \\ & + \frac{1}{16} [1 - (\tau_A \tau_B + 4D)^4] i_{d1}^2 \end{aligned}$$

$$\begin{aligned} & + 4\pi_A \tau_A a_1 d_1 + 4D a_1 a_2 + 4\tau_B D a_1 d_2 - 4\tau_B \pi_A a_1 i_{aa} \\ & - 2(\pi_A \tau_B^2 + 4D^2) a_1 i_{ad} - 2[2\tau_B \pi_A \tau_A + (1-2\tau_A^2) D] a_1 i_{da} \\ & - 2[\tau_A \pi_A \tau_B^2 - (\tau_A^2 - 4\pi_A) \tau_B D - 4\tau_A D^2] a_1 i_{dd} \\ & + 4\tau_A D d_1 a_2 - 4D(\tau_A \tau_B - 2D) d_1 d_2 - 4\pi_A (\tau_A \tau_B + 2D) d_1 i_{aa} \\ & - 2\tau_B \pi_A (\tau_A \tau_B + 4D) d_1 i_{ad} - 2[\tau_B \pi_A (\tau_A^2 + 1) + \tau_A^3 D] d_1 i_{da} \\ & - [\pi_A \tau_B^2 (1 + \tau_A^2) - 2\tau_A^3 \tau_B D - 4\tau_A^2 D^2] d_1 i_{dd} \\ & + 4\tau_B \pi_B a_2 d_2 - 4\tau_A \pi_B a_2 i_{aa} \\ & - 2[2\tau_A \pi_B \tau_B + (1-2\tau_B^2) D] a_2 i_{ad} - 2(\pi_B \tau_A^2 + 4D^2) a_2 i_{da} \\ & - 2[\tau_B \pi_B \tau_A^2 - (\tau_B^2 - 4\pi_B) \tau_A D - 4\tau_B D^2] a_2 i_{dd} \\ & - 4\pi_B (\tau_A \tau_B + 2D) d_2 i_{aa} \\ & - 2[\tau_A \pi_B (\tau_B^2 + 1) + \tau_B^3 D] d_2 i_{ad} - 2\tau_A \pi_B (\tau_A \tau_B + 4D) d_2 i_{da} \\ & - [\pi_B \tau_A^2 (1 + \tau_B^2) - 2\tau_B^3 \tau_A D - 4\tau_B^2 D^2] d_2 i_{dd} \\ & + 2[\tau_B (\pi_A + 2\pi_B \tau_A^2) + \tau_A (1 - 3\tau_B^2) D - 4\tau_B D^2] i_{aa} i_{ad} \\ & + 2[\tau_A (\pi_B + 2\pi_A \tau_B^2) + \tau_B (1 - 3\tau_A^2) D - 4\tau_A D^2] i_{aa} i_{da} \\ & + \frac{1}{2} (\tau_A \tau_B + 2D) [1 - (\tau_A \tau_B + 4D)^2] i_{aa} i_{dd} \\ & + \frac{1}{2} [\tau_A \tau_B + 2D - \tau_A \tau_B (\tau_A \tau_B + 4D)^2] i_{ad} i_{da} \\ & + \frac{1}{2} [\tau_A \tau_B + 2D - \tau_A \tau_B (\tau_A \tau_B + 4D)^2] i_{ad} i_{dd} \\ & + \frac{1}{4} [\tau_A - \tau_B (\tau_A \tau_B + 4D)^3] i_{ad} i_{dd} + \frac{1}{4} [\tau_B - \tau_A (\tau_A \tau_B + 4D)^3] i_{da} i_{dd}. \end{aligned} \quad (C1)$$

The total genetic variance is composed of variances and covariances of different genetic parameters. With inbreeding, the genetic variance becomes even more complicated and is provided by WEIR and COCKERHAM (1977).

