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, R.O.C. †
Program in Statistical Genetics, Department of Statistics, North Carolina State University,
Raleigh, NC 27695-8203, U.S.A.

Running head: Cocherham's model

Corresponding address: Dr. Chen-Hung Kao
Institute of Statistical Science
Academia Sinica, Taipei 11529
Taiwan, ROC
Telephone number: (02) 2783-5611 ext 415

Fax number: (02) 2783-1523
E-mail: chkao@stat.sinica.edu.tw

Key words: Epistasis, Cockerham's model,
Orthogonal contrasts, Least squares,
Normal equations, Linkage disequilibrium,
Variance components, QTL mapping

ABSTRACT

We use the orthogonal contrast scales proposed by Cockerham to construct a model for studying epistasis between genes and defining gene effects in the F_2 population. This model is called Cocherham's model. The properties of Cockerham's model in modeling the relationship between a quantitative trait and epistatic genes under linkage equilibrium and disequilibrium are investigated and discussed. Cockerham's model has several advantages in partitioning the genetic variance of a quantitative trait, interpreting and estimating gene effects because of its orthogonal property when compared to other models, and thus it can facilitate the study of QTL epistasis and be readily used in QTL mapping. A numerical example is used to illustrate Cockerham's model for the analysis of epistasis bewteen genes. The extension of Cockerham's model to model epistasis in the backcross population and between multiple loci is also presented.

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 controlling qualitative trait has been demonstrated for a long time since Greger 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 et al., 1961; Lee et al. 1968; Stuber and Moll, 1971), epistasis bewteen 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 being constrcuted 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). The problem of epistasis in QTL has been considered in some QTL mapping studies (*e.g.*, Stuber et al. 1992, Doebley et al. 1995, Cockerham and Zeng, 1996; Kao et al. 1999; Zeng et al. 2000), but not sufficiently, and many theoretical and statistical studies when epistasis involves have not been discussed. Here, we adopt Cockerham's idea (1954) about partitioning genetic variance into components to propose a model, termed Cockerham's model, and investigate its properties when applying it to the study of QTL epistasis and QTL mapping.

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 (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: (a) the pure-line-metric or F_∞ (corresponding to Mather and Jink (1982)): The parameters in pure-line-metric show orthogonality with respect to genotypic frequencies of an F_∞ population with linkage equilibrium. (b) the F_2 metric (corresponding to Cockerham's model): The parameters in F_2-metric are mutually orthogonal with respect to genotypic frequencies of an F_2 population with linkage equilibrium, and (c) the mixed-metric (corresponding to Hayman and Mathers' model). The mixed-metric can be transformed to the F_2-metric and becomes orthogonal in an F_2 population by subtraction of the mean. Although these three metrics can be translated to each other by addition or subtraction of a constant (see Table 2 of Van Der Veen's paper), they

show different meaning in interpreting gene effects and different properties in estimation which may affect the study on QTL as shown in this paper. Cheverud and Routman (1995; Routman and Cheverud, 1997) presented another model to measure gene effects.

In this paper, we start from the traditional partitioning of genetic variance into variance components for the F_2 population using Cockerham's orthogonal contrast scales, then lead up to a definition of the genetic parameters of gene effects and present Cockerham's epistasis model. The advantages and properties of Cockerham's model in the F_2 population are investigated and discussed when genes are in linkage equilibrium and disequilibrium. A numerical example is used to illustrate Cockerham's model in the analysis of epistasis between genes. Finally, we generalize Cockerham's model to model epistasis in the backcross population, between multiple loci and discuss its applications in QTL mapping.

COCKERHAM'S EPISTASIS MODEL

Cockerham (1954) defined eight othogonal contrast scales to partition the genetic variance contributed by the two genes. His idea of partitioning genetic variance by the scales leads the way to propose a genetic model for interpreting gene effects.

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 diploid organism using the orthogonal contrast scales in Table 2 of his paper. To implement Cockerham's idea for partitioning the genetic variance of two genes in linkage equilibrium in the F_2 popultion (p_A=p_a=p_B=p_b=0.5), the orthogonal contrast scales for the nine distinguishable genotypes with segragation ratio $AABB$ 1: $AABb$ 2: $AAbb$ 1: $AaBB$ 2: $AaBb$ 4: $Aabb$ 2: $aaBB$ 1: $aaBb$ 2: $aabb$ 1 are modified and listed in Table 1 (see also Cockerham and Zeng (1995)). The scales, $W_t's$, satisfy two requirements

$$\sum_{i,j} f_{ij} W_{tij} = 0 \qquad \text{and} \qquad \sum_{i,j} f_{ij} W_{tij} W_{t'ij} = 0.$$

where i (j) indexed by 2, 1 or 0 refers to the genotype AA, Aa or aa at locus A (B), and W_{tij} is the scale component of genotype ij for tth scale. The first requirement ensures that deviations around the mean are compared (the scales W_{tij} are contrasts). The second requirement ensures that the contrasts are orthogonal. In Table 1, W_1 and W_2 (W_3 and W_4) are the linear and quadratic orthogonal contrasts for gene A (gene 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 × quadratic contrast. The statistical terms linear and quadratic correspond to the genetical terms additive and dominance, respectively. Cockerham's orthogonal contrast scales serve the same purpose as the orthogonal contrasts in partitioning the sum of squares due to treatment into independent single-degree-of freedom components in experimental design. Here, Cockerham's scales are used to define the genetic parameters of gene effects, and to partition the genetic variance into independent variance components.

Definition of genetic parameters: Using the orthogonal contrast scales, Cockerham interpreted the genotypic value G_{ij} of a genotype ij with the following relation;

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

where E_t is the corresponding coefficient. In matrix notation, Equation (1) is equivalent to

$$\begin{bmatrix} G_{22} \\ G_{21} \\ G_{20} \\ G_{12} \\ G_{11} \\ G_{10} \\ G_{02} \\ G_{01} \\ G_{00} \end{bmatrix} = \begin{bmatrix} 1 & 1 & -\frac{1}{2} & 1 & -\frac{1}{2} & 1 & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{4} \\ 1 & 1 & -\frac{1}{2} & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & -\frac{1}{4} \\ 1 & 1 & -\frac{1}{2} & -1 & -\frac{1}{2} & -1 & -\frac{1}{2} & \frac{1}{2} & \frac{1}{4} \\ 1 & 0 & \frac{1}{2} & 1 & -\frac{1}{2} & 0 & 0 & \frac{1}{2} & -\frac{1}{4} \\ 1 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{4} \\ 1 & 0 & \frac{1}{2} & -1 & -\frac{1}{2} & 0 & 0 & -\frac{1}{2} & -\frac{1}{4} \\ 1 & -1 & -\frac{1}{2} & 1 & -\frac{1}{2} & -1 & \frac{1}{2} & -\frac{1}{2} & \frac{1}{4} \\ 1 & -1 & -\frac{1}{2} & 0 & \frac{1}{2} & 0 & -\frac{1}{2} & 0 & -\frac{1}{4} \\ 1 & -1 & -\frac{1}{2} & -1 & -\frac{1}{2} & 1 & \frac{1}{2} & \frac{1}{2} & \frac{1}{4} \end{bmatrix} \begin{bmatrix} E_0 \\ E_1 \\ E_2 \\ E_3 \\ E_4 \\ E_5 \\ E_6 \\ E_7 \\ E_8 \end{bmatrix} \tag{2}$$

The unique solutions of $E's$ are

$$E_0 = \frac{G_{22}}{16} + \frac{G_{21}}{8} + \frac{G_{20}}{16} + \frac{G_{12}}{8} + \frac{G_{11}}{4} + \frac{G_{10}}{8} + \frac{G_{02}}{16} + \frac{G_{01}}{8} + \frac{G_{00}}{16}, \tag{3}$$

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

$$E_2 = \frac{G_{11}}{4} + \frac{G_{12}}{8} + \frac{G_{10}}{8} - \frac{G_{22}}{16} - \frac{G_{21}}{8} - \frac{G_{20}}{16} - \frac{G_{02}}{16} - \frac{G_{01}}{8} - \frac{G_{00}}{16}, \tag{5}$$

$$E_3 = \frac{G_{22}}{8} + \frac{G_{12}}{4} + \frac{G_{02}}{8} - \frac{G_{20}}{8} - \frac{G_{10}}{4} - \frac{G_{00}}{8}, \tag{6}$$

$$E_4 = \frac{G_{11}}{4} + \frac{G_{21}}{8} + \frac{G_{01}}{8} - \frac{G_{22}}{16} - \frac{G_{12}}{8} - \frac{G_{02}}{16} - \frac{G_{20}}{16} - \frac{G_{10}}{8} - \frac{G_{00}}{16}, \tag{7}$$

$$E_5 = \frac{(G_{22} - G_{02}) - (G_{20} - G_{00})}{4} = \frac{(G_{22} - G_{20}) - (G_{02} - G_{00})}{4}, \tag{8}$$

$$E_6 = \frac{(2G_{21} - G_{22} - G_{20}) - (2G_{01} - G_{02} - G_{00})}{4}, \tag{9}$$

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

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

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

Under the segregation ratio 1:2:1:2:4:2:1:2:1 for the nine genotypes of two genes with linkage equilibrium in an F_2 population, E_0 is the mean of the genotypic values and can be denoted as $\overline{G}_{..}$. Coefficient E_1 is equivalent to $(\overline{G}_{2.} - \overline{G}_{0.})/2$, which is one-half of the difference in genotypic value between two homozygote means of AA and aa, and thus is defined as the genetic parameter a_1 of additive effects of gene A. Coefficient E_2 is equivalent to $(2\overline{G}_{1.} - \overline{G}_{2.} - \overline{G}_{0.})/2$, which represents the departure in genotypic value of the heterozygote mean Aa from the midpoint between the two homozygote means AA and aa, and thus is defined as the genetic parameter, d_1, which is the dominance effect of gene A. The same argument leads to define coefficients E_3 and E_4 as the genetic parameters, a_2 and d_2, of additive and dominance effects of gene B. If the genetic effects of one gene are not the same at different genotypes of another gene, there is an interaction between two genes in the usual sense. Coefficient E_5 quantifies the difference of additive effects of gene A (gene B), $(G_{2*} - G_{0*})/2$ $((G_{*2} - G_{*0})/2)$, at two different homozygotes of gene B (gene A), BB and bb (AA and aa). The larger the difference is, the stronger the epistasis between the additive effects is. Therefore, coefficient E_5 can be defined as the genetic parameter of additive $\times$ additive epistatic effect i_{aa}. The same argument leads to define 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 loci A and B. The definitions of these nine genetic parameters are summarized in Table 2.

Partition of the genetic variance: The genetic variance caused by two genes in linkage equilibrium can be partitioned into eight independent genetic variance components using Cocherham's orthogonal contrast scales. The variance component contributed by W_t is

$$V_{W_t} = \frac{(\sum_{i,j} f_{ij} G_{ij} W_{tij})^2}{\sum_{i,j} f_{ij} W_{tij}^2}$$

More specifically, the variance components contributed by the additive effects of genes A and B are $V_{W_1} = a_1^2/2$ and $V_{W_3} = a_2^2/2$, respectively. The variance components contributed by the dominance effects of genes A and B are $V_{W_2} = d_1^2/4$ and $V_{W_4} = d_2^2/4$, respectively. The variance component contributed by additive $\times$ additive effect is $V_{W_5} = i_{aa}^2/4$. The variance

component contributed by additive × dominance effect is $V_{W_6} = i^2_{ad}/8$. The variance component contributed by dominance × additive effect is $V_{W_7} = i^2_{da}/8$. The variance component contributed by dominance × dominance effect is $V_{W_8} = i^2_{dd}/16$. Totally, the genetic variance contributed by two genes in linkage equilibrium can be partitioned into eight independent components with no covariance

$$V_G = \sum_{i,j} p_{ij}(G_{ij} - \mu)^2 = \frac{1}{2}a_1^2 + \frac{1}{4}d_1^2 + \frac{1}{2}a_2^2 + \frac{1}{4}d_2^2 + \frac{1}{4}i_{aa}^2 + \frac{1}{8}i_{ad}^2 + \frac{1}{8}i_{da}^2 + \frac{1}{16}i_{dd}^2. \qquad (12)$$

Each variance component is contributed by its own genetic parameters, for example, the component contributed by a_1 is $a_1^2/2$, and there is no genetic covariance with other effects.

Cockerham's epistasis model: Equation (1) gives the relationship between genetic parameters and genotypic values.Each of the nine genotypic values can be expressed in terms of the nine genetic parameters as summarized in Table 3, or more succinctly expressed by the following equation

$$G_{ij} = \mu + a_1x_1 + d_1z_1 + a_2x_2 + d_2z_2 + i_{aa}w_{aa} + i_{ad}w_{ad} + i_{da}w_{da} + i_{dd}w_{dd}, \qquad (13)$$

where the coded variables

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

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

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

such that each genotype corresponds to its genotypic value in Table 3. The model in Equation (13) is termed Cockerham's model for the F_2 population.

Linkage equilibrium and disequilibrium: The coded variables for genotypes in Cockerham's model (or the othogonal contrast scales W_{tij} and $W_{t'ij}$) are orthogonal to each other when the nine genotypic ratio is 1:2:1:2:4:2:1:2:1 in a population. In the F_2 population, unlinked genes will have genotypic ratio 1:2:1:2:4:2:1:2:1, and the definition of the genetic parameters in Table 1 is appropriate for describing the gene effecs and the genetic variance is partitioned (Equation (12)). If genes are linked, the genotypic ratio will deviate from 1:2:1:2:4:2:1:2:1 and depends on strength of linkage (Table 4). Consequently, the genetic parameters defined in Table 2 are not appropriate and the genetic variance will be composed of

genetic variances and covariances. To implement Cockerham's model to linked and unlinked genes, *statistical parameters* based on the least squares principle are introduced below to contrast genetic parameters in interpreting gene effects, and the genetic variance, which is composed of variances and covariances components, is derived.

MODELING QTL EPISTASIS IN THE F_2 POPULATION

If Cockerham's model in Equation (13) is used to analyze a quantitative trait, controlled by two epistatic genes, from a sample with size n in an F_2 population, the trait value of the kth individual with genotype ij can be modelled as

$$\begin{aligned} y_{ijk} &= G_{ij} + \epsilon_{ijk} \\ &= \mu + a_1 x_1 + d_1 z_1 + a_2 x_2 + d_2 z_2 + i_{aa} w_{aa} + i_{ad} w_{ad} + i_{da} w_{da} + i_{dd} w_{dd} + \epsilon_{ijk}, \end{aligned} \quad (14)$$

where ϵ_{ijk} is an environmental error. The observed frequency and sample size of genotype ij are denoted as $\tilde{P}_{ij}$ and n_{ij} where $n_{ij} = n \times \tilde{P}_{ij}$. 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. The genotypic frequencies P_{ij}'s depend on the genetic distance between genes and are listed in Table 4. The (genotypic) ratio of P_{ij}'s reduces to 1:2:1:2:4:2:1:2:1 if genes are unlinked.

Least squares estimators of genetic parameters: The least squares estimators (LSEs) of the genetic parameters in Equation (14) have the same formulation as equations (3) to (11) except that G_{ij} is replaced with $\overline{Y}_{ij.}$ regardless of whether the genes are in linkage equilibrium or not. For example, the LSE of a_1 is

$$\hat{a}_1 = \frac{\overline{y}_{22.}}{8} + \frac{\overline{y}_{21.}}{4} + \frac{\overline{y}_{20.}}{8} - \frac{\overline{y}_{02.}}{8} - \frac{\overline{y}_{01.}}{4} - \frac{\overline{y}_{00.}}{8} \quad (15)$$

For unlinked genes with genotypic ratio 1:2:1:2:4:2:1:2:1 in the F_2 population, the expectation of $\hat{a}_1$ is

$$E(\hat{a}_1) = \frac{\overline{G}_{2.} - \overline{G}_{0.}}{2} \quad (16)$$

which corresponds to the additive effect of gene A. Likewise, the LSEs of the genetic parameters are appropriate estimators for the nine gene effects when genes are unlinked. Howeve, when genes are linked, the genotypic ratio is no longer 1:2:1:2:4:2:1:2:1, and consequently some of the genetic parameters are no longer appropriate to represent the gene effects. For example, the expected value of $\hat{a}_1$ is not a measure of the difference of two homozygote

means. To remedy this problem, the *statistical parameters* are introduced to contrast with genetic parameters for the interpretation of the gene effects.

Statistical parameters of gene effects: The LSEs of the nine genetic parameters in Equation (14) can be obtained by minimizing the error sum of squares. The derivatives of the error sum of squares with respect to every genetic parameters in turn are set equal to zero, and gives the nine normal equations. For example, the normal equation with respect to a_1 is

$$\begin{aligned} y_{2..} - y_{0..} &= (n_{22} + n_{21} + n_{20} - n_{02} - n_{01} - n_{00})\hat{\mu} + (n_{22} + n_{21} + n_{20} + n_{02} + n_{01} + n_{00})\hat{a}_1 \\ &+ (-n_{22} - n_{21} - n_{20} + n_{02} + n_{01} + n_{00})\frac{\hat{d}_1}{2} + (n_{22} - n_{20} - n_{02} + n_{00})\hat{a}_2 \\ &+ (-n_{22} + n_{21} - n_{20} + n_{02} - n_{01} + n_{00})\frac{\hat{d}_2}{2} + (n_{22} - n_{20} + n_{02} - n_{00})\hat{i}_{aa} \\ &+ (-n_{22} + n_{21} - n_{20} - n_{02} + n_{01} - n_{00})\frac{\hat{i}_{ad}}{2} + (-n_{22} + n_{20} + n_{02} - n_{00})\frac{\hat{i}_{da}}{2} \\ &+ (n_{22} - n_{21} + n_{20} - n_{02} + n_{01} - n_{00})\frac{\hat{i}_{dd}}{4}. \end{aligned} \tag{17}$$

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 in Appendix A. Further, for simplicity, the left-hand sides of expected normal equation from Equations (31) to (39) are denoted as β_0, β_1 ... β_8, which are functions of genotypic values and frequencies. In Equation (31),

$$\beta_0 = P_{22}G_{22} + P_{21}G_{21} + P_{20}G_{20} + P_{12}G_{12} + P_{11}G_{11} + P_{10}G_{10} + P_{02}G_{02} + P_{01}G_{01} + P_{00}G_{00}$$

is the mean of genotypic values. In Equation (32),

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

reduces to a_1 when genes are unlinked. When genes are linked, $P_{22} + P_{21} + P_{20}$=$P_{02} + P_{01} + P_{00}$=1/4. Therefore,

$$\beta_1 = \frac{\frac{P_{22}G_{22}+P_{21}G_{21}+P_{20}G_{20}}{P_{22}+P_{21}+P_{20}} - \frac{P_{02}G_{02}+P_{01}G_{01}+P_{00}G_{00}}{P_{02}+P_{01}+P_{00}}}{2} = \frac{\overline{G}_{2.} - \overline{G}_{0.}}{2}$$

is still half of the difference in genotypic value of two homozygote means for gene A as $P_{22} + P_{21} + P_{20}$=$P_{02} + P_{01} + P_{00}$=1/4. That is β_1 is a quantity to measure the additive

effect of gene A in the usual sense no matter whether genes are in linkage equilibrium or not. Furthre, as $2P_{2.} = 2P_{0.} = P_{1.} = 1/2$ for linked and unlinked genes in an F_2 population,

$$\begin{aligned}
\beta_2 &= 2(P_{12}G_{12} + P_{11}G_{11} + P_{10}G_{10} - P_{22}G_{22} - P_{21}G_{21} - P_{20}G_{20} - P_{02}G_{02} - P_{01}G_{01} \\
&\quad -P_{00}G_{00}) \\
&= \frac{2\frac{P_{12}G_{12}+P_{11}G_{11}+P_{10}G_{10}}{P_{12}+P_{11}+P_{10}} - \frac{P_{22}G_{22}+P_{21}G_{21}+P_{20}G_{20}}{P_{22}+P_{21}+P_{20}} - \frac{P_{02}G_{02}+P_{01}G_{01}+P_{00}G_{00}}{P_{02}+P_{01}+P_{00}}}{2} \\
&= \frac{2\overline{G}_{1.} - \overline{G}_{2.} - \overline{G}_{0.}}{2} \qquad (18)
\end{aligned}$$

which can measure the dominance effect of gene A despite linkage disequilibrium. For epistasis,

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

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

In the F_2 population, the genotypic frequency P_{ij} (ratio) varies according to different strengths of linkage between genes (Table 4). Given genotypic values G's, the genetic parameters E's depend on the (unlinked) ratio 1:2:1:2:4:2:1:2:1 and will not change as the strength of linkage change (Equation (3) to (11)). On the contrary, the quantities $\beta's$ depend on the P_{ij}'s and will change as the genotypic ratio changes. Therefore, the quantities $\beta's$ are called *statistical parameters* to contrast with the genetic parameters of gene effects E's. The genetic parameters E's can be directly estimated from the model in Equation (13), but the statistical parameters β's can not be estimated directly. Fortunately, a one-to-one relation exists between the two kinds of parameters as shown below. It allows that once the genetic parameters are estimated the statistical parameters can be obtained.

The 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). That is

$$p_{AB} = p_A p_B + D$$

The linkage disequilibrium coefficient D is equivalent to $(1-2r)/4$ where r is the recombination fraction between genes A and B. If the union of gametes is random, genotypic frequencies $P'_{ij}s$ can be expressed in terms of allele frequencies $p's$ and linkage disequilibrium coefficient D as shown in Table 4. The expected normal equations from Equations (31) to (39) can be further expressed in terms of genetic parameters ($E's$), statistical parameters ($\beta's$), population allele frequencies ($p's$) and the linkage disequilibrium coefficient D as shown in Equations (40) to (48) (Appendix B). For example, the expected normal equation with respect to a_1 (Equation (32)) can be written as

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

where $\tau_A = p_a - p_A$, $\tau_B = p_b - p_B$, $\pi_A = p_A p_a$ and $\pi_B = p_B p_b$ (Weir, 1996). In the F_2 population, the allele frequencies p_A, p_a, p_B and p_b are one-half ($\tau_A = \tau_B = 0$. $\pi_A = \pi_B = 1/4$). The nine expected normal equations in Appendix B reduce to the following.

$$\beta_0 = \mu + 2Di_{aa} + (4D)^2\frac{i_{dd}}{4} \tag{19}$$

$$\beta_1 = a_1 + 4Da_2 - 16D^2\frac{i_{ad}}{2} - 4D\frac{i_{da}}{2} \tag{20}$$

$$\beta_3 = 4Da_1 + a_2 - 4D\frac{i_{ad}}{2} - 16D^2\frac{i_{da}}{2} \tag{21}$$

$$\frac{\beta_2}{2} = \frac{d_1}{2} + (4D)^2\frac{d_2}{2} - 2Di_{aa} \tag{22}$$

$$\frac{\beta_4}{2} = (4D)^2\frac{d_1}{2} + \frac{d_2}{2} - 2Di_{aa} \tag{23}$$

$$\beta_5 = \frac{1}{\frac{1}{4}+4D^2}[2D\mu - 2D\frac{d_1}{2} - 2D\frac{d_2}{2} + (\frac{1}{4}+4D^2)i_{aa} + 2D\frac{i_{dd}}{4}] \tag{24}$$

$$\beta_6 = 4(-8D^2a_1 - 2Da_2 + \frac{1}{2}\frac{i_{ad}}{2} + 2D\frac{i_{da}}{2}) \tag{25}$$

$$\beta_7 = 4(-2Da_1 - 8D^2a_2 + 2D\frac{i_{ad}}{2} + \frac{1}{2}\frac{i_{da}}{2}) \tag{26}$$

$$\beta_8 = 4[(4D)^2\mu + 2Di_{aa} + \frac{i_{dd}}{4}] \tag{27}$$

The statistical parameters ($\beta's$) are composed of genetic parameters ($E's$) *via* linkage disequilibrium, and vise versa. Obviously, if genes are in linkage equilibrium ($D = 0$), the two

kinds parameters are equivalent. If genes are not in linkage equilibrium ($D = 0$), the two kinds parameters are different, but one-to-one transferable by a nonsingular disequilibrium coefficient matrix T as shown below.

The expected normal equations (19) to (27) can be written in matrix notation as

$$\mathbf{B} = \mathbf{T}\ \mathbf{E}, \tag{28}$$

where

$$\mathbf{B}' = [\beta_0 \quad \beta_1 \quad \frac{\beta_2}{2} \quad \beta_3 \quad \frac{\beta_4}{2} \quad (1/4 + 4D^2)\beta_5 \quad \frac{\beta_6}{2} \quad \frac{\beta_7}{2} \quad \frac{\beta_8}{4}],$$

is a cloumn vector of statistical parameters,

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

is a cloumn vector of genetic parameters, and

$$\mathbf{T} = \begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 2D & 0 & 0 & (4D)^2 \\
0 & 1 & 0 & 4D & 0 & 0 & -(4D)^2 & -4D & 0 \\
0 & 0 & 1 & 0 & (4D)^2 & -2D & 0 & 0 & 0 \\
0 & 4D & 0 & 1 & 0 & 0 & -4D & -(4D)^2 & 0 \\
0 & 0 & (4D)^2 & 0 & 1 & -2D & 0 & 0 & 0 \\
2D & 0 & -2D & 0 & -2D & \frac{1}{4} + 4D^2 & 0 & 0 & 2D \\
0 & -(4D)^2 & 0 & -4D & 0 & 0 & 1 & 4D & 0 \\
0 & -4D & 0 & -(4D)^2 & 0 & 0 & 4D & 1 & 0 \\
(4D)^2 & 0 & 0 & 0 & 0 & 2D & 0 & 0 & 1
\end{bmatrix}$$

is a symmetric matrix consisted of linkage disequilibrium coefficient D. Matrix T is nonsingular with inverse

$$T^{-1} = \frac{1}{(1-(4D)^2)^2} \begin{bmatrix}
\frac{1}{1+(4D)^2} & 0 & \frac{-(4D)^2}{1+(4D)^2} & 0 & \frac{-(4D)^2}{1+(4D)^2} & -8D & 0 & 0 & \frac{(4D)^4}{1+(4D)^2} \\
0 & 1 & 0 & -4D & 0 & 0 & -(4D)^2 & 4D & 0 \\
\frac{-(4D)^2}{1+(4D)^2} & 0 & \frac{1}{1+(4D)^2} & 0 & \frac{(4D)^4}{1+(4D)^2} & 8D & 0 & 0 & \frac{-(4D)^2}{1+(4D)^2} \\
0 & -4D & 0 & 1 & 0 & 0 & 4D & -(4D)^2 & 0 \\
\frac{-(4D)^2}{1+(4D)^2} & 0 & \frac{(4D)^4}{1+(4D)^2} & 0 & \frac{1}{1+(4D)^2} & 8D & 0 & 0 & \frac{-(4D)^2}{1+(4D)^2} \\
-8D & 0 & 8D & 0 & 8D & 4(1+(4D)^2) & 0 & 0 & -8D \\
0 & -(4D)^2 & 0 & 4D & 0 & 0 & 1 & -4D & 0 \\
0 & 4D & 0 & -(4D)^2 & 0 & 0 & -4D & 1 & 0 \\
\frac{(4D)^4}{1+(4D)^2} & 0 & \frac{-(4D)^2}{1+(4D)^2} & 0 & \frac{-(4D)^2}{1+(4D)^2} & -8D & 0 & 0 & \frac{1}{1+(4D)^2}
\end{bmatrix}$$

(by MATHEMATICA, Wolfram Research, Inc, 1992). Since $\mathbf{T}$ is nonsingular, it means that the two kinds of parameters have a one-to-one relationship and are transferable given a linkage disequilibrium coefficient D. When genes are in linkage equilibrium ($D = 0$), $\mathbf{T}$ is

diagonal, and $\beta's$ correspond to $E's$. When genes are in linkage disequilibrium ($D \neq 0$), the directly estimable $E's$ can be used to obtained $\beta's$ by Equation (28).

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 disequlibrium coefficient in generation F_t is

$$D_t = (1-r)^{t-2}D$$

where $t \geq 2$ is the number of generations. As t gets larger, D_t approaches zero. Linkage equilibrium will be gradually attained in later generations by random mating. After random mating, D_t changes (becomes smaller), so do the genotypic frequencies ($P'_{ij}s$), and accordingly the statistical parameters ($\beta's$) change and become closer to the genetic parameters ($E's$). Therefore, the statistical parameters ($\beta's$) depends on the ($P'_{ij}s$) and will vary with generations. When D_t approaches 0, genes approach linkage equilibrium and the ratio of the nine genotypic frequencies approaches 1:2:1:2:4:2:1:2:1. Finally, the statistical parameters ($\beta's$) approach genetic parameters ($E's$). Therefore, 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

$$\begin{aligned} Var(G) &= \frac{1}{2}a_1^2 + \frac{1}{4}d_1^2 + \frac{1}{2}a_2^2 + \frac{1}{4}d_2^2 + \frac{1}{4}i_{aa}^2 + \frac{1}{8}i_{ad}^2 + \frac{1}{8}i_{da}^2 + \frac{1}{16}[1-(4D)^4]i_{dd}^2 \\ &\quad +4Da_1a_2 - 8D^2a_1i_{ad} - 2Da_1i_{da} + 8D^2d_1d_2 - 2Dd_1i_{aa} - 2Da_2i_{ad} \\ &\quad -8D^2a_2i_{da} - 2Dd_2i_{aa} + D[1-(4D)^2]i_{aa}i_{dd} + Di_{ad}i_{da} \end{aligned} \tag{29}$$

(Appendix C). The genetic variance is composed of the variance and covariance genetic parameters *via* the linkage disequilibrium coefficient D. If genes are in linkage equilibrium, or attain equilibrium in later generations by random mating ($D = 0$), the covariances disappear and the genetic variance is partitioned into 8 independent variance componenets.

Advantages of Cockerham's Model in the F_2 Population: When using Cockerham's model to study epistasis in the F_2 population, it has the following advantages as compared to other models. when genes are in linkage equilibrium:

(1) The genetic variance is partitioned into eight independent components (Equation (12)). There are no genetic covariances due to orthogonality. Each component is contributed to

by its corresponding genetic parameter. This is a desirable property. On the contrary, the pure-line-metric and mixed-metric models do not have this property.

(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. However, in the F_∞-metric, the marginal means of gene A are $(a_1 + d_2/2 + i_{ad}/2)$, $(d_1 + d_2/2 + i_{dd}/2)$ and $(-a_1 + d_2/2 - i_{ad}/2)$, which involve the dominance genetic parameter of locus B (d_2) and epistasis parameters i_{ad} and i_{dd}. In the mixed-metric, the marginal means of gene A are $a_1 + d_2/2$ and $d_1 + d_2/2$ and $-a_1 + d_2/2$, which involve the dominance genetic parameters of locus B (d_2). They deviate from the one locus analysis.

(3) The difference of the two homozygote means $(G_{2.} - G_{0.})/2$ $((G_{.2} - G_{.0})/2)$ estimates the genetic parameter a_1 (a_2) of gene A (B), and the departure of the heterozygote mean from the midpoint between 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 gene A (B). They follow the definition of additive and dominance effects in a one-locus analysis. But, in the F_∞ metric model, the two contrasts of gene A (B) estimate $a_1 + i_{ad}/2$ $(a_2 + i_{da}/2)$ and $d_1 + i_{dd}/2$ $(d_2 + i_{dd}/2)$, and they deviate from the definition in the one locus analysis.

(4) Genetic parameters of epistasis do not affect the estimation of the marginal effects and the location of epistatic QTL in QTL mapping. This is a desirable property in QTL mapping. This property ensures that QTL mapping could firstly be performed without taking epistasis into account without causing a problem for unlinked QTL. The other models do not possess this property.

The above advantages are attributed to the orthogonal property of Cockerham's model.

EXAMPLE

The data: The data from Doebley et al. (1995) were used as an example to illustrate the use of Cockerham's model. Doebley et al. crossed two corn inbred lines *Terosinte-M1L* $\times$ *Terosinte-M3L* to generate 183 F_2 progeny, and concluded that two unlinked markers UMC107 and BV302 are the candidate QTL for trait LBIL (average length of vegetative internodes in the primary lateral branch) in QTL analysis. The markers UMC107 and BV302 are regarded as QTL A and B in the following analysis. Among the 183 progeny, 21 individuals have trait missing and one individual has the marker missingare. The 161

individuals with complete trait and marker information are used in the analysis.

Genetic parameters: The observed genotypic means ($\overline{Y}'_{ij.}s$) and frequencies ($\overline{P}_{ij.}$'s) of the sample are shown in Table 5. Using Cockerham's genetic model, the LSEs of the genetic parameters are shown in Table 6B. The estimates of a_1 and d_1 are 15.11 (p-value 0.0008) and -3.92 (p-value 0.5035), respectively. The estimates of a_2 and d_2 are 19.46 (p-value 0.0001) and -5.66 (p-value 0.3336), respectively. It shows that the marginal effects of the QTL A and B are mostly additive. The estimate of i_{aa} is 2.68 with p-value 0.7054. Analytically, it means that the additive effects of QTL B (A) in the background AA (BB) and aa (bb) of QTL A, which are $(\overline{Y}_{22} - \overline{Y}_{20})/2 = 20.27$ $((\overline{Y}_{22} - \overline{Y}_{02})/2 = 26.93)$ and $(\overline{Y}_{02} - \overline{Y}_{00})/2 = 14.91$ $((\overline{Y}_{20} - \overline{Y}_{00})/2 = 21.57)$ respectively, differ by 2.68, and the difference is not statistically significant at the 5% level. In Figure 1A, the two near parallel lines illustrate that the difference between the additive effects of gene B in the background AA and aa is not significant. The estimate of i_{ad} is -18.28 with p-value=0.0411. Analytically, it means that the dominance effects of QTL B in the background of AA and aa, which are 21.68 ($(2\overline{Y}_{01}$ - $\overline{Y}_{02}$ - $\overline{Y}_{00})/2$) and -14.88 ($(2\overline{Y}_{21}$ - $\overline{Y}_{22}$ - $\overline{Y}_{20})/2$) respectively, are different, and the difference quantified by $i_{ad} = -18.28$ is significant at the 5% level. In Figure 1B, the cross between two lines illustrates that genotype Bb performs better than BB in the background of aa, but, it does worse in the backgound of AA. Likewise, the nonsignificant estimate of i_{da}, 3.75, which is not significant at the 5% level (p-value 0.6725), is illustrated by the three nearly parallel lines in Figure 1C. The estimate of i_{dd}, -18.13, is not significant (p-value 0.1227), although Figure 1D shows that a cross between two lines. The estimated dominace effects of QTL B are -14.88, -14.72 and 21.68 on AA, Aa and aa backgrounds, respectively, are not atatistically differenct at level 0.05.

Genetic variance and variance components: The proportion of genetic variance in the total variance (R-square of the model) is 23.66% (Table 6A). Since two QTL are unlinked, there is no genetic covariances in expectation. The variance components can be estimated by Equation (12). The additive and dominance parameters, a_1 and d_1, of QTL A contributes about 29.9% and 1.00% of the total genetic variance. The additive and dominance parameters, a_2 and d_2, of QTL B contributes about 49.5% and 2.10% of the total genetic variance. The epistaic parameters i_{aa}, i_{ad}, i_{da} and i_{dd} contribute about 0.5%, 11.00%, 0.5% and 5.4% of the total genetic variance, respectively. Totally, the epistasis contributes about

17.4% of the total genetic variance.

CONCLUSION AND DISCUSSION

We use the orthogonal contrast scales by Cockerham (1954) to define the genetic parameters of gene effects and to construct a model, called Cockerham's epistasis model, for studying digenic epistasis in the F_2 population. Primarily, Cockerham's model is derived for analyzing genes in linkage equilibrium. when genes are in linkage equilibrium, the genetic parameters of the model are approaprite in interpreting the gene effects. When genes are in linkage disequilibrium, some genetic parameters are not approaprite, but statistical parameters, which is derived based on least squares principle, can be used to interpret gene effects. The genetic and statistical parameters are equivalent when genes are in linkage equilibrium. But, they are different and one-to-one transferable when genes are not in linkage equilibrium. The variance components of genetic variance under Cockerham's model are also derived. There are several advantages of using Cockerham's model in modeling epistasis because of orthogonality. The advantages can also benefit the study of QTL mapping. A real example is used to illustrate the use of Cockerham's model in quantitative trait analysis.

Parameterization of epistasis: Using Cockerham's model, different types and degrees of epistasis can be quantified by the genetic parameters. For example, if $a_1 = a_2 = d_1 = d_2 = 3i_{aa}/2 = 3i_{ad}/2 = 3i_{da}/2 = 3i_{dd}/2$, the genotypes show classical complementary interaction with a 9:7 ratio among different genotypic values. The marginal effects of A and B contribute 42.86% each and the epistatic effects contribute 14.28% to the total genetic variance. If $a_1 = a_2 = d_1 = d_2 = -i_{aa}/2 = -i_{ad}/2 = -i_{da}/2 = -i_{dd}/2$, the genotypes show classical duplicate interaction with a 15:1 ratio. The contribution of marginal effects of A and B and epistatic effects are 20%, 20% and 60%. Other classical interactions, such as recessive, dominant and supression, in qualitative traits can also be quantified by the nine genetic parameters. The magnitude of epistasis can be measured by the genetic paramters as well as by the proportion of epistatic variance. The parameterization of epistasis can facilitate the study of epistasis in quantitative traits.

The backcross population: Cockerham's model for the backcross population can be also derived following the same orthogonal contrast principle. In a backcross population, the F_1 progeny produce four different digeneic gametes AB, Ab, aB and ab, with expected frequencies, $(1-r)/2$, $r/2$, $r/2$ and $(1-r)/2$. The allele frequencies p_A, p_a, p_B and p_b in the gametes are one-half. The scales for partitioning the genetic variance of two unlinked genes

can be constructed (Table 7) and used to derive Cockerham's digenic epistasis model for the backcross population (Table 8). Table 8 can be expressed succintly as

$$G_{ij} = \mu + a_1 x_1 + a_2 x_2 + w_{12}(x_1 x_2), \tag{30}$$

where i (j) can be 1 or 0 to denote the genotype AA (BB) or Aa (Bb) of QTL A (B), and x_1 (x_2) is 1/2 or -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

$$\begin{aligned}
\mu &= \frac{G_{11} + G_{10} + G_{01} + G_{00}}{4} \\
a_1 &= \frac{G_{11} + G_{10} - G_{01} - G_{00}}{2} \\
a_2 &= \frac{G_{11} - G_{10} + G_{01} - G_{00}}{2} \\
w_{12} &= (G_{11} - G_{10}) - (G_{01} - G_{00}) = (G_{11} - G_{01}) - (G_{10} - G_{00}).
\end{aligned}$$

When genes are unlinked, the parameters $\mu = \overline{G}_{..}$, $a_1 = \overline{G}_{1.} - \overline{G}_{0.}$, $a_1 = \overline{G}_{.1} - \overline{G}_{.0}$ are mean and marginal effects, and w_{12} measure the difference of genotypic values between genotypes AA (BB) and Aa (Bb) in two different backgrounds of BB (AA) and Bb (Aa). Therefore, 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 the epistatic effects. It should be noted that the a_1 and a_2 for the backcross population have different meaning than the a_1 and a_2 in Equation (13) for the F_2 population. The advantages and properties of Cockerham's model for the F_2 population are also true for the backcross population. Statistical parameters can also be derived to contrast with genetic parameters for interpretation of gene effects in linkage disequilibrium.

Epistasis between multiple loci: It is straightforward to extend Cockerham's model to 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 principle, the model considering m loci for a beckcross population can be written as

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

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 multiple loci model in the F_2 population is also straightforward, but trivial.

Mapping epistatic QTL: Cockerham's model can be readily used to map for epistatic QTL. Given m QTL, Q_1, Q_2, $\cdots$, and Q_m, located in m different marker intervals, I_1, I_2, $\cdots$, and I_m, along the genome for sample size n in a backcross population , the following model

$$y_i = \mu + \sum_{j=1}^{m} a_j x_{ij}^* + \sum_{j \neq k}^{m} \delta_{jk}(w_{jk} x_{ij}^* x_{ik}^*) + \epsilon_i, \qquad i = 1, 2 \cdots n,$$

can be used to analyze epistasis and search for epistatic QTL jointly or separately. By taking epistasis and multiple QTL into account in the model, QTL mapping could be more powerful, precise and accurate in the estimation of QTL effects and locations as shown by the method of multiple interval mapping (Kao et al., 1999). The advantages of Cockerham's model due to its orthogonal property can help outlining QTL mapping strategy and facilitate the search for QTL.

ACKNOWLEGEMENT

This paper is in memory of Dr. C. Clark Cockerham. The authors are grateful to Dr. Christopher Basten for his suggestions. C.-H.K. was supported by grants NSC89-2313-B-001-006 from the National Science Council, Taiwan, Republic of China. Z.-B.Z. was funded by GM45344 from the National Institutes of Health and no. 9600645 from the United States Department of Agriculture, Plant Genome Project.

LITERATURE CITED

Brim, C. A. and C. C. Cockerham. 1961. Inheritance of quantitative characters in soybeans. *Crop Science* **1**: 187-190.

Cheverud, J. M. and E. J. Routman. 1995. Epistasis and its contribution to genetic variance components. *Genetics* **139**: 1455-1461.

Routman, E.J. and J. M. Cheverud. 1997. Gene effects on a quantitative trait: Two locus epistatic effects measured at microsatellite markers and at estimated QTL. *Evolution* 51(5):1654-1662.

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 Zeng, Z.-B. 1995. Design III with marker loci. *Genetics* **143**: 1437-1456.

Doebley, J., A. Stec and C. Gustus. 1995. *teosinte branched1* and Original 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. *Preceedings of the Royal Society of Edinburgh.*

Hayman, B. I. and K. Mather. 1955. The description of genetic interactions in continuous variation. *Biometrics 11*, No 1: 69-82.

Kao, C.-H., Z.-B. Zeng and R. D. Teasdale. 1999. Multiple interval mapping for quantitative trait loci. *GENETICS* **152**:1203-1216.

Kempthorne, O. 1957. *An Introduction to Genetic Statistics.* John Wiley & Sons (New York).

Mather, K. 1941. Variation and selection of polygenic characters. *J. Genet.* **41**: 159-193.

Mather, K. 1967. Complementary and duplicate gene interactions in biometrical genetics. *Heredity* **22**: 97-103.

Mather, K. and Jinks, J, L. 1982. *Biometrical genetics.* Third edition. Chapman & Hall Landon.

Kao, C.-H., Z.-B. Zeng and R. D. Teasdale. 1999. Multiple interval mapping for quantitative

trait loci. *GENETICS* **152**:1203-1216.

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.

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., Lincoln, S. E., Woff, D. W., Helentjaris, T. and Lander, E. S. (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 XXX:* 201-232.

Weir, B. S. 1996. *Genetic Data Analysis II.* Sinauer Associates, Inc. Sunderland, Massachusetts, USA.

Weir, B. S. and C. C. Cockerham. 1977. Two-locus theory in quantitative genetics. *Proceedings of the international conference on quantitative genetics.*

Wolfram, S. 1992. *Mathematica.* Wolfram Research, Inc. Illinois, USA.

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, and C. C. Laurie. 2000. Genetic Architecture of a Morphological Shape Difference between Two Drosophila species. *Genetics* **154**: 299-310.

APPENDIX A

By taking the expectation of the nine normal equations of the statistical model in equation (14), the expected normal equations are shown in the following since the expectations of components $E(n_{ij})=nP_{ij}$ and $E(y_{ij.})=nP_{ij}G_{ij}$.

$$
\begin{aligned}
& P_{22}G_{22}+P_{21}G_{21}+P_{20}G_{20}+P_{12}G_{12}+P_{11}G_{11}+P_{10}G_{10}+P_{02}G_{02}+P_{01}G_{01} \\
& +P_{00}G_{00} \\
= \;& \mu-(-P_{22}-P_{21}-P_{20}+P_{02}+P_{01}+P_{00})a_1 \\
& -(P_{22}+P_{21}+P_{20}-P_{12}-P_{11}-P_{10}+P_{02}+P_{01}+P_{00})\frac{d_1}{2} \\
& -(-P_{22}+P_{20}-P_{12}+P_{10}-P_{02}+P_{00})a_2 \\
& -(P_{22}-P_{21}+P_{20}+P_{12}-P_{11}+P_{10}+P_{02}-P_{01}+P_{00})\frac{d_2}{2} \\
& -(P_{22}-P_{20}-P_{02}+P_{00})i_{aa}-(P_{22}-P_{21}+P_{20}-P_{02}+P_{01}-P_{00})\frac{i_{ad}}{2} \\
& -(P_{22}-P_{20}-P_{12}+P_{10}+P_{02}-P_{00})\frac{i_{da}}{2} \\
& -(-P_{22}+P_{21}-P_{20}+P_{12}-P_{11}+P_{10}-P_{02}+P_{01}-P_{00})\frac{i_{dd}}{4} \qquad (31)
\end{aligned}
$$

$$
\begin{aligned}
& \frac{(P_{22}G_{22}+P_{21}G_{21}+P_{20}G_{20})-(P_{02}G_{02}+P_{01}G_{01}+P_{00}G_{00})}{(P_{22}+P_{21}+P_{20}+P_{02}+P_{01}+P_{00})} \\
= \;& \frac{1}{(P_{22}+P_{21}+P_{20}+P_{02}+P_{01}+P_{00})}\times \\
& \{(P_{22}+P_{21}+P_{20}-P_{02}-P_{01}-P_{00})\mu+(P_{22}+P_{21}+P_{20}+P_{02}+P_{01}+P_{00})a_1 \\
& +(-P_{22}-P_{21}-P_{20}+P_{02}+P_{01}+P_{00})\frac{d_1}{2} \\
& +(P_{22}-P_{20}-P_{02}+P_{00})a_2+(-P_{22}+P_{21}-P_{20}+P_{02}-P_{01}+P_{00})\frac{d_2}{2} \\
& +(P_{22}-P_{20}+P_{02}-P_{00})i_{aa}+(-P_{22}+P_{21}-P_{20}-P_{02}+P_{01}-P_{00})\frac{i_{ad}}{2} \\
& +(-P_{22}+P_{20}+P_{02}-P_{00})\frac{i_{da}}{2}+(P_{22}-P_{21}+P_{20}-P_{02}+P_{01}-P_{00})\frac{i_{dd}}{4}\}, \qquad (32)
\end{aligned}
$$

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

$$
\begin{aligned}
&+(P_{22}-P_{20}-P_{02}+P_{00})a_1+(-P_{22}+P_{20}+P_{12}-P_{10}-P_{02}+P_{00})\frac{d_1}{2}\\
&+(P_{22}+P_{20}+P_{12}+P_{10}+P_{02}+P_{00})a_2\\
&+(-P_{22}+P_{20}-P_{12}+P_{10}-P_{02}+P_{00})\frac{d_2}{2}\\
&+(P_{22}+P_{20}-P_{02}-P_{00})i_{aa}+(-P_{22}+P_{20}+P_{02}-P_{00})\frac{i_{ad}}{2}\\
&+(-P_{22}-P_{20}+P_{12}+P_{10}-P_{02}-P_{00})\frac{i_{da}}{2}\\
&+(P_{22}-P_{20}-P_{12}+P_{10}+P_{02}-P_{00})\frac{i_{dd}}{4}\},
\end{aligned}
\tag{33}
$$

$$
\begin{aligned}
&(P_{12}G_{12}+P_{11}G_{11}+P_{10}G_{10})-(P_{22}G_{22}+P_{21}G_{21}+P_{20}G_{20})\\
&-(P_{02}G_{02}+P_{01}G_{01}+P_{00}G_{00})\\
=\;&(-P_{22}-P_{21}-P_{20}+P_{12}+P_{11}+P_{10}-P_{02}-P_{01}-P_{00})\mu\\
&+(-P_{22}-P_{21}-P_{20}+P_{02}+P_{01}+P_{00})a_1+\frac{d_1}{2}\\
&+(-P_{22}+P_{20}+P_{12}-P_{10}-P_{02}+P_{00})a_2\\
&+(P_{22}-P_{21}+P_{20}-P_{12}+P_{11}-P_{10}+P_{02}-P_{01}+P_{00})\frac{d_2}{2}\\
&+(-P_{22}+P_{20}+P_{02}-P_{00})i_{aa}+(P_{22}-P_{21}+P_{20}-P_{02}+P_{01}-P_{00})\frac{i_{ad}}{2}\\
&+(P_{22}-P_{20}+P_{12}-P_{10}+P_{02}-P_{00})\frac{i_{da}}{2}\\
&+(-P_{22}+P_{21}-P_{20}-P_{12}+P_{11}-P_{10}-P_{02}+P_{01}-P_{00})\frac{i_{dd}}{4},
\end{aligned}
\tag{34}
$$

$$
\begin{aligned}
&(P_{21}G_{21}+P_{11}G_{11}+P_{01}G_{01})-(P_{22}G_{22}+P_{12}G_{12}+P_{02}G_{02})\\
&-(P_{20}G_{20}+P_{10}G_{10}+P_{00}G_{00})\\
=\;&(-P_{22}+P_{21}-P_{20}-P_{12}+P_{11}-P_{10}-P_{02}+P_{01}-P_{00})\mu\\
&+(-P_{22}+P_{21}-P_{20}+P_{02}-P_{01}+P_{00})a_1\\
&+(P_{22}-P_{21}+P_{20}-P_{12}+P_{11}-P_{10}+P_{02}-P_{01}+P_{00})\frac{d_1}{2}\\
&+(-P_{22}+P_{20}-P_{12}+P_{10}-P_{02}+P_{00})a_2+\frac{d_2}{2}\\
&+(-P_{22}+P_{20}+P_{02}-P_{00})i_{aa}+(P_{22}+P_{21}+P_{20}-P_{02}-P_{01}-P_{00})\frac{i_{ad}}{2}\\
&+(P_{22}-P_{20}-P_{12}+P_{10}+P_{02}-P_{00})\frac{i_{da}}{2}\\
&+(-P_{22}-P_{21}-P_{20}+P_{12}+P_{11}+P_{10}-P_{02}-P_{01}-P_{00})\frac{i_{dd}}{4},
\end{aligned}
\tag{35}
$$

$$
\begin{aligned}
&\frac{P_{22}G_{22} - P_{20}G_{20} + P_{02}G_{02} - P_{00}G_{00}}{P_{22} + P_{20} + P_{02} + P_{00}} \\
= \;& \frac{1}{(P_{22} + P_{20} + P_{02} + P_{00})} \times \\
& \{(P_{22} - P_{20} - P_{02} + P_{00})\mu + (P_{22} - P_{20} + P_{02} - P_{00})a_1 \\
& +(-P_{22} + P_{20} + P_{02} - P_{00})\frac{d_1}{2} + (P_{22} + P_{20} - P_{02} - P_{00})a_2 \\
& +(-P_{22} + P_{20} + P_{02} - P_{00})\frac{d_2}{2} + (P_{22} + P_{20} + P_{02} + P_{00})i_{aa} \\
& +(-P_{22} + P_{20} - P_{02} + P_{00})\frac{i_{ad}}{2} + (-P_{22} - P_{20} + P_{02} + P_{00})\frac{i_{da}}{2} \\
& +(P_{22} - P_{20} - P_{02} + P_{00})\frac{i_{dd}}{4}\}, \qquad (36)
\end{aligned}
$$

$$
\begin{aligned}
&\frac{P_{21}G_{21} - P_{22}G_{22} - p_{20}G_{20} - p_{01}G_{01} + P_{02}G_{02} + P_{00}G_{00}}{P_{22} + P_{21} + P_{20} + P_{02} + P_{01} + P_{00}} \\
= \;& \frac{1}{(P_{22} + P_{21} + P_{20} + P_{02} + P_{01} + P_{00})} \times \\
& \{(-P_{22} + P_{21} - P_{20} + P_{02} - P_{01} + P_{00})\mu \\
& +(-P_{22} + P_{21} - P_{20} - P_{02} + P_{01} - P_{00})a_1 \\
& +(P_{22} - P_{21} + P_{20} - P_{02} + P_{01} - P_{00})\frac{d_1}{2} \\
& +(-P_{22} + P_{20} + P_{02} - P_{00})a_2 + (P_{22} + P_{21} + P_{20} - P_{02} - P_{01} - P_{00})\frac{d_2}{2} \\
& +(-P_{22} + P_{20} - P_{02} + P_{00})i_{aa} + (P_{22} + P_{21} + P_{20} + P_{02} + P_{01} + P_{00})\frac{i_{ad}}{2} \\
& +(P_{22} - P_{20} - P_{02} + P_{00})\frac{i_{da}}{2} + (-P_{22} - P_{21} - P_{20} + P_{02} + P_{01} + P_{00})\frac{i_{dd}}{4}\}, \qquad (37)
\end{aligned}
$$

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

$$
\begin{aligned}
&+(P_{22}+P_{20}+P_{12}+P_{10}+P_{02}+P_{00})\frac{i_{da}}{2}\\
&+(-P_{22}+P_{20}-P_{12}+P_{10}-P_{02}+P_{00})\frac{i_{dd}}{4}\},
\end{aligned}
\tag{38}
$$

$$
\begin{aligned}
&4(P_{22}G_{22}-P_{21}G_{21}+P_{20}G_{20}-P_{12}G_{12}+P_{11}G_{11}-P_{10}G_{10}+P_{20}G_{20}\\
&-P_{01}G_{01}+P_{00}G_{00})\\
=\;& 4\{(P_{22}-P_{21}+P_{20}-P_{12}+P_{11}-P_{10}+P_{02}-P_{01}+P_{00})\mu\\
&+(P_{22}-P_{21}+P_{20}-P_{02}+P_{01}-P_{00})a_1\\
&+(-P_{22}+P_{21}-P_{20}-P_{12}+P_{11}-P_{10}-P_{02}+P_{01}-P_{00})\frac{d_1}{2}\\
&+(P_{22}-P_{20}-P_{12}+P_{10}+P_{02}-P_{00})a_2\\
&+(-P_{22}-P_{21}-P_{20}+P_{12}+P_{11}+P_{10}-P_{02}-P_{01}-P_{00})\frac{d_2}{2}\\
&+(P_{22}-P_{20}-P_{02}+P_{00})i_{aa}+(-P_{22}-P_{21}-P_{20}+P_{02}+P_{01}+P_{00})\frac{i_{ad}}{2}\\
&+(-P_{22}+P_{20}-P_{12}+P_{10}-P_{02}+P_{00})\frac{i_{da}}{2}\\
&+(P_{22}+P_{21}+P_{20}+P_{12}+P_{11}+P_{10}+P_{02}+P_{01}+P_{00})\frac{i_{dd}}{4}\},
\end{aligned}
\tag{39}
$$

Therefore, the expected normal equations are function of genotypic values, genotypic frequencies and genetic parameters.

APPENDIX B

When the genotypic frequencies are expressed in terms of allele frequencies and a linkage disequilibrium coefficient as shown in Table 4 and the left hand sides of the expected normal equations in Appendix A are replaced with statistical parameters, the nine expected normal equations in Appendix A are equivalent to the following equations.

$$
\begin{aligned}
\beta_0 \;=\;& \mu-\tau_A a_1-\tau_A^2\frac{d_1}{2}-\tau_B a_2-\tau_B^2\frac{d_2}{2}+(\tau_A\tau_B+2D)i_{aa}+(\tau_A\tau_B^2+4\tau_B D)\frac{i_{ad}}{2}\\
&+(\tau_A^2\tau_B+4\tau_A D)\frac{i_{da}}{2}+(\tau_A\tau_B+4D)^2\frac{i_{dd}}{4},
\end{aligned}
\tag{40}
$$

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).

$$
\beta_1 \;=\; \frac{1}{1-2\pi_A}\{-\tau_A\mu+(1-2\pi_A)a_1+\tau_A\frac{d_1}{2}+(\tau_A\tau_B+2D)a_2
$$

$$+(\tau_A\tau_B^2+4D\tau_B)\frac{d_2}{2}-[(1-2\pi_A)\tau_B+2D\tau_A]i_{aa}-(\tau_A\tau_B^2+4D\tau_B)\frac{i_{dd}}{4}\},\qquad(41)$$

$$\begin{aligned}\beta_3 &= \frac{1}{1-2\pi_B}\{-\tau_B\mu+(\tau_A\tau_B+2D)a_1+(\tau_B\tau_A^2+4D\tau_A)\frac{d_1}{2}+(1-2\pi_B)a_2\\ &+\tau_B\frac{d_2}{2}-(\tau_A(1-2\tau_B)+2D\tau_B)i_{aa}-(\tau_A\tau_B+2D)\frac{i_{ad}}{2}\\ &-[(1-2\pi_B)\tau_A^2+4D\tau_A\tau_B+8D^2]\frac{i_{da}}{2}-(\tau_B\tau_A^2+4D\tau_A)\frac{i_{dd}}{4}\},\end{aligned}\qquad(42)$$

$$\begin{aligned}\frac{\beta_2}{2} &= -\tau_A^2\mu+\tau_Aa_1+\frac{d_1}{2}+(\tau_A^2\tau_B+4\tau_AD)a_2+(\tau_A\tau_B+4D)^2\frac{d_2}{2}\\ &-(\tau_A\tau_B+2D)i_{aa}-(\tau_A\tau_B^2+4D\tau_B)\frac{i_{ad}}{2}-\tau_B\frac{i_{da}}{2}-\tau_B^2\frac{i_{dd}}{4},\end{aligned}\qquad(43)$$

$$\begin{aligned}\frac{\beta_4}{2} &= -\tau_B^2\mu+(\tau_B^2\tau_A+4\tau_BD)a_1+(\tau_A\tau_B+4D)^2\frac{d_1}{2}+\tau_Ba_2+\frac{d_2}{2}\\ &-(\tau_A\tau_B+2D)i_{aa}-\tau_A\frac{i_{ad}}{2}--(\tau_B\tau_A^2+4D\tau_A)\frac{i_{da}}{2}-\tau_A^2\frac{i_{dd}}{4}\end{aligned}\qquad(44)$$

$$\begin{aligned}\beta_5 &= \frac{1}{(1-2\pi_A)(1-2\pi_B)+2\tau_A\tau_BD+4D^2}\{(\tau_A\tau_B+2D)\mu\\ &-[(1-2\pi_A)\tau_B+2\tau_AD]a_1-(\tau_A\tau_B+2D)\frac{d_1}{2}-[\tau_A(1-2\pi_B)+2D\tau_B]a_2\\ &-(\tau_A\tau_B+2D)\frac{d_2}{2}+[(1-2\pi_A)(1-2\pi_B)+2\tau_A\tau_BD+4D^2]i_{aa}\\ &+[(1-2\pi_A)\tau_B+2\tau_AD]\frac{i_{ad}}{2}+[(1-2\pi_B)\tau_A+2\tau_BD]\frac{i_{da}}{2}\\ &+(\tau_A\tau_B+2D)\frac{i_{dd}}{4}\},\end{aligned}\qquad(45)$$

$$\begin{aligned}\beta_6 &= \frac{2}{1-2\pi_A}\{(\tau_A\tau_B^2+4D\tau_B)\mu-[(1-2\pi_A)\tau_B^2+4\tau_A\tau_BD+8D^2)]a_1\\ &-(\tau_A\tau_B^2+4D\tau_B)\frac{d_1}{2}-(\tau_A\tau_B+2D)a_2-\tau_A\frac{d_2}{2}+[(1-2\pi_A)\tau_B+2\tau_AD]i_{aa}\\ &+(1-2\pi_A)\frac{i_{ad}}{2}+(\tau_A\tau_B+2D)\frac{i_{da}}{2}+\tau_A\frac{i_{dd}}{4}\},\end{aligned}\qquad(46)$$

$$\begin{aligned}\beta_7 &= \frac{2}{1-2\pi_B}\{(\tau_B\tau_A^2+4D\tau_A)\mu-(\tau_A\tau_B+2D)a_1-\tau_B\frac{d_1}{2}-[(1-2\pi_A)\tau_B^2\\ &+4\tau_A\tau_BD+8D^2)]a_2-(\tau_B\tau_A^2+4D\tau_A)\frac{d_2}{2}+[(1-2\pi_B)\tau_A+2\tau_BD]i_{aa}\end{aligned}$$

$$+(\tau_A\tau_B+2D)\frac{i_{ad}}{2}+(1-2\pi_B)\frac{i_{da}}{2}+\tau_B\frac{i_{dd}}{4}\}, \tag{47}$$

$$\begin{aligned}\beta_8 &= 4\{(\tau_A\tau_B^2+4D)^2\mu-(\tau_A\tau_B^2+2D\tau_B)a_1-\tau_B^2\frac{d_1}{2}-[\tau_A^2\tau_B+4\tau_AD)]a_2-\tau_A^2\frac{d_2}{2}\\ &+(\tau_A\tau_B+2D)\hat{i_{aa}}+\tau_A\frac{i_{ad}}{2}+\tau_B\frac{i_{da}}{2}+\frac{i_{dd}}{4}\}. \end{aligned}\tag{48}$$

APPENDIX C

If genes are not in linkage disequilibrium and do not have allele frequencies one-half, the genetic variance is no longer partitioned into eight independent components and becomes

$$\begin{aligned}V_G &= \sum P_{ij}(G_{ij}-\mu)^2\\ &= 2\pi_Aa_1^2+\pi_A(1+\tau_A^2)d_1^2+2\pi_Ba_2^2+\pi_B(1+\tau_B^2)d_2^2\\ &+2(\pi_A\tau_B^2+\pi_B\tau_A^2+2\pi_A\pi_B-\tau_A\tau_BD)i_{aa}^2+\frac{1}{4}[(1-2\pi_A)-\tau_B^2(\tau_A\tau_B+4D)^2]i_{ad}^2\\ &+\frac{1}{4}[(1-2\pi_B)-\tau_A^2(\tau_A\tau_B+4D)^2]i_{da}^2+\frac{1}{16}[1-(\tau_A\tau_B+4D)^4]i_{dd}^2+4\pi_A\tau_Aa_1d_1\\ &+4Da_1a_2+4\tau_BDa_1d_2-4\tau_B\pi_Aa_1i_{aa}-2(\pi_A\tau_B^2+4D^2)a_1i_{ad}\\ &-2[2\tau_B\pi_A\tau_A+(1-2\tau_A^2)D]a_1i_{da}-2[\tau_A\pi_A\tau_B^2-(\tau_A^2-4\pi_A)\tau_BD-4\tau_AD^2]a_1i_{dd}\\ &+4\tau_ADd_1a_2-4D(\tau_A\tau_B-2D)d_1d_2-4\pi_A(\tau_A\tau_B+2D)d_1i_{aa}\\ &-2\tau_B\pi_A(\tau_a\tau_B+4D)d_1i_{ad}-2[\tau_B\pi_A(\tau_A^2+1)+\tau_A^3D]d_1i_{da}\\ &-[\pi_A\tau_B^2(1+\tau_A^2)-2\tau_A^3\tau_BD-4\tau_A^2D^2]d_1i_{dd}+4\tau_B\pi_Ba_2d_2-4\tau_A\pi_Ba_2i_{aa}\\ &-2[2\tau_A\pi_B\tau_B+(1-2\tau_B^2)D]a_2i_{ad}-2(\pi_B\tau_A^2+4D^2)a_2i_{da}\\ &-2[\tau_B\pi_B\tau_A^2-(\tau_B^2-4\pi_B)\tau_AD-4\tau_BD^2]a_2i_{dd}-4\pi_B(\tau_A\tau_B+2D)d_2i_{aa}\\ &-2[\tau_A\pi_B(\tau_B^2+1)+\tau_B^3D]d_2i_{ad}-2\tau_A\pi_B(\tau_A\tau_B+4D)d_2i_{da}\\ &-[\pi_B\tau_A^2(1+\tau_B^2)-2\tau_B^3\tau_AD-4\tau_B^2D^2]d_2i_{dd}\\ &+2[\tau_B(\pi_A+2\pi_B\tau_A^2)+\tau_A(1-3\tau_B^2)D-4\tau_BD^2]i_{aa}i_{ad}\\ &+2[\tau_A(\pi_B+2\pi_A\tau_B^2)+\tau_B(1-3\tau_A^2)D-4\tau_AD^2]i_{aa}i_{da}\\ &+\frac{1}{2}(\tau_A\tau_B+2D)[1-(\tau_A\tau_B+4D)^2]i_{aa}i_{dd}\\ &+\frac{1}{2}[\tau_A\tau_B+2D-\tau_A\tau_B(\tau_A\tau_B+4D)^2]i_{ad}i_{da}\end{aligned}$$

$$+\frac{1}{4}[\tau_A - \tau_B(\tau_A\tau_B + 4D)^3]i_{ad}i_{dd} + \frac{1}{4}[\tau_B - \tau_A(\tau_A\tau_B + 4D)^3]i_{da}i_{dd}. \quad (49)$$

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).

Table 1

Genotypic values, frequencies, and eight orthogonal scales for F_2 populations which are utilized in partitioning the total genetic value and variance of two unlinked genes

Scale	AABB	AABb	AAbb	AaBB	AaBb	Aabb	aaBB	aaBb	aabb
G	G_{22}	G_{21}	G_{20}	G_{12}	G_{11}	G_{10}	G_{02}	G_{01}	G_{00}
f	1/16	1/8	1/16	1/8	1/4	1/8	1/16	1/8	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	$\frac{1}{4}$	$-\frac{1}{4}$	$\frac{1}{4}$	$-\frac{1}{4}$	$\frac{1}{4}$	$-\frac{1}{4}$	$\frac{1}{4}$	$-\frac{1}{4}$	$\frac{1}{4}$

TABLE 2

Definition of genetic parameters

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

E's are the solution in equation (1).

TABLE 3

Cockerham's Model (F_2-metric Model)

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

The marginal means $\overline{G}_{i.}$ and $\overline{G}_{.j}$ are calculated for genes in linkage equilibrium.

TABLE 4

Genotypic Frequencies in terms of allele frequencies and the linkage disequilibrium coefficient

	AA	Aa	aa	Total
	P_{22}	P_{12}	P_{02}	$P_{.2}$
BB	$(p_A p_B + D)^2$	$2(p_A p_B + D)(p_a p_B - D)$	$(p_a p_B - D)^2$	p_B^2
	P_{21}	P_{11}	P_{01}	$P_{.1}$
Bb	$2(p_A p_B + D)(p_A p_b - D)$	$2(p_A p_b - D)(p_a p_B - D)$ $+2(p_A p_B + D)(p_a p_b + D)$	$2(p_a p_B - D)(p_a p_b + D)$	$2p_B p_b$
	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	$P_{2.} \quad p_A^2$	$P_{1.} \quad 2p_A p_a$	$P_{0.} \quad p_a^2$	1

Assuming that G_{AB}^{Ab} and G_{Ab}^{AB}, G_{aB}^{ab} and G_{ab}^{aB}, G_{Ab}^{ab} and G_{ab}^{Ab}, G_{AB}^{aB} and G_{aB}^{AB}, and G_{aB}^{Ab}, G_{ab}^{AB}, G_{Ab}^{aB} and G_{ab}^{AB} are not distinguishable, so the number of genotypes reduces from 16 to 9. 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.

TABLE 5

Means and sample sizes of different genotypes for two unlinked QTL.

Frequencies		AA	Aa	aa	Marginal Total
BB	n	8 (9)	20 (21)	11 (10)	39
	mean	101.65	83.62	47.80	77.21
Bb	n	22 (17)	42 (45)	21 (22)	85
	mean	66.50	47.554	54.57	54.19
bb	n	3 (8)	24 (20)	10 (10)	37
	mean	61.11	40.94	17.98	36.37
Marginal	Total	33	86	42	161
		74.52	54.09	44.08	55.67

QTL A and B represents UMC107 and BV302, respectively.

Numbers in the brackets are the expected sample size.

TABLE 6A

Two Way ANOVA Analysis.

Source	Df	Sum of Square	Mean Square	F-value	p-value
A	2	16995.08	8497.54	6.89	0.0014
B	2	19227.48	9613.74	7.80	0.0006
A*B	4	10921.70	2730.42	2.21	0.0701
Error	152	187440.72	1233.16		
Total	160	245527.72			

R-square is 0.236582.

TABLE 6B

Results of estimation using Cockerham's Model.

Source	Df	Parameter Estimate	Standard Error	Test for H_0 Parameter=0	p-value
intercept	1	56.87	2.92	19.482	0.0001
a_1	1	15.11	4.47	3.41	0.0008
d_1	1	-3.92	5.84	-0.67	0.5035
a_2	1	19.46	4.42	4.40	0.0001
d_2	1	-5.66	5.84	-0.97	0.3336
i_{aa}	1	2.68	7.07	0.38	0.7054
i_{ad}	1	-18.28	8.87	-2.06	0.0411
i_{da}	1	3.75	8.85	0.42	0.6725
i_{dd}	1	-18.13	11.68	-1.55	0.1227

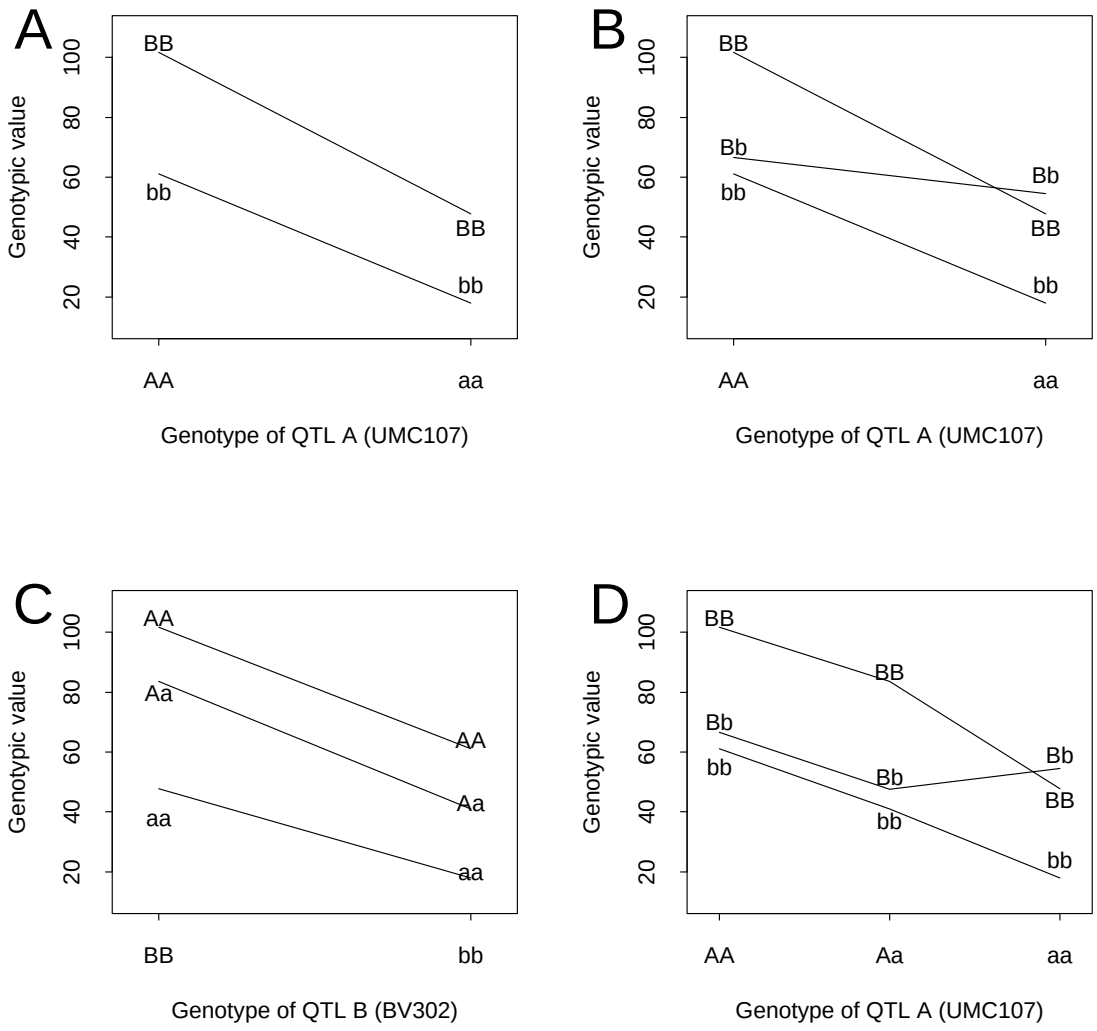
A
Genotypic value
20
40
60
80
100
BB
bb
BB
bb
AA
aa
Genotype of QTL A (UMC107)
B
Genotypic value
20
40
60
80
100
BB
Bb
bb
Bb
BB
bb
AA
aa
Genotype of QTL A (UMC107)
C
Genotypic value
20
40
60
80
100
AA
Aa
aa
AA
Aa
aa
BB
bb
Genotype of QTL B (BV302)
D
Genotypic value
20
40
60
80
100
BB
Bb
bb
BB
Bb
bb
Bb
BB
bb
AA
Aa
aa
Genotype of QTL A (UMC107)

Table 7

Genotypic values, frequencies, and orthogonal scales for backcross populations which are utilized in partitioning the genetic value and variance of two unlinked genes.

Scale	AABB	AABb	AaBB	AaBb
G	G_{11}	G_{10}	G_{01}	G_{00}
f	$1/4$	$1/4$	$1/4$	$1/4$
W_1	$1/2$	$1/2$	$-1/2$	$-1/2$
W_2	$1/2$	$-1/2$	$1/2$	$-1/2$
W_3	$1/4$	$-1/4$	$-1/4$	$1/4$

TABLE 8

Cockerham's epistasis model for the backcross population

	AA	Aa	Mean
	G_{11}	G_{01}	$\overline{G}_{.1}$
BB	$\frac{a_1}{2} + \frac{a_2}{2} + \frac{w_{12}}{4}$	$-\frac{a_1}{2} + \frac{a_2}{2} - \frac{w_{12}}{4}$	$\frac{a_2}{2}$
	G_{10}	G_{00}	$\overline{G}_{.0}$
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	$\overline{G}_{1.}$	$\overline{G}_{0.}$	
	$\frac{a_1}{2}$	$-\frac{a_1}{2}$	

The marginal means $\overline{G}_{i.}$ and $\overline{G}_{.j}$ are calculated for genes in linkage equilibrium.