

Research article

Open Access

# Models and partition of variance for quantitative trait loci with epistasis and linkage disequilibrium

Tao Wang<sup>1,3</sup> and Zhao-Bang Zeng\*<sup>1,2</sup>

Address: <sup>1</sup>Bioinformatics Research Center & Department of Statistics, North Carolina State University, Raleigh, NC 27695, USA, <sup>2</sup>Department of Genetics, North Carolina State University, Raleigh, NC 27695, USA and <sup>3</sup>Division of Biostatistics & Human Molecular Genetics Center, Medical College of Wisconsin, Milwaukee, WI 53226, USA

Email: Tao Wang - taowang@mcw.edu; Zhao-Bang Zeng\* - zeng@stat.ncsu.edu

\* Corresponding author

Published: 10 February 2006

Received: 29 June 2005

BMC Genetics 2006, 7:9 doi:10.1186/1471-2156-7-9

Accepted: 10 February 2006

This article is available from: <http://www.biomedcentral.com/1471-2156/7/9>

© 2006 Wang and Zeng; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## Abstract

**Background:** A genetic model about quantitative trait loci (QTL) provides a basis to interpret the genetic basis of quantitative traits in a study population, such as additive, dominance and epistatic effects of QTL and the partition of genetic variance. The standard quantitative genetics model is based on the least squares partition of genetic effects and also genetic variance in an equilibrium population. However, over years many specialized QTL models have also been proposed for applications in some specific populations. How are these models related? How to analyze and partition a QTL model and genetic variance when both epistasis and linkage disequilibrium are considered?

**Results:** Starting from the classical description of Cockerham genetic model, we first represent the model in a multiple regression setting by using indicator variables to describe the segregation of QTL alleles. In this setting, the definition of additive, dominance and epistatic effects of QTL and the basis for the partition of genetic variance are elaborated. We then build the connection between this general genetic model and a few specialized models (a haploid model, a diploid  $F_2$  model and a general two-allele model), and derive the genetic effects and partition of genetic variance for multiple QTL with epistasis and linkage disequilibrium for these specialized models.

**Conclusion:** In this paper, we study extensively the composition and property of the genetic model parameters, such as genetic effects and partition of genetic variance, when both epistasis and linkage disequilibrium are considered. This is the first time that both epistasis and linkage disequilibrium are considered in modeling multiple QTL. This analysis would help us to understand the structure of genetic parameters and relationship of various genetic quantities, such as allelic frequencies and linkage disequilibrium, on the definition of genetic effects, and will also help us to understand and properly interpret estimates of the genetic effects and variance components in a QTL mapping experiment.

## Background

Modeling quantitative trait loci (QTL) started with Yule [1,2] and Pearson [3] (see [4,5] for the early history of quantitative genetics). However, it was Fisher [6] who laid the firm foundation for quantitative genetics. Fisher defined gene effects (additive, dominance and epistatic effects) based on the partition of genetic variance. He par-

tioned the genetic variance into a portion due to additive effects (averaged allelic substitution effects), a portion due to dominance effects (allelic interactions), and a portion due to epistatic effects (non-allelic interactions) of genes. He then studied the correlation between relatives using the model. Cockerham [7] used the orthogonal contrasts to redefine the additive and dominance effects of QTL

and, by extending the contrasts to include epistatic effects, he partitioned the epistatic variance of two loci into those due to additive × additive, additive × dominance, dominance × additive and dominance × dominance effects of QTL. Cockerham then generalized the model to multiple loci. This was further generalized by Kempthorne [8,9] to multiple alleles. This model has been used as the basis for studying quantitative genetics ever since.

However, over years, many specialized models have also been proposed. Some are just special cases of the general genetic model and some are simplified variants tailored for particular applications or interpretations. With the propagation of numerous quantitative genetic models, there have also been some confusions in literature on the definition and interpretation of additive, dominance and epistatic effects of QTL and their relationship to the partition of genetic variance. Also, there has never been a study that considers both epistasis and linkage disequilibrium in the partition of genetic variance for multiple QTL. In this paper, we try to build the connection between the general genetic model and a few other commonly used genetic models to clarify the basis for the interpretation of different genetic models.

We start with an introduction of the genetic model as expressed in [7] in the context of variance components. Then by introducing an indicator variable for each QTL allele, we represent the model in a multiple regression setting and examine the definition and meaning of the genetic effects (additive, dominance and epistatic effects) of QTL and partition of the genetic variance in an equilibrium population and also in a disequilibrium population. Most of previous studies on modeling QTL discuss epistasis only in reference to an equilibrium population. An examination of properties of a model with both epistasis and linkage disequilibrium is important for QTL analysis in both experimental and natural populations. This is another goal of the paper and is studied in great detail here. We discuss a few reduced models used for QTL analysis, such as backcross model (essentially a haploid model) and  $F_2$  model. We also give details for a general two-allele model which may be useful for studying the genetic architecture in a natural population using single nucleotide polymorphisms (SNPs).

Previously, in [10], we compared  $F_2$  model and the general two-allele model with another commonly used genetic model, called  $F_\infty$  model. By specifying the basis of definition for each model, we compared the properties of these models in the estimation and interpretation of QTL effects including epistasis and discussed a few potential problems of using  $F_\infty$  model in a segregating population for QTL analysis. Similarly, we also compared these models with another model proposed by Cheverud [11,12].

An important result of [10] is that the genetic effects defined in reference to an equilibrium population also apply to a disequilibrium population. The partial regression coefficients, that define the genetic effects in a disequilibrium population, equal to the simple regression coefficients in a corresponding equilibrium population – the usual basis to define and interpret a genetic effect including an epistatic effect. Hardy-Weinberg and linkage disequilibria only introduce covariances between different genetic effects. With this result, in this paper our discussion on epistasis and linkage disequilibrium is focused on the partition and composition of genetic variances and covariances between different genetic effects in different populations.

## Results

### The genetic model

A general genetic model for the partition of genetic variance (particularly epistatic variance) in a random mating population was first given by Cockerham [7,13] and extended to multiple alleles by Kempthorne [8,9], following the basic genetic model formulated by Fisher [6]. The model for two loci **A** and **B** with multiple alleles was expressed as follows

$$G_{ij}^{ik} = \mu + \alpha^i + \alpha_j + \delta_j^i + \beta^k + \beta_l + \gamma_l^k + (\alpha^i \beta^k) + (\alpha^i \beta_l) + (\alpha_j \beta^k) + (\alpha_j \beta_l) + (\alpha^i \gamma_l^k) + (\alpha_j \gamma_l^k) + (\delta_j^i \beta^k) + (\delta_j^i \beta_l) + (\delta_j^i \gamma_l^k) \tag{1}$$

where the genotypic value  $G_{ij}^{ik}$  is the expected phenotype of an individual carrying alleles  $A_i, A_j, B_k,$  and  $B_l$  with phased genotype  $A_i B_k / A_j B_l$  formed by the union of a paternal gamete  $A_i B_k$  and a maternal gamete  $A_j B_l$ . The model partitions the total genotypic value into a number of genetic effects which include additive effects of each allele ( $\alpha$ 's and  $\beta$ 's), dominance effects between two alleles at each locus ( $\delta$ 's and  $\gamma$ 's), additive × additive interactions between two alleles at two loci ( $(\alpha\beta)$ 's), additive × dominance interactions involving three alleles ( $(\alpha\gamma)$ 's and  $(\delta\beta)$ 's), and dominance × dominance interaction involving all four alleles ( $(\delta\gamma)$ 's).

As an ANOVA model, it is known that not all the parameters in model (1) are estimable. A number of constraint conditions on these parameters are therefore needed. Let  $p^i, q^k$  denote allelic frequencies for alleles on paternal gametes, and  $p_j, q_l$  allelic frequencies for alleles on maternal gametes. It is usually assumed that a weighted summation of genetic effects is zero over any index for each genetic component as a deviation from the mean. Some examples are

$$\begin{aligned} \sum_i p^i \alpha^i &= 0, \quad \sum_i p^i \delta_j^i = 0, \quad \sum_j p_j \delta_j^i = 0, \\ \sum_i p^i (\alpha^i \beta^k) &= 0, \quad \sum_k p^k (\alpha^i \beta^k) = 0, \\ \sum_j p_j (\delta_j^i \beta^k) &= 0, \quad \sum_k q^k (\delta_j^i \beta^k) = 0, \dots \end{aligned} \tag{2}$$

Under the assumption of *random mating and linkage equilibrium* and allowing for different allelic frequencies in paternal and maternal gametes, the mean and genetic effects can be expressed as follows based on the least squares principle:

$$\mu = \sum_{i,j,k,l} p^i p_j q^k q_l G_{jl}^{ik} \tag{3}$$

$$\begin{aligned} \alpha^i &= G_{..}^i - G_{..}, \quad \beta^k = G_{..}^k - G_{..}, \\ \alpha_j &= G_{.j} - G_{..}, \quad \beta_l = G_{.l} - G_{..}, \\ \delta_j^i &= G_{.j}^i - G_{..}^i - G_{.j} + G_{..}, \\ \gamma_l^k &= G_{.l}^k - G_{..}^k - G_{.l} + G_{..}, \\ (\alpha^i \beta^k) &= G_{..}^{ik} - G_{..}^i - G_{..}^k + G_{..}, \\ (\alpha^i \beta_l) &= G_{.l}^i - G_{..}^i - G_{.l} + G_{..}, \\ (\alpha_j \beta^k) &= G_{.j}^k - G_{.j} - G_{..}^k + G_{..}, \\ (\alpha_j \beta_l) &= G_{.jl} - G_{.j} - G_{.l} + G_{..}, \\ (\alpha^i \gamma_l^k) &= G_{.l}^{ik} - G_{..}^{ik} - G_{.l}^i - G_{.l}^k + G_{..}^i + G_{..}^k \\ &\quad + G_{.l} - G_{..}, \\ (\alpha_j \gamma_l^k) &= G_{.jl}^k - G_{.j}^k - G_{.l}^k - G_{.jl} + G_{.j} + G_{..}^k \\ &\quad + G_{.l} - G_{..}, \\ (\delta_j^i \beta^k) &= G_{.j}^{ik} - G_{..}^{ik} - G_{.j}^i - G_{.j}^k + G_{..}^i + G_{..}^k \\ &\quad + G_{.j} - G_{..}, \\ (\delta_j^i \beta_l) &= G_{.jl}^i - G_{.j}^i - G_{.l}^i - G_{.jl} + G_{.j} + G_{..}^i \\ &\quad + G_{.l} - G_{..}, \\ (\delta_j^i \gamma_l^k) &= G_{.jl}^{ik} - G_{.j}^{ik} - G_{.l}^{ik} - G_{.jl}^i + G_{.jl}^k + G_{..}^{ik} \\ &\quad + G_{.jl} + G_{.j}^i + G_{.l}^k + G_{.j}^k + G_{.l}^i - G_{..}^i \\ &\quad - G_{..}^k - G_{.j} - G_{.l} + G_{..} \end{aligned}$$

where  $G_{..}^i = \sum_{i,j,k,l} p^i p_j q^k q_l G_{jl}^{ik}$ ,  $G_{..}^k = \sum_{j,k,l} p_j q^k q_l G_{jl}^{ik}$ , and so on. The total genetic variance is  $V_G = \sum_{i,j,k,l} p^i p_j q^k q_l (G_{jl}^{ik} - \mu)^2$ , and has an orthogonal partition under random mating and linkage equilibrium

$$V_G = V_{A_1} + V_{A_2} + V_{D_1} + V_{D_2} + V_{A_1 A_2} + V_{A_1 D_2} + V_{D_1 A_2} + V_{D_1 D_2} \tag{4}$$

with

$$\begin{aligned} V_{A_1} &= \sum_i p^i (\alpha^i)^2 + \sum_j p_j (\alpha_j)^2 \\ V_{D_1} &= \sum_{i,j} p^i p_j (\delta_j^i)^2 \\ V_{A_2} &= \sum_k q^k (\beta^k)^2 + \sum_l q_l (\beta_l)^2 \\ V_{D_2} &= \sum_{k,l} q^k q_l (\gamma_l^k)^2 \\ V_{A_1 A_2} &= \sum_{i,k} p^i q^k (\alpha^i \beta^k)^2 + \sum_{j,l} p_j q_l (\alpha_j \beta_l)^2 \\ &\quad + \sum_{i,l} p^i q_l (\alpha^i \beta_l)^2 + \sum_{j,k} p_j q^k (\alpha_j \beta^k)^2 \\ V_{A_1 D_2} &= \sum_{i,k,l} p^i q^k q_l (\alpha^i \gamma_l^k)^2 + \sum_{j,k,l} p_j q^k q_l (\alpha_j \gamma_l^k)^2 \\ V_{D_1 A_2} &= \sum_{i,j,k} p^i p_j q^k (\delta_j^i \beta^k)^2 + \sum_{i,j,l} p^i p_j q_l (\delta_j^i \beta_l)^2 \\ V_{D_1 D_2} &= \sum_{i,j,k,l} p^i p_j q^k q_l (\delta_j^i \gamma_l^k)^2 \end{aligned}$$

Using indicator variables, we can represent model (1) in another form. Assume that the two loci A and B have alleles  $A_i$ ,  $i = 1, 2, \dots, n_1$ ; and  $B_k$ ,  $k = 1, 2, \dots, n_2$ , respectively. We define the following indicator variables to represent the segregation of alleles in a population.

$$z_{M_i}^{(1)} = \begin{cases} 1, & \text{for } A_i \text{ allele from paternal gamete} \\ 0, & \text{otherwise.} \end{cases}$$

$$z_{F_j}^{(1)} = \begin{cases} 1, & \text{for } A_j \text{ allele from maternal gamete} \\ 0, & \text{otherwise.} \end{cases}$$

for  $i, j = 1, 2, \dots, n_1$  at locus A, and

$$z_{M_k}^{(2)} = \begin{cases} 1, & \text{for } B_k \text{ allele from paternal gamete} \\ 0, & \text{otherwise.} \end{cases}$$

$$z_{F_l}^{(2)} = \begin{cases} 1, & \text{for } B_l \text{ allele from maternal gamete} \\ 0, & \text{otherwise.} \end{cases}$$

for  $k, l = 1, 2, \dots, n_2$  at locus B. In terms of these indicator variables, we have the following.

- Hardy-Weinberg equilibrium (HWE) implies that  $\{z_{M_i}^{(1)}, i = 1, 2, \dots, n_1\}$  are independent of  $\{z_{F_j}^{(1)}, j = 1, 2, \dots, n_1\}$ ,

and  $\{z_{M_k}^{(2)}, k = 1, 2, \dots, n_2\}$  are independent of  $\{z_{F_l}^{(2)}, l = 1, 2, \dots, n_2\}$ .

- Linkage equilibrium (LE) implies that  $\{z_{M_i}^{(1)}, i = 1, 2, \dots, n_1\}$  are independent of  $\{z_{M_k}^{(2)}, k = 1, 2, \dots, n_2\}$ , and  $\{z_{F_j}^{(1)}, j = 1, 2, \dots, n_1\}$  are independent of  $\{z_{F_l}^{(2)}, l = 1, 2, \dots, n_2\}$ .
- There is another type of disequilibrium; i.e., the so-called genotypic disequilibrium [14] for two alleles on different gametes and at different loci. So, the genotypic equilibrium (GE) here means that  $\{z_{M_i}^{(1)}, i = 1, 2, \dots, n_1\}$  are independent of  $\{z_{F_l}^{(2)}, l = 1, 2, \dots, n_2\}$ , and  $\{z_{F_j}^{(1)}, j = 1, 2, \dots, n_1\}$  are independent of  $\{z_{M_k}^{(2)}, k = 1, 2, \dots, n_2\}$ .

It is known that under random mating we have both HWE and GE, which together are called gametic phase equilibrium. Now, let  $G$  denote the genotypic value of a progeny drawn randomly from the current population. Based on Cockerham model,  $G$  can be expressed as

$$\begin{aligned}
 G = & \mu + \sum_{i=1}^{n_1} \alpha^i z_{M_i}^{(1)} + \sum_{j=1}^{n_1} \alpha_j z_{F_j}^{(1)} + \sum_{i,j} \delta_j^i z_{M_i}^{(1)} z_{F_j}^{(1)} \\
 & + \sum_{k=1}^{n_2} \beta^k z_{M_k}^{(2)} + \sum_{l=1}^{n_2} \beta_l z_{F_l}^{(2)} + \sum_{k,l} \gamma_l^k z_{M_k}^{(2)} z_{F_l}^{(2)} \\
 & + [\sum_{i,k} (\alpha^i \beta^k) z_{M_i}^{(1)} z_{M_k}^{(2)} + \sum_{i,l} (\alpha^i \beta_l) z_{M_i}^{(1)} z_{F_l}^{(2)} \\
 & + \sum_{j,k} (\alpha_j \beta^k) z_{F_j}^{(1)} z_{M_k}^{(2)} + \sum_{j,l} (\alpha_j \beta_l) z_{F_j}^{(1)} z_{F_l}^{(2)}] \\
 & + [\sum_{i,k,l} (\alpha^i \gamma_l^k) z_{M_i}^{(1)} z_{M_k}^{(2)} z_{F_l}^{(2)} + \sum_{j,k,l} (\alpha_j \gamma_l^k) z_{F_j}^{(1)} z_{M_k}^{(2)} z_{F_l}^{(2)}] \\
 & + [\sum_{i,j,k} (\delta_j^i \beta^k) z_{M_i}^{(1)} z_{F_j}^{(1)} z_{M_k}^{(2)} + \sum_{i,j,l} (\delta_j^i \beta_l) z_{M_i}^{(1)} z_{F_j}^{(1)} z_{F_l}^{(2)}] \\
 & + \sum_{i,j,k,l} (\delta_j^i \gamma_l^k) z_{M_i}^{(1)} z_{F_j}^{(1)} z_{M_k}^{(2)} z_{F_l}^{(2)} \quad (5)
 \end{aligned}$$

This is simply a different presentation of Cockerham model with the same constraint conditions applied on the coefficient parameters. For a given individual with genotype  $A_i B_k / A_j B_l$ ,  $G$  will take the same value of  $G_{jl}^{ik}$  as before. However, this expression is helpful for us to understand some details about each component of genetic effects. We

can see this more clearly in the examination of some reduced models later.

In general, the genetic effects can be defined separately for alleles that are paternally and maternally transmitted to account for possible biological differences. As a fully parameterized model for  $G_{jl}^{ik}$ , model (3) may give  $G_{jl}^{ik} \neq G_{jk}^{il} \neq G_{il}^{jk} \neq G_{ik}^{jl}$  depending on how genetic effects are defined. If locus **A** has  $n_1$  alleles, and locus **B** has  $n_2$  alleles, there are  $N = n_1^2 n_2^2$  possible phased genotypes in total with the partition of the degrees of freedom given in Table 1.

If we assume that the union of paternal gamete  $A_i B_k$  with maternal gamete  $A_j B_l$  have the same mean effect as that of paternal gamete  $A_j B_l$  with maternal gamete  $A_i B_k$  (i.e.,  $G_{jl}^{ik} = G_{ik}^{jl}$ ), the coupling and repulsion heterozygotes have the same genotypic value (i.e.,  $G_{jl}^{ik} = G_{jk}^{il}$ ), and paternal and maternal gametes have the same gametic frequency distribution, we do not need to distinguish paternal and maternal effects. In this case, the two loci can be regarded as 2 factors and each factor has  $n_i^2$  ( $i = 1, 2$ ) levels produced by the allelic combinations of  $n_i$  alleles (cf. [7]). The total number of genotypes is  $N = n_1(n_1 + 1)n_2(n_2 + 1)/4$  and the partition of degrees of freedom is shown in Table 2. Since in this case,  $\alpha^i = \alpha_i$ ,  $\beta^k = \beta_k$ , ..., and so on, the model can also be expressed as follows

$$\begin{aligned}
 G = & \mu + \sum_{i=1}^{n_1} \alpha_i (z_{M_i}^{(1)} + z_{F_i}^{(1)}) + \sum_{i,j} \delta_j^i z_{M_i}^{(1)} z_{F_j}^{(1)} \\
 & + \sum_{k=1}^{n_2} \beta_k (z_{M_k}^{(2)} + z_{F_k}^{(2)}) + \sum_{k,l} \gamma_l^k z_{M_k}^{(2)} z_{F_l}^{(2)} \\
 & + \sum_{i,k} (\alpha_i \beta_k) (z_{M_i}^{(1)} + z_{F_i}^{(1)}) (z_{M_k}^{(2)} + z_{F_k}^{(2)}) \\
 & + \sum_{i,k,l} (\alpha_i \gamma_l^k) (z_{M_i}^{(1)} + z_{F_i}^{(1)}) z_{M_k}^{(2)} z_{F_l}^{(2)} \\
 & + \sum_{i,j,k} (\delta_j^i \beta_k) z_{M_i}^{(1)} z_{F_j}^{(1)} (z_{M_k}^{(2)} + z_{F_k}^{(2)}) \\
 & + \sum_{i,j,k,l} (\delta_j^i \gamma_l^k) z_{M_i}^{(1)} z_{F_j}^{(1)} z_{M_k}^{(2)} z_{F_l}^{(2)} \quad (6)
 \end{aligned}$$

For the case of an arbitrary number of loci, the situation will become more complicated. In addition to the addi-

**Table 1: Partition of degrees of freedom for two loci with number of alleles  $n_1$  and  $n_2$  (a general case)**

Source	Degrees of Freedom
additive ( $\alpha$ )	$2(n_1 - 1)$
dominance ( $\delta$ )	$(n_1 - 1)^2$
additive ( $\beta$ )	$2(n_2 - 1)$
dominance ( $\gamma$ )	$(n_2 - 1)^2$
additive $\times$ additive ( $\alpha\beta$ )	$4(n_1 - 1)(n_2 - 1)$
additive $\times$ dominance ( $\alpha\gamma$ )	$2(n_1 - 1)(n_2 - 1)^2$
dominance $\times$ additive ( $\delta\beta$ )	$2(n_1 - 1)^2(n_2 - 1)$
dominance $\times$ dominance ( $\delta\gamma$ )	$(n_1 - 1)^2(n_2 - 1)^2$
total	$n_1^2 n_2^2 - 1$

tive and dominance effects at each locus and two locus interactions (additive  $\times$  additive, additive  $\times$  dominance, dominance  $\times$  additive, dominance  $\times$  dominance, with a total number of  $2^2$  terms), there are 3 locus interactions (additive  $\times$  additive  $\times$  additive, additive  $\times$  additive  $\times$  dominance, ..., with a total number of  $2^3$  terms), 4 locus interactions (additive  $\times$  additive  $\times$  additive  $\times$  additive, ..., with a total number of  $2^4$  terms), and so on. Though the extension is straightforward, the total number of terms will increase dramatically. We will show some models with multiple loci in later examples by ignoring trigenic and higher order epistasis.

**Effects and variance components**

Let  $p^i, p_j$  ( $i, j = 1, 2, \dots, n_1$ ) be allelic frequencies of paternal and maternal gametes at locus A, respectively. Let also  $q^k, q_l$  ( $k, l = 1, 2, \dots, n_2$ ) denote allelic frequencies of paternal and maternal gametes at locus B, respectively. In the analysis of variance for the model, it is convenient to use deviations of the indicator variables  $z_{M_i}^{(1)}, z_{F_i}^{(1)}, z_{M_j}^{(2)}$  and  $z_{F_j}^{(2)}$  from their expected values. That is

$$x_{M_i}^{(1)} = z_{M_i}^{(1)} - E(z_{M_i}^{(1)}) = z_{M_i}^{(1)} - p^i$$

$$= \begin{cases} 1 - p^i, & \text{for } A_i \text{ allele from paternal gamete} \\ -p^i, & \text{otherwise} \end{cases}$$

Similarly, define

$$x_{F_j}^{(1)} = z_{F_j}^{(1)} - E(z_{F_j}^{(1)}) = z_{F_j}^{(1)} - p_j$$

$$x_{M_k}^{(2)} = z_{M_k}^{(2)} - E(z_{M_k}^{(2)}) = z_{M_k}^{(2)} - q^k$$

$$x_{F_l}^{(2)} = z_{F_l}^{(2)} - E(z_{F_l}^{(2)}) = z_{F_l}^{(2)} - q_l$$

Taking the constraint conditions on the genetic effects into account, we can show that,

$$\sum_{i=1}^{n_1} \alpha^i x_{M_i}^{(1)} = \sum_{i=1}^{n_1} \alpha^i z_{M_i}^{(1)}$$

$$\sum_{i,j} \delta_j^i x_{M_i}^{(1)} x_{F_j}^{(1)} = \sum_{i,j} \delta_j^i z_{M_i}^{(1)} z_{F_j}^{(1)}$$

$$\sum_{i,k,l} (\alpha^i \gamma_l^k) x_{M_i}^{(1)} x_{M_k}^{(2)} x_{F_l}^{(2)} = \sum_{i,k,l} (\alpha^i \gamma_l^k) z_{M_i}^{(1)} z_{M_k}^{(2)} z_{F_l}^{(2)}$$

$$\sum_{i,j,k,l} (\delta_j^i \gamma_l^k) x_{M_i}^{(1)} x_{F_j}^{(1)} x_{M_k}^{(2)} x_{F_l}^{(2)} = \sum_{i,j,k,l} (\delta_j^i \gamma_l^k) z_{M_i}^{(1)} z_{F_j}^{(1)} z_{M_k}^{(2)} z_{F_l}^{(2)}$$

and so on. For example,

$$\sum_{i=1}^{n_1} \alpha^i x_{M_i}^{(1)} = \sum_{i=1}^{n_1} \alpha^i (z_{M_i}^{(1)} - p^i)$$

$$= \sum_{i=1}^{n_1} \alpha^i z_{M_i}^{(1)} - \sum_{i=1}^{n_1} \alpha^i p^i = \sum_{i=1}^{n_1} \alpha^i z_{M_i}^{(1)}$$

as  $\sum_{i=1}^{n_1} \alpha^i p^i = 0$  by the constrain condition (2). Therefore, model (5) can be rewritten as

$$G = \mu + \sum_{i=1}^{n_1} \alpha^i x_{M_i}^{(1)} + \sum_{j=1}^{n_1} \alpha^j x_{F_j}^{(1)} + \sum_{i,j} \delta_j^i x_{M_i}^{(1)} x_{F_j}^{(1)}$$

$$+ \sum_{k=1}^{n_2} \beta^k x_{M_k}^{(2)} + \sum_{l=1}^{n_2} \beta_l x_{F_l}^{(2)} + \sum_{k,l} \gamma_l^k x_{M_k}^{(2)} x_{F_l}^{(2)}$$

$$+ [\sum_{i,k} (\alpha^i \beta^k) x_{M_i}^{(1)} x_{M_k}^{(2)} + \sum_{i,l} (\alpha^i \beta_l) x_{M_i}^{(1)} x_{F_l}^{(2)}$$

$$+ \sum_{j,k} (\alpha_j \beta^k) x_{F_j}^{(1)} x_{M_k}^{(2)} + \sum_{j,l} (\alpha_j \beta_l) x_{F_j}^{(1)} x_{F_l}^{(2)}]$$

$$+ [\sum_{i,k,l} (\alpha^i \gamma_l^k) x_{M_i}^{(1)} x_{M_k}^{(2)} x_{F_l}^{(2)} + \sum_{j,k,l} (\alpha_j \gamma_l^k) x_{F_j}^{(1)} x_{M_k}^{(2)} x_{F_l}^{(2)}$$

$$+ [\sum_{i,j,k} (\delta_j^i \beta^k) x_{M_i}^{(1)} x_{F_j}^{(1)} x_{M_k}^{(2)} + \sum_{i,j,l} (\delta_j^i \beta_l) x_{M_i}^{(1)} x_{F_j}^{(1)} x_{F_l}^{(2)}]$$

$$+ \sum_{i,j,k,l} (\delta_j^i \gamma_l^k) x_{M_i}^{(1)} x_{F_j}^{(1)} x_{M_k}^{(2)} x_{F_l}^{(2)} \tag{7}$$

When we deal with some reduced models in the next section, we will find that the model of this form is especially helpful as it makes the model parameter constraints built into regression variables which is suited for genetic interpretation. The form of model (7) can also facilitate the demonstration that under Hardy-Weinberg, linkage and genotypic equilibria, the regression coefficients (genetic effects) are Cockerham's least squares effects (3) (Appen-

**Table 2: Partition of degrees of freedom for two loci with number of alleles  $n_1$  and  $n_2$  (a simplified case without distinguishing the paternal and maternal origins)**

Source	Degrees of Freedom
additive ( $\alpha$ )	$n_1 - 1$
dominance ( $\delta$ )	$n_1(n_1 - 1)/2$
additive ( $\beta$ )	$n_2 - 1$
dominance ( $\gamma$ )	$n_2(n_2 - 1)/2$
additive $\times$ additive ( $\alpha\beta$ )	$(n_1 - 1)(n_2 - 1)$
additive $\times$ dominance ( $\alpha\gamma$ )	$(n_1 - 1)n_2(n_2 - 1)/2$
dominance $\times$ additive ( $\delta\beta$ )	$n_1(n_1 - 1)(n_2 - 1)/2$
dominance $\times$ dominance ( $\delta\gamma$ )	$n_1(n_1 - 1)n_2(n_2 - 1)/4$
total	$\frac{n_1(n_1 + 1)}{2} \cdot \frac{n_2(n_2 + 1)}{2} - 1$

dix A), and the genotypic variance  $V_G$  has the orthogonal partition (4) (Appendix B).

Now we discuss the properties of model in a disequilibrium situation. As stated in [14], there are three types of disequilibria

- Type1: between alleles on the same gametes but at different loci
- Type2: between alleles at the same locus but on different gametes
- Type3: between alleles on different gametes and at different loci.

If we denote  $P_{jl}^{ik}$  as the genotypic frequency of  $A_i B_k / A_j B_l$ ,  $P_j^i$  as the genotypic frequency of  $A_i / A_j$ , and so on, following [14], the digenic disequilibria can be written as

$$D_{..}^{ik} = \text{Cov}(z_{M_i}^{(1)}, z_{M_k}^{(2)}) = E(x_{M_i}^{(1)} x_{M_k}^{(2)}) = P_{..}^{ik} - p^i q^k$$

$$D_{jl}^i = \text{Cov}(z_{F_j}^{(1)}, z_{F_l}^{(2)}) = E(x_{F_j}^{(1)} x_{F_l}^{(2)}) = P_{jl}^i - p_j q_l$$

$$D_{.j}^i = E(x_{M_i}^{(1)} x_{F_j}^{(1)}) = P_{.j}^i - p^i p_j$$

$$D_{.l}^k = E(x_{M_k}^{(2)} x_{F_l}^{(2)}) = P_{.l}^k - q^k q_l$$

$$D_{.j}^i = E(x_{M_i}^{(1)} x_{F_j}^{(2)}) = P_{.j}^i - p^i q_l$$

$$D_{.j}^k = E(x_{F_j}^{(1)} x_{M_k}^{(2)}) = P_{.j}^k - p_j q^k.$$

And the trigenic disequilibria

$$D_{.j}^{ik} = E(x_{M_i}^{(1)} x_{F_j}^{(1)} x_{M_k}^{(2)})$$

$$= P_{.j}^{ik} - p^i P_{.j}^k - q^k P_{.j}^i - p_j P_{..}^{ik} + 2p^i p_j q^k$$

$$D_{.l}^{ik} = E(x_{M_i}^{(1)} x_{M_k}^{(2)} x_{F_l}^{(2)})$$

$$= P_{.l}^{ik} - p^i P_{.l}^k - q^k P_{.l}^i - q_l P_{..}^{ik} + 2p^i q^k q_l$$

$$D_{jl}^i = E(x_{M_i}^{(1)} x_{F_j}^{(1)} x_{F_l}^{(2)})$$

$$= P_{jl}^i - p^i P_{jl}^i - p_j P_{..}^{ik} - q_l P_{.j}^i + 2p^i p_j q_l$$

$$D_{jl}^k = E(x_{F_j}^{(1)} x_{M_k}^{(2)} x_{F_l}^{(2)})$$

$$= P_{jl}^k - p_j P_{.l}^k - q^k P_{.j}^i - q_l P_{.j}^k + 2p_j q^k q_l$$

Similarly for the quadrigenic disequilibrium, we may define

$$D_{jl}^{ik} = E(x_{M_i}^{(1)} x_{F_j}^{(1)} x_{M_k}^{(2)} x_{F_l}^{(2)})$$

$$= P_{jl}^{ik} - p^i P_{jl}^k - p_j P_{.l}^{ik} - q^k P_{.j}^i - q_l P_{.j}^{ik}$$

$$+ p^i p_j P_{.l}^k - p^i q^k P_{jl}^i + p^i q_l P_{.j}^k + p_j q^k P_{.l}^i$$

$$+ p_j q_l P_{..}^{ik} + q^k q_l P_{.j}^i - 3p^i p_j q^k q_l.$$

If we express  $D_{jl}^{ik}$  as a function of lower-order linkage disequilibria, we have

$$D_{jl}^{ik} = P_{jl}^{ik} - p^i D_{jl}^k - p_j D_{.l}^{ik} - q^k D_{.j}^i - q_l D_{.j}^{ik}$$

$$- p^i p_j D_{.l}^k - p^i q^k D_{.j}^i - p^i q_l D_{.j}^k - p_j q^k D_{.l}^i$$

$$- p_j q_l D_{..}^{ik} - q^k q_l D_{.j}^i - p^i p_j q^k q_l.$$

This definition is the same as that given by [15,16]. Note that  $\sum_i z_{M_i}^{(1)} = 1$ . Then, we have

$$\sum_i D_{jl}^{ik} = E[(\sum_i x_{M_i}^{(1)}) x_{F_j}^{(1)} x_{M_k}^{(2)} x_{F_l}^{(2)}]$$

$$= E[\sum_i (z_{M_i}^{(1)} - p^i) x_{F_j}^{(1)} x_{M_k}^{(2)} x_{F_l}^{(2)}] = 0$$

In general,  $D_{jl}^{ik}$  is summed to zero over any allele involved, so are  $D_{.j}^{ik}$  and other disequilibrium measurements.

With Hardy-Weinberg and genotypic equilibria but linkage disequilibrium, model (7) leads to the following expression for the overall mean

$$E(G) = \mu + \sum_{i,k} (\alpha^i \beta^k) D_{..}^{ik} + \sum_{j,l} (\alpha_j \beta_l) D_{j\bar{l}} + \sum_{i,j,k,l} (\delta_j^i \gamma_l^k) D^{ik} D_{j\bar{l}}$$

where  $\mu$  is the mean genotypic value under linkage equilibrium, and

$$\Delta_\mu = \sum_{i,k} (\alpha^i \beta^k) D_{..}^{ik} + \sum_{j,l} (\alpha_j \beta_l) D_{j\bar{l}} + \sum_{i,j,k,l} (\delta_j^i \gamma_l^k) D^{ik} D_{j\bar{l}}$$

represents the departure from  $\mu$  due to linkage disequilibrium and epistasis. If there is no epistasis, linkage disequilibrium does not affect the mean genotypic value. Similar results were given by [17]. Note that for marginal means of the genotypic values, we have

$$G_{..} = \sum_{i,j,k,l} P_{ji}^{ik} P_{jl}^{ik} G_{jl}^{ik}, \quad G_{..}^i = \frac{1}{p^i} \sum_{j,k,l} P_{ji}^{ik} P_{jl}^{ik} G_{jl}^{ik},$$

$$G_{..}^{ik} = \sum_{j,l} P_{ji}^{ik} G_{jl}^{ik}, \quad G_{j.}^i = \frac{1}{p^i p_j} \sum_{k,l} P_{jl}^{ik} P_{..}^{ik} G_{jl}^{ik},$$

$$G_{j.}^{ik} = \frac{1}{p_j} \sum_l P_{jl}^{ik} G_{jl}^{ik}, \dots$$

and so on.

Then the question is what the genetic effects are in a disequilibrium population. Do Hardy-Weinberg and linkage disequilibria change the definition and values of genetic effects? The short answer to this question is "no" in a fully characterized model, but "yes" in a model that ignores some QTL or genetic effects. This is proved and discussed in [10]. With Hardy-Weinberg and linkage disequilibria, the genetic effects no longer correspond to the deviations from marginal means of genotypic values in a disequilibrium population. In a multiple regression model (7), the genetic effects are partial regression coefficients. These partial regression coefficients correspond to the simple regression coefficients, or deviations from marginal means of genotypic values, only in an equilibrium population. In a disequilibrium population, a direct analysis on the partial regression coefficients can be very complex (see the appendix of [10] for a relatively simple example). However, in a full model which includes all relevant loci and genetic effects, the model parameters depend only on how the regressors, i.e.  $x$  variables in (7), are defined and are independent of correlations between  $x$  variables, i.e. Hardy-Weinberg and linkage disequilibria. So, the genetic effects are still the same as those defined in the equilibrium population, although the population mean and

marginal means of genotypic values are changed in a disequilibrium population.

Hardy-Weinberg and linkage disequilibria introduce correlation between  $x$  variables, thus covariances between different genetic effect components. Define

$$A_1 = \sum_{i=1}^{n_1} \alpha^i x_{M_i}^{(1)} + \sum_{j=1}^{n_1} \alpha_j x_{F_j}^{(1)}$$

$$D_1 = \sum_{i,j} \delta_j^i x_{M_i}^{(1)} x_{F_j}^{(1)}$$

$$A_2 = \sum_{k=1}^{n_2} \beta^k x_{M_k}^{(2)} + \sum_{l=1}^{n_2} \beta_l x_{F_l}^{(2)}$$

$$D_2 = \sum_{k,l} \gamma_l^k x_{M_k}^{(2)} x_{F_l}^{(2)}$$

$$A_1 A_2 = \sum_{i,k} (\alpha^i \beta^k) x_{M_i}^{(1)} x_{M_k}^{(2)} + \sum_{i,l} (\alpha^i \beta_l) x_{M_i}^{(1)} x_{F_l}^{(2)} + \sum_{j,k} (\alpha_j \beta^k) x_{F_j}^{(1)} x_{M_k}^{(2)} + \sum_{j,l} (\alpha_j \beta_l) x_{F_j}^{(1)} x_{F_l}^{(2)}$$

$$A_1 D_2 = \sum_{i,k,l} (\alpha^i \gamma_l^k) x_{M_i}^{(1)} x_{M_k}^{(2)} x_{F_l}^{(2)} + \sum_{j,k,l} (\alpha_j \gamma_l^k) x_{F_j}^{(1)} x_{M_k}^{(2)} x_{F_l}^{(2)}$$

$$A_2 D_1 = \sum_{i,j,k} (\delta_j^i \beta^k) x_{M_i}^{(1)} x_{F_j}^{(1)} x_{M_k}^{(2)} + \sum_{i,j,l} (\delta_j^i \beta_l) x_{M_i}^{(1)} x_{F_j}^{(1)} x_{F_l}^{(2)}$$

$$D_1 D_2 = \sum_{i,j,k,l} (\delta_j^i \gamma_l^k) x_{M_i}^{(1)} x_{F_j}^{(1)} x_{M_k}^{(2)} x_{F_l}^{(2)}$$

Then we can write

$$G = \mu + A_1 + A_2 + D_1 + D_2 + A_1 A_2 + A_1 D_2 + A_2 D_1 + D_1 D_2$$

In a disequilibrium population, the partition of the genotypic variance becomes

$$V_G = \sum_{i=1}^8 \sum_{j=1}^8 V_{ij} = \mathbf{1}^T \mathbf{V} \mathbf{1}$$

where

$$V = (V_{ij})_{8 \times 8}$$

It is a symmetric matrix. In Appendix C, we give the detailed result for each component of the matrix with linkage disequilibrium, but assuming Hardy-Weinberg equilibrium.

For the rest of paper, when we discuss disequilibrium, we mainly discuss linkage disequilibrium and assume Hardy-Weinberg and genotypic equilibria which can be achieved by random mating in one generation. Hardy-Weinberg disequilibrium can be taken into account which will make results more complex and is thus omitted.

**Reduced models**

In many genetic applications, experimental population has some regular genetic structure by design. In these cases, the genetic model can be further simplified to reflect the experimental design structure. Also sometimes we may want to simplify the genetic model by imposing certain constraints or assumptions, such as the number of alleles, to increase the feasibility of analysis. In this section, we give a few reduced genetic models that are relevant to many genetic applications.

*1. Backcross population or recombinant inbred population (haploid model)*

Backcross population or recombinant inbred population is a common experimental design for QTL mapping study. By crossing two inbred lines, we can create a  $F_1$  population. If we randomly backcross  $F_1$  to one of the inbred lines, we have a backcross population. Let us assume that the cross is AA (paternal)  $\times$  Aa (maternal). In a random-mating backcross population, there are only two possible genotypes at each segregating locus  $A_rA_r$  or  $A_rA_r$ , for  $r = 1, 2, \dots, m$ , where  $m$  is the number of QTL. Since for the paternal gametes,

$$z_{M_1}^{(r)} = \begin{cases} 1, & \text{for } A_r \text{ allele from paternal gamete} \\ 0, & \text{otherwise} \end{cases} = 1,$$

and

$$z_{M_2}^{(r)} = \begin{cases} 1, & \text{for } a_r \text{ allele from paternal gamete} \\ 0, & \text{otherwise} \end{cases} = 0,$$

thus  $x_{M_1}^{(r)} = z_{M_1}^{(r)} - 1 = 0$  and  $x_{M_2}^{(r)} = z_{M_2}^{(r)} = 0$  for  $r = 1, 2, \dots,$

$m$ . For maternal gametes however,

$$x_{F_1}^{(r)} = \begin{cases} 1/2, & \text{for } A_r \text{ from maternal gamete} \\ -1/2, & \text{otherwise} \end{cases} \\ = -x_{F_2}^{(r)}, \text{ for } r = 1, 2, \dots, m.$$

Thus the model becomes

$$G = \mu + \sum_{r=1}^m a_r x_{F_1}^{(r)} + \sum_{r<s} b_{rs} (x_{F_1}^{(r)} x_{F_1}^{(s)}) \\ + \sum_{r<s<t} c_{rst} (x_{F_1}^{(r)} x_{F_1}^{(s)} x_{F_1}^{(t)}) + \dots \tag{8}$$

where  $a_r = \alpha_1^{(r)} - \alpha_2^{(r)}$  is the substitution effect between homozygote genotype  $A_rA_r$  and heterozygote genotype  $A_rA_r, b_{rs} = (\alpha_1^{(r)} \alpha_1^{(s)}) - (\alpha_1^{(r)} \alpha_2^{(s)}) - (\alpha_2^{(r)} \alpha_1^{(s)}) + (\alpha_2^{(r)} \alpha_2^{(s)})$  is the interaction effect between loci  $r$  and  $s, c_{rst} = (\alpha_1^{(r)} \alpha_1^{(s)} \alpha_1^{(t)}) - (\alpha_1^{(r)} \alpha_1^{(s)} \alpha_2^{(t)}) - (\alpha_1^{(r)} \alpha_2^{(s)} \alpha_1^{(t)}) + (\alpha_1^{(r)} \alpha_2^{(s)} \alpha_2^{(t)}) - (\alpha_2^{(r)} \alpha_1^{(s)} \alpha_1^{(t)}) + (\alpha_2^{(r)} \alpha_1^{(s)} \alpha_2^{(t)}) - (\alpha_2^{(r)} \alpha_2^{(s)} \alpha_1^{(t)}) + (\alpha_2^{(r)} \alpha_2^{(s)} \alpha_2^{(t)})$ ,  $\dots$ , and so on. Taking constraint conditions into account, we have  $\alpha_1 = -\alpha_2, \beta_1 = -\beta_2$ , and so on. Then,  $a_r = 2\alpha_1^{(r)}, b_{rs} = 4(\alpha_1^{(r)} \alpha_1^{(s)})$ , and  $c_{rst} = 8(\alpha_1^{(r)} \alpha_1^{(s)} \alpha_1^{(t)})$ , and so on. With linkage equilibrium, the genetic effects as the partial regression coefficients of the model correspond to the simple regression coefficients. For example, for the substitution effect of locus  $r, a_r$ , it is the covariance between genotypic value  $G$  and substitution effect design variable  $x_{F_1}^{(r)}$  divided by the variance of  $x_{F_1}^{(r)}$ . So in general, we have

$$a_r = E[(G - \mu)x_{F_1}^{(r)}] / E(x_{F_1}^{(r)2}) \\ = E[G(z_{F_1}^{(r)} - 1/2)] / (1/4) \\ = 2[E(G | z_{F_1}^{(r)} = 1) - E(G)] \\ b_{rs} = E[(G - \mu)x_{F_1}^{(r)} x_{F_1}^{(s)}] / [E(x_{F_1}^{(r)2})E(x_{F_1}^{(s)2})] \\ = 4[E(G | z_{F_1}^{(r)} = z_{F_1}^{(s)} = 1) - E(G | z_{F_1}^{(r)} = 1) \\ - E(G | z_{F_1}^{(s)} = 1) + E(G)] \\ c_{rst} = E[(G - \mu)x_{F_1}^{(r)} x_{F_1}^{(s)} x_{F_1}^{(t)}] / \\ [E(x_{F_1}^{(r)2})E(x_{F_1}^{(s)2})E(x_{F_1}^{(t)2})] \\ = 8[E(G | z_{F_1}^{(r)} = z_{F_1}^{(s)} = z_{F_1}^{(t)} = 1) \\ - E(G | z_{F_1}^{(r)} = z_{F_1}^{(s)} = 1) \\ - E(G | z_{F_1}^{(r)} = z_{F_1}^{(t)} = 1) \\ - E(G | z_{F_1}^{(s)} = z_{F_1}^{(t)} = 1) \\ + E(G | z_{F_1}^{(r)} = 1) + E(G | z_{F_1}^{(s)} = 1) \\ + E(G | z_{F_1}^{(t)} = 1) - E(G)]$$

The orthogonal partition of the genotypic variance in an equilibrium population is

$$V_G = \frac{1}{4} \sum_{r=1}^m a_r^2 + \frac{1}{4^2} \sum_{r<s} b_{rs}^2 + \frac{1}{4^3} \sum_{r<s<t} c_{rst}^2 + \dots \tag{9}$$



As noted above, linkage disequilibrium does not change the values of genetic effects in a full model. The model parameters are still the same as those defined in the equilibrium population. However, in this case there is a simple relationship between the substitution effects at multiple loci and marginal means of genotypic values in a disequilibrium population [18]. This is noted here. Let  $P_{rs} = P\{z_{F_1}^{(r)} = z_{F_1}^{(s)} = 1\}$ , and the digenic linkage disequilibrium be defined as

$$D_{rs} = \text{Cov}(z_{F_1}^{(r)}, z_{F_1}^{(s)}) = E(x_{F_1}^{(r)} x_{F_1}^{(s)})$$

Ignoring trigenic and higher order linkage disequilibria, we have

$$E\left[(G - \mu)x_{F_1}^{(r)}\right] = \frac{1}{4}a_r + \sum_{r' \neq r} D_{rr'}a_{r'}$$

$$E[(G - \mu)x_{F_1}^{(r)}x_{F_1}^{(s)}] = \frac{1}{4^2}b_{rs}$$

Therefore, the digenic interaction effects can be expressed as

$$b_{rs} = 4^2[P_{rs}E(G | z_{F_1}^{(r)} = z_{F_1}^{(s)} = 1) - D_{rs}\mu] - 4[E(G | z_{F_1}^{(r)} = 1) + E(G | z_{F_1}^{(s)} = 1) - E(G)]$$

Then the substitution effects can be expressed as a function of marginal means in the disequilibrium population as

$$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_l \end{pmatrix} = (I + 4D)^{-1}(2\mathbf{q})$$

where  $I$  is a  $m \times m$  identity matrix,  $D = (D_{ij})_{m \times m}$  with all diagonal elements being zeros;

$\mathbf{q} = (q_1, q_2, \dots, q_m)^T$ , with  $q_i = E(G | z_{F_1}^{(i)} = 1) - E(G)$ , for  $i = 1, 2, \dots, m$ .

The partition of genetic variance with linkage disequilibrium is complex. Here we give details of the partition of genotypic variance for the following model

$$G = \mu + \sum_{r=1}^m a_r x_{F_1}^{(r)} + \sum_{r<s} b_{rs} (x_{F_1}^{(r)} x_{F_1}^{(s)}) \quad (10)$$

Let  $x_r = x_{F_1}^{(r)}$  and  $x_s = x_{F_1}^{(s)}$  to simplify the notation here.

The genotypic variance is

$$\begin{aligned} V_G &= V\left(\sum_r a_r x_r\right) + 2\text{Cov}\left(\sum_r a_r x_r, \sum_{r<s} b_{rs} x_r x_s\right) \\ &+ V\left(\sum_{r<s} b_{rs} x_r x_s\right) \\ &= \sum_r a_r^2 p_r (1 - p_r) + 2 \sum_{r<s} a_r a_s D_{rs} \\ &+ 2 \sum_{r<s} [a_r b_{rs} (1 - 2p_r) D_{rs} + a_s b_{rs} (1 - 2p_s) D_{rs}] \\ &+ 2 \sum_{r<s<t} (a_r b_{st} + a_s b_{rt} + a_t b_{rs}) D_{rst} \\ &+ \sum_{r<s} b_{rs}^2 [(1 - 2p_r)(1 - 2p_s) D_{rs} \\ &+ p_r (1 - p_r) p_s (1 - p_s) - D_{rs}^2] \\ &+ 2 \sum_{r<s<t} \{b_{rs} b_{rt} [(1 - 2p_r) D_{rst} \\ &- p_r (1 - p_r) D_{st} - D_{rs} D_{rt}] \\ &+ b_{rs} b_{st} [(1 - 2p_s) D_{rst} + p_s (1 - p_s) D_{rt} - D_{rs} D_{st}] \\ &+ b_{rt} b_{st} [(1 - 2p_t) D_{rst} \\ &+ p_t (1 - p_t) D_{rs} - D_{rt} D_{st}]\} \\ &+ 2 \sum_{r<s<t<u} [b_{rt} b_{su} (D_{rstu} - D_{rt} D_{su}) \\ &+ b_{ru} b_{st} (D_{rstu} - D_{ru} D_{st}) \\ &+ b_{rs} b_{tu} (D_{rstu} - D_{rs} D_{tu})] \end{aligned} \quad (11)$$

where

$$D_{rst} = E(x_r x_s x_t) \text{ and } D_{rstu} = E(x_r x_s x_t x_u)$$

are three locus and four locus linkage disequilibria. This is a general partition of genetic variance for a haploid model.

For the backcross population, it can be shown that  $D_{rst} = 0$  (see Appendix D for both backcross and  $F_2$  populations) and  $D_{rstu} = D_{rs} D_{tu}$  for loci  $r, s, t$  and  $u$  in this order under the assumption of no crossing-over interference. Also with this assumption,  $D_{rt} = 4D_{rs} D_{st}$  and  $D_{rs} = (1 - 2\lambda_{rs})/4$ , where  $\lambda_{rs}$  is the recombination frequency between loci  $r$  and  $s$ . Since,  $p_r = p_s = 1/2$ , the variance becomes

$$\begin{aligned}
 V_G = & \frac{1}{4} \sum_r a_r^2 + 2 \sum_{r<s} a_r a_s D_{rs} + \\
 & \frac{1}{16} \sum_{r<s} b_{rs}^2 (1 - 16D_{rs}^2) \\
 & + \frac{1}{2} \sum_{r<s<t} \left[ b_{rs} b_{rt} (1 - 16D_{rs}^2) D_{st} \right. \\
 & \left. + b_{rt} b_{st} (1 - 16D_{st}^2) D_{rs} \right] \\
 & + 2 \sum_{r<s<t<u} (b_{rt} b_{su} + b_{ru} b_{st}) \\
 & (1 - 16D_{st}^2) D_{rs} D_{tu} \quad (12)
 \end{aligned}$$

In this partition of variance, the first summation term is the genetic variance due to the substitution effect of each QTL, the second summation term is the covariance between substitution effects of QTL pairs due to linkage disequilibrium, the third summation term is the genetic variance due to epistatic effects of QTL, and the fourth and fifth summation terms are the covariance between different epistatic effects of QTL due to linkage disequilibrium. There is no covariance between the main substitution effects and epistatic effects (see also [19]).

For a backcross population, the genetic interpretation of the substitution effect  $a_r$  depends on which parental line is backcrossed. In one backcross  $AA \times Aa$ , the substitution effect is traditionally defined as the difference between the additive effect and dominance effects, and in the other backcross  $Aa \times aa$ , it is the sum of the additive and dominance effects. Only with both backcrosses, can one estimate both additive and dominance effects separately (for example [20]).

The same model also applies to a recombinant inbred population which is another very popular experimental design for QTL mapping study. For a recombinant inbred population, the substitution effects of QTL are the additive effects and the epistatic effects are the additive  $\times$  additive interaction effects. Statistical methods to map QTL and to estimate various components of the genetic variance due to QTL including epistasis has been developed through the maximum likelihood approach [19,21]. In a few cases where the method was applied, we estimated, for the first time, how the quantitative genetic variance was partitioned into various components in designed experimental populations. For example, Weber *et al.* [22] reported the result of QTL mapping for wing shape on the third chromosome of *Drosophila melanogaster* from a cross of divergent selection lines. From 519 recombinant inbred lines, 11 QTL were mapped on the third chromosome. Nine QTL pairs showed significant epistatic effects. The total genetic variance amounts to 95.5% of the phenotypic variance in the recombinant inbred lines with phenotypes measured and averaged over 50 male flies for each recombinant

inbred line. The partition of the genetic variance is as follows (see Table 6 and 7 of [22]): 27.4% due to the variances of additive effects (equivalent to the first summation term of (12)); 67.3% due to the covariances between additive effects (the second summation term); 7.2% due to the variances of epistatic effects (the third summation term); and -6.0% due to the covariances between epistatic effects (the fourth and fifth summation terms). The covariances between additive and epistatic effects, expected to be 0, account for -0.4% due to sampling. Similar kind of partition of the genetic variance is also observed in a group of 701 second chromosome recombinant inbred lines from a cross of the same divergent selection lines (see Table 4 and 5 of [23]). See also [20] for another example.

## 2. F<sub>2</sub> population

F<sub>2</sub> is created from a cross between pairs of F<sub>1</sub> individuals. It is also a very popular experimental design for QTL mapping study. The advantage of this design is that both additive and dominance effects of a QTL can be estimated as well as various epistatic effects. The design also has more statistical power for QTL detection as compared to a backcross population. In a random-mating F<sub>2</sub> population, there are only two alleles at each segregating locus and allelic frequencies are expected to be one half if there is no segregation distortion.

Let us consider only two loci first. Let  $A$  and  $a$  denote the two alleles at locus 1, and  $B$  and  $b$  at locus 2. In this case,  $x_{M_1}^{(1)} = -x_{M_2}^{(1)}$  and  $x_{F_1}^{(1)} = -x_{F_2}^{(1)}$ . Assuming  $G_{jl}^{ij} = G_{ik}^{jl} = G_{jk}^{il} = G_{il}^{jk}$ , it also holds that  $\alpha^1 = \alpha_1 = -\alpha^2 = -\alpha_2$ ,  $\delta_1^1 = \delta_2^2 = -\delta_1^2$ , and so on. The additive term for locus 1 then becomes

$$A_1 = \alpha^1 (x_{M_1}^{(1)} - x_{M_2}^{(1)}) + \alpha_1 (x_{F_1}^{(1)} - x_{F_2}^{(1)}) = 2\alpha_1 w_1$$

with

$$w_1 = x_{M_1}^{(1)} + x_{F_1}^{(1)} = \begin{cases} 1, & \text{for } AA \text{ at locus 1} \\ 0, & \text{for } Aa \text{ at locus 1} \\ -1, & \text{for } aa \text{ at locus 1} \end{cases} \quad (13)$$

and the dominance term is

$$D_1 = 2(\delta_1^1)(x_{M_1}^{(1)} x_{F_1}^{(1)} + x_{M_1}^{(1)} x_{F_1}^{(1)}) = (-2)(\delta_1^1) v_1$$

with

$$v_1 = (-2)x_{M_1}^{(1)} x_{F_1}^{(1)} = \begin{cases} -1/2, & \text{for } AA \text{ at locus 1} \\ 1/2, & \text{for } Aa \text{ at locus 1} \\ -1/2, & \text{for } aa \text{ at locus 1} \end{cases} \quad (14)$$

Note that the  $v$  variable in this section for  $F_2$  differ, by a factor -2, from the  $v$  variable in the next section for a general two-allele model to conform to the usual definition for the  $F_2$  model. Similarly, for locus 2

$$A_2 = 2\beta_1 w_2 \text{ and } D_2 = (-2)(\gamma_1^2) v_2$$

with

$$w_2 = \begin{cases} 1, & \text{for } BB \text{ at locus 2} \\ 0, & \text{for } Bb \text{ at locus 2} \\ -1, & \text{for } bb \text{ at locus 2} \end{cases}$$

and

$$v_2 = \begin{cases} -1/2, & \text{for } BB \text{ at locus 2} \\ 1/2, & \text{for } Bb \text{ at locus 2} \\ -1/2, & \text{for } bb \text{ at locus 2} \end{cases} \quad (15)$$

The model can then be written as

$$G = \mu + a_1 w_1 + d_1 v_1 + a_2 w_2 + d_2 v_2 + (aa)_{12}(w_1 w_2) + (ad)_{12}(w_1 v_2) + (da)_{12}(v_1 w_2) + (dd)_{12}(v_1 v_2) \quad (16)$$

where the parameters are related as  $a_1 = 2\alpha_1$ ,  $a_2 = 2\beta_2$ ,  $d_1 = -2\delta_1^1$ ,  $d_2 = -2\gamma_1^1$ ,  $(aa)_{12} = 4(\alpha_1\beta_1)$ ,  $(ad)_{12} = 8(\alpha_1\gamma_1^1)$ ,  $(da)_{12} = 8(\delta_1^1\beta_1)$ ,  $(dd)_{12} = 16(\delta_1^1\gamma_1^1)$ . With random mating and linkage equilibrium, we have

$$a_1 = E[(G - \mu)w_1] / E(w_1^2) = 2(G_{i\cdot}^1 - G_{\cdot\cdot})$$

$$d_1 = E[(G - \mu)v_1] / E(v_1^2) = (-2)(G_{i\cdot}^1 - 2G_{i\cdot} + G_{\cdot\cdot})$$

$$a_2 = 2(G_{\cdot\cdot} - G_{\cdot\cdot})$$

$$d_2 = (-2)(G_{\cdot\cdot}^1 - 2G_{\cdot\cdot} + G_{\cdot\cdot})$$

$$(aa)_{12} = E[(G - \mu)w_1 w_2] / E(w_1^2 w_2^2) = 4(G_{\cdot\cdot}^{11} - G_{i\cdot} - G_{\cdot\cdot} + G_{\cdot\cdot})$$

$$(ad)_{12} = E[(G - \mu)w_1 v_2] / E(w_1^2 v_2^2) = (-4)(G_{i\cdot}^1 - 2G_{i\cdot} - G_{\cdot\cdot}^1 + G_{i\cdot} + 2G_{\cdot\cdot} - G_{\cdot\cdot})$$

$$(da)_{12} = E[(G - \mu)v_1 w_2] / E(v_1^2 w_2^2) = (-4)(G_{i\cdot}^1 - 2G_{i\cdot} - G_{\cdot\cdot}^1 + G_{i\cdot} + 2G_{\cdot\cdot} - G_{\cdot\cdot})$$

$$(dd)_{12} = E[(G - \mu)v_1 v_2] / E(v_1^2 v_2^2) = 4(G_{i\cdot}^{11} - 2G_{i\cdot}^1 - 2G_{i\cdot} + G_{i\cdot}^1 + G_{\cdot\cdot}^1 + 4G_{i\cdot} - 2G_{i\cdot} - 2G_{\cdot\cdot} + G_{\cdot\cdot})$$

The orthogonal partition of the genotypic variance is

$$V_G = \frac{1}{2}a_1^2 + \frac{1}{4}d_1^2 + \frac{1}{2}a_2^2 + \frac{1}{4}d_2^2 + \frac{1}{4}(aa)_{12}^2 + \frac{1}{8}(ad)_{12}^2 + \frac{1}{8}(da)_{12}^2 + \frac{1}{16}(dd)_{12}^2. \quad (17)$$

Recently, Kao and Zeng [18] have examined many genetic and statistical issues of the above  $F_2$  model and the effects of linkage disequilibrium. As we have shown here, the  $F_2$  model is a special case of Cockerham model with two alleles at each locus and all allelic frequencies being 1/2.

Now we give the partition of genetic variance for  $m$  loci with epistasis and linkage disequilibrium in the  $F_2$  population. Generalizing model (16) to  $m$  loci and ignoring the trigenic and higher order epistasis, we have the following model

$$G = \mu + \sum_{r=1}^m a_r w_r + \sum_{r=1}^m d_r v_r + \sum_{r<s} (aa)_{rs}(w_r w_s) + \sum_{r \neq s} (ad)_{rs}(w_r v_s) + \sum_{r<s} (dd)_{rs}(v_r v_s) = \mu + A + D + AA + Ad + DD \quad (18)$$

The partition of genetic variance for this model under the assumption of Hardy-Weinberg equilibrium is

$$V_G = V_A + V_D + V_{AA} + V_{AD} + V_{DD} + 2Cov(A, D) + 2Cov(A, AA) + 2Cov(A, AD) + 2Cov(A, DD) + 2Cov(D, AA) + 2Cov(D, AD) + 2Cov(D, DD) + 2Cov(AA, AD) + 2Cov(AA, DD) + 2Cov(AD, DD). \quad (19)$$

The detail of each component is presented in Appendix D.

The  $F_2$  model is a special case of the general two-allele model with  $p_r = 1/2$ . Note the difference on the  $v$  variable used for the  $F_2$  model and for the general two-allele model below. This partition of genetic variance can provide a basis for the interpretation of genetic variance estimation by multiple interval mapping in a  $F_2$  population [19,21].

### 3. A general two-allele model

Here, we provide details of a general two-allele model for multiple loci. This model is probably useful for studying genetic architecture of a quantitative trait in natural populations. Let the two alleles at locus  $r$  be  $A_r$  and  $a_r$ , for  $r = 1, 2, \dots, m$  with  $m$  the number of QTL. Assume that the frequencies and genetic effects of alleles are the same for both paternal and maternal gametes. Let  $p_r$  denote the frequency of allele  $A_r$  at locus  $r$ . Note that in this case

$$z_{M_1}^{(r)} = 1 - z_{M_2}^{(r)}, \quad r = 1, 2, \dots, m.$$

Also  $x_{M_1}^{(r)} = z_{M_1}^{(r)} - E[z_{M_1}^{(r)}] = (1 - z_{M_2}^{(r)}) - E(1 - z_{M_2}^{(r)}) = -x_{M_2}^{(r)}$ .

Similarly,  $x_{F_2}^{(r)} = -x_{F_1}^{(r)}$ . Ignoring higher order epistasis involving at least three loci, we can define a two-allele model as

$$G = \mu + \sum_{r=1}^m a_r w_r + \sum_{r=1}^m d_r v_r + \sum_{r < s} (aa)_{rs} (w_r w_s) + \sum_{r \neq s} (ad)_{rs} (w_r v_s) + \sum_{r < s} (dd)_{rs} (v_r v_s) \quad (20)$$

where

$$w_r = x_{M_1}^{(r)} + x_{F_1}^{(r)} = \begin{cases} (1 - p_r) & \text{for } A_r A_r \text{ at locus } r \\ 1 - 2p_r & \text{for } A_r a_r \text{ at locus } r \\ -2p_r & \text{for } a_r a_r \text{ at locus } r \end{cases} \quad (21)$$

$$v_r = x_{M_1}^{(x)} x_{F_1}^{(x)} = \begin{cases} (1 - p_r)^2 & \text{for } A_r A_r \text{ at locus } r \\ -p_r(1 - p_r) & \text{for } A_r a_r \text{ at locus } r \\ p_r^2 & \text{for } a_r a_r \text{ at locus } r \end{cases} \quad (22)$$

for  $r = 1, 2, \dots, m$ . The coefficients are associated with the original parameters in Cockerham model as follows.

$$\begin{aligned} a_r &= \alpha_1^{(r)} - \alpha_2^{(r)} \\ d_r &= \delta_1^{1(r)} - \delta_2^{1(r)} - \delta_1^{2(r)} + \delta_2^{2(r)} \\ (aa)_{rs} &= (\alpha_1^{(r)} \alpha_1^{(s)}) - (\alpha_1^{(r)} \alpha_2^{(s)}) - (\alpha_2^{(r)} \alpha_1^{(s)}) + (\alpha_2^{(r)} \alpha_2^{(s)}) \\ (ad)_{rs} &= (\alpha_1^{(r)} \delta_1^{1(s)}) - (\alpha_1^{(r)} \delta_2^{1(s)}) - (\alpha_1^{(r)} \delta_1^{2(s)}) + (\alpha_1^{(r)} \delta_2^{2(s)}) - (\alpha_2^{(r)} \delta_1^{1(s)}) + (\alpha_2^{(r)} \delta_2^{1(s)}) + (\alpha_2^{(r)} \delta_1^{2(s)}) - (\alpha_2^{(r)} \delta_2^{2(s)}) \\ (dd)_{rs} &= (\delta_1^{1(r)} \delta_1^{1(s)}) - (\delta_1^{1(r)} \delta_2^{1(s)}) - (\delta_1^{1(r)} \delta_1^{2(s)}) + (\delta_1^{1(r)} \delta_2^{2(s)}) - (\delta_2^{1(r)} \delta_1^{1(s)}) + (\delta_2^{1(r)} \delta_2^{1(s)}) + (\delta_2^{1(r)} \delta_1^{2(s)}) - (\delta_2^{1(r)} \delta_2^{2(s)}) - (\delta_1^{2(r)} \delta_1^{1(s)}) + (\delta_1^{2(r)} \delta_2^{1(s)}) - (\delta_1^{2(r)} \delta_1^{2(s)}) + (\delta_1^{2(r)} \delta_2^{2(s)}) - (\delta_2^{2(r)} \delta_1^{1(s)}) + (\delta_2^{2(r)} \delta_2^{1(s)}) - (\delta_2^{2(r)} \delta_1^{2(s)}) + (\delta_2^{2(r)} \delta_2^{2(s)}) \end{aligned}$$

The constraint conditions further lead to

$$a_r = \alpha_1^{(r)} - \frac{(-p_r)}{(1 - p_r)} \alpha_1^{(r)} = \frac{1}{(1 - p_r)} \alpha_1^{(r)}$$

$$d_r = \frac{1}{(1 - p_r)^2} \delta_1^{1(r)}$$

$$(aa)_{rs} = \frac{1}{(1 - p_r)(1 - p_s)} (\alpha_1^{(r)} \alpha_1^{(s)})$$

$$(ad)_{rs} = \frac{1}{(1 - p_r)(1 - p_s)^2} (\alpha_1^{(r)} \delta_1^{1(s)})$$

$$(dd)_{rs} = \frac{1}{(1 - p_r)^2(1 - p_s)^2} (\delta_1^{1(r)} \delta_1^{1(s)})$$

With Hardy-Weinberg, linkage and genotypic equilibria, the partial regression coefficients in the above model correspond to the simple regression coefficients

$$a_r = \frac{1}{(1 - p_r)} [E(G | z_{M_1}^{(r)} = 1) - E(G)]$$

$$d_r = \frac{1}{(1 - p_r)^2} [E(G | z_{M_1}^{(r)} = z_{F_1}^{(r)} = 1) - 2E(G | z_{M_1}^{(r)} = 1) + E(G)]$$

$$(aa)_{rs} = \frac{1}{(1 - p_r)(1 - p_s)} [E(G | z_{M_1}^{(r)} = z_{M_1}^{(s)} = 1) - E(G | z_{M_1}^{(r)} = 1) - E(G | z_{M_1}^{(s)} = 1) + E(G)]$$

$$(ad)_{rs} = \frac{1}{(1 - p_r)(1 - p_s)^2} [E(G | z_{M_1}^{(r)} = z_{M_1}^{(s)} = z_{F_1}^{(s)} = 1) - 2E(G | z_{M_1}^{(r)} = z_{M_1}^{(s)} = 1) - E(G | z_{M_1}^{(s)} = z_{F_1}^{(s)} = 1) + E(G | z_{M_1}^{(r)} = 1) + 2E(G | z_{M_1}^{(s)} = 1) - E(G)]$$

$$(dd)_{rs} = \frac{1}{(1 - p_r)^2(1 - p_s)^2} [E(G | z_{M_1}^{(r)} = z_{F_1}^{(r)} = z_{M_1}^{(s)} = z_{F_1}^{(s)} = 1) - 2E(G | z_{M_1}^{(r)} = z_{F_1}^{(r)} = z_{M_1}^{(s)} = 1) - 2E(G | z_{M_1}^{(r)} = z_{M_1}^{(s)} = z_{F_1}^{(s)} = 1) - E(G | z_{M_1}^{(r)} = z_{F_1}^{(r)} = 1) + E(G | z_{M_1}^{(s)} = z_{F_1}^{(s)} = 1) + 4E(G | z_{M_1}^{(r)} = z_{F_1}^{(r)} = 1) - 2E(G | z_{M_1}^{(r)} = 1) - 2E(G | z_{M_1}^{(s)} = 1) + E(G)]$$

Note that in this case the genetic effects in the original model are

$$\begin{aligned}
 \alpha_1^{(r)} &= E(G | z_{M_1}^{(r)} = 1) - E(G) \\
 \delta_1^{1(r)} &= E(G | z_{M_1}^{(r)} = z_{F_1}^{(r)} = 1) \\
 &\quad - 2E(G | z_{M_1}^{(r)} = 1) + E(G) \\
 (\alpha_1^{(r)} \alpha_1^{(s)}) &= E(G | z_{M_1}^{(r)} = z_{M_1}^{(s)} = 1) \\
 &\quad - E(G | z_{M_1}^{(r)} = 1) - E(G | z_{M_1}^{(s)} = 1) \\
 &\quad + E(G) \\
 (\alpha_1^{(r)} \delta_1^{1(s)}) &= E(G | z_{M_1}^{(r)} = z_{M_1}^{(s)} = z_{F_1}^{(s)} = 1) \\
 &\quad - 2E(G | z_{M_1}^{(r)} = z_{M_1}^{(s)} = 1) \\
 &\quad - E(G | z_{M_1}^{(s)} = z_{F_1}^{(s)} = 1) \\
 &\quad + E(G | z_{M_1}^{(r)} = 1) + 2E(G | z_{M_1}^{(s)} = 1) \\
 &\quad - E(G) \\
 (\delta_1^{1(r)} \delta_1^{1(s)}) &= E(G | z_{M_1}^{(r)} = z_{F_1}^{(r)} = z_{M_1}^{(s)} = z_{F_1}^{(s)} = 1) \\
 &\quad - 2E(G | z_{M_1}^{(r)} = z_{F_1}^{(r)} = z_{M_1}^{(s)} = 1) \\
 &\quad - 2E(G | z_{M_1}^{(r)} = z_{M_1}^{(s)} = z_{F_1}^{(s)} = 1) \\
 &\quad + E(G | z_{M_1}^{(r)} = z_{F_1}^{(r)} = 1) \\
 &\quad + E(G | z_{M_1}^{(s)} = z_{F_1}^{(s)} = 1) \\
 &\quad + 4E(G | z_{M_1}^{(r)} = z_{F_1}^{(s)} = 1) \\
 &\quad - 2E(G | z_{M_1}^{(r)} = 1) - 2E(G | z_{M_1}^{(s)} = 1) \\
 &\quad + E(G)
 \end{aligned}$$

They are the same as the least squares definition.

Yet another form of this result is shown in Table 1 of [10]. Zeng et al. [10] also show that linkage disequilibrium does not change the values of genetic effects in a full model. This means that the partial regression coefficients in a disequilibrium population equal to the simple regression coefficients in a corresponding equilibrium population with the same allelic frequency configuration.

The partition of genotypic variance in an equilibrium population is

$$\begin{aligned}
 V_G &= 2 \sum_{r=1}^m p_r (1-p_r) a_r^2 + \sum_{r=1}^m p_r^2 (1-p_r)^2 d_r^2 \\
 &\quad + 4 \sum_{r<s} p_r p_s (1-p_r)(1-p_s) (aa)_{rs}^2 \\
 &\quad + 2 \sum_{r \neq s} p_r (1-p_r) p_s^2 (1-p_s)^2 (ad)_{rs}^2 \\
 &\quad + \sum_{r<s} p_r^2 p_s^2 (1-p_r)^2 (1-p_s)^2 (dd)_{rs}^2
 \end{aligned} \tag{23}$$

or

$$\begin{aligned}
 V_G &= \frac{2p_r}{(1-p_r)} \sum_{r=1}^m (\alpha_1^{(r)})^2 + \frac{p_r^2}{(1-p_r)^2} \sum_{r=1}^m (\delta_1^{1(r)})^2 \\
 &\quad + \frac{4p_r p_s}{(1-p_r)(1-p_s)} \sum_{r<s} (\alpha_1^{(r)} \alpha_1^{(s)})^2 \\
 &\quad + \frac{2p_r p_s^2}{(1-p_r)(1-p_s)^2} \sum_{r \neq s} (\alpha_1^{(r)} \delta_1^{1(s)})^2 \\
 &\quad + \frac{p_r^2 p_s^2}{(1-p_r)^2 (1-p_s)^2} \sum_{r<s} (\delta_1^{1(r)} \delta_1^{1(s)})^2
 \end{aligned} \tag{24}$$

The partition of the genetic variance with epistasis and linkage disequilibrium is complex. We give the result with trigenic and quadrigenic linkage disequilibria included as well. The partition of variance has a similar form as (19). The detail of each component is presented in Appendix E.

### Discussion

In this paper we explore various properties of the standard quantitative genetic model with multiple interacting loci in linkage equilibrium and disequilibrium. Starting from the traditional least squares model, we represent it in the setting of multiple regression with standardized allelic indicator variables and their products as the independent variables and the trait value as the dependent variable. Then the partial regression coefficients associated with these indicator variables define the additive, dominance and epistatic effects for QTL. This is the original definition of QTL effects introduced by Fisher [6] and extended to epistasis by Cockerham [7]. We examine the properties and meaning of these QTL effects in an equilibrium population and also in a disequilibrium population. We show details of the partition of genetic variance for both equilibrium and disequilibrium populations in terms of QTL effects, allelic frequencies and disequilibrium measures. Moreover, we relate this general model to several reduced models used for QTL mapping analysis in cross populations from inbred lines, such as  $F_2$ , backcross and recombinant inbred lines. The detailed partition of genetic variance in these populations can provide a basis for the interpretation of genetic variance component estimates from multiple interval mapping [21].

The purpose of modeling QTL is to provide a meaningful and convenient framework and basis to infer and interpret relative significance of each QTL and intricate inter-relationship among QTL on a set of quantitative traits in an experimental or natural population for genetic study. The linear model provides a framework to study the effects of QTL on the mean and variance of the distribution of a trait or multiple traits in a population. With the assumption of a normal distribution for both genotypic and phenotypic values of a quantitative trait or multiple traits, this analysis on the first and second order statistics is sufficient to characterize the relationship between QTL and trait (s). Otherwise, it is an approximation on the relationship. In this model, the model parameters are partitioned into two parts: one is the effects of QTL (additive, dominance and epistatic effects), and the other is frequencies and correlations (Hardy-Weinberg and linkage disequilibria) of QTL alleles. Together they characterize the genetic architecture of quantitative traits in a population.

This linear model also provides a framework for statistical inference of genetic model parameters. If QTL genotypes are known and directly observed, a regression analysis of trait phenotype on QTL genotypes would provide a direct estimation of the genetic model parameters. However, if QTL genotypes are unknown and are only indirectly observed through molecular markers, the statistical inference of QTL and model parameters becomes more complicated. Statistically, we can regard QTL genotypes as missing data with trait phenotypes and marker genotypes as observed data and use a mixture model through the maximum likelihood analysis to infer the conditional distribution of missing data and through that to infer QTL parameters which also include the number and genomic position of QTL [19,24,25]. The likely positions of QTL are searched in the whole genome if data permit and the number of significant QTL positions can be estimated through some model selection procedure.

On modeling QTL, the consistence of model parameters in a multiple-locus setting is an important consideration. It is important for a model to be multiple-locus consistent, and the relationship within and between loci can be clearly and readily analyzed, estimated and interpreted. Here the consistence means that the effect of a QTL is consistently defined in a reference equilibrium population for one, two or multiple loci. In statistics, this is the property of orthogonality. This property is particularly important for the study of epistasis. With that the additive, dominance and epistatic effects can be independently and consistently estimated for one, two, three or multiple loci in the reference population where the model is defined and interpreted. Thus, if the number of QTL is incorrectly identified which seems to be always the case in practice, the parameter values for those identified QTL can still be

consistently estimated. However, the situation would certainly be different and complicated if the population is not at equilibrium, for example for QTL in linkage disequilibrium. Linkage disequilibrium would complicate the partition of genetic variance, and could certainly bias the estimation of parameter values for those identified QTL if the QTL model (number and genomic position of QTL) is miss-identified.

In this paper, we study extensively the composition and property of the genetic model parameters, such as genetic effects and partition of genetic variance, when both epistasis and linkage disequilibrium are considered. This would help us to understand the relationship of various genetic quantities, such as allelic frequencies and linkage disequilibrium, on the definition of genetic effects. It would also help us to understand and properly interpret estimates of the genetic effects and variance components in a QTL mapping experiment. It is important to emphasize that modeling QTL is inherently population based as it defines the variation of QTL in reference to a population, either a study population, cross population or natural population. The very basic concept of additive effect of a QTL is a population concept and is population dependent. It depends on the genotypes at other loci and depends on the genetic structure of the population (allelic frequencies, Hardy-Weinberg and linkage disequilibria).

We also clarify the connection between the general genetic model and some reduced models. By restricting the number of alleles at each locus to two and setting allelic frequencies to half, the general genetic model is reduced to the  $F_2$  model. This simplification reduces the partition of genetic variance enormously.

Another property for this  $F_2$  population is that, if there is no crossing-over interference, the three-locus linkage disequilibrium is expected to be zero regardlessly whether the loci are linked. Also the four-locus disequilibrium is reduced to the product of the two-locus disequilibria for the two non-adjacent locus pairs. If there is crossing-over interference, the three-locus linkage disequilibrium would be a good measure of the interference. As many QTL mapping experiments are performed in a  $F_2$  population, this reduced model is very relevant to QTL mapping analysis for the interpretation of genetic architecture in a  $F_2$  population. Another reduced model is the backcross model which is essentially a haploid model.

We give many details for a general two-allele model with epistasis and linkage disequilibrium. Research on QTL mapping analysis has been shifted in recent years from inbred line crosses to natural populations. With the availability of very dense SNP markers, it is now possible to use SNP for fine mapping of QTL in a natural population.

Currently most QTL fine mapping studies are concentrated on candidate genes. It will be increasingly possible to have genome-wide SNP data for a sample of individuals from a natural population. The general two-allele model can be used as a framework to interpret and estimate the genome-wide genetic architecture for a quantitative trait in a natural population. The model can be extended to multiple alleles to take haplotypes into account if needed.

**Authors' contributions**

TW conducted the initial derivation of models and partition of variance and drafted the initial manuscript. ZBZ extended the derivation of models and partition of variance, revised and finalized the manuscript. Both authors have read and approved the final manuscript. Comments and requests should be addressed to ZBZ or TW.

**Appendix**

**A. Cockerham least squares estimates**

In this appendix, we show that the regression coefficients (genetic effects) in model (7) are Cockerham least squares estimates under Hardy-Weinberg, linkage and genotypic equilibria. First, note that at each locus only one allele is present on a gamete. That is if an individual inherits an allele  $A_i$  from a parental gamete,  $z_{M_i}^{(1)} = 1$  and the other

$$z_{M_j}^{(1)} = 0 \text{ for } j \neq i \text{ Therefore, when } i \neq j, \text{ we have } E(z_{M_i}^{(1)} z_{M_j}^{(1)}) = 0 \text{ or } E(x_{M_i}^{(1)} x_{M_j}^{(1)}) = E(z_{M_i}^{(1)} z_{M_j}^{(1)}) - E(z_{M_i}^{(1)})E(z_{M_j}^{(1)}) = -p^i p^j .$$

Using these relationships, we can show the following for model (7).

- Additive effects: For  $i = 1, 2, \dots, n_1$ , we can show

$$E[(G - \mu)x_{M_i}^{(1)}] = E[(\sum_{i'} \alpha^{i'} x_{M_{i'}}^{(1)})x_{M_i}^{(1)}] = \alpha^i E(x_{M_i}^{(1)2}) + \sum_{i' \neq i} \alpha^{i'} E(x_{M_i}^{(1)} x_{M_{i'}}^{(1)}) = \alpha^i p^i (1 - p^i) + \sum_{i' \neq i} \alpha^{i'} (-p^{i'} p^i) = \alpha^i p^i + (-p^i) \sum_{i'} \alpha^{i'} p^{i'} = \alpha^i p^i ,$$

On the other hand,

$$E[(G - \mu)x_{M_i}^{(1)}] = E(G x_{M_i}^{(1)}) = E[G(z_{M_i}^{(1)} - p^i)] = E(Gz_{M_i}^{(1)}) - p^i E(G) = E[E(Gz_{M_i}^{(1)} | z_{M_i}^{(1)})] - p^i G_{..} = p^i E(G | z_{M_i}^{(1)} = 1) - p^i G_{..} = p^i (G_{.i} - G_{..}) .$$

Therefore,  $\alpha^i = G_{.i} - G_{..}$  for  $i = 1, 2, \dots, n_1$ . Similarly, we can show that  $\alpha_j = G_{.j} - G_{..}$  for  $j = 1, 2, \dots, n_1$ , and

$$\beta^k = G_{..} - G_{.k} , \beta_l = G_{.l} - G_{..} \text{ for } k, l = 1, 2, \dots, n_2 .$$

- Dominance effects: For locus 1, we have

$$\begin{aligned} E[(G - \mu)x_{M_i}^{(1)} x_{F_j}^{(1)}] &= E[(\sum_{i',j'} \delta_{i',j'}^{i'} x_{M_{i'}}^{(1)} x_{F_{j'}}^{(1)})x_{M_i}^{(1)} x_{F_j}^{(1)}] \\ &= \sum_{j'} \delta_{j'}^i E(x_{M_i}^{(1)2}) E(x_{F_j}^{(1)} x_{F_{j'}}^{(1)}) + \sum_{i' \neq i} \sum_{j'} \delta_{i',j'}^{i'} E(x_{M_i}^{(1)} x_{M_{i'}}^{(1)}) E(x_{F_j}^{(1)} x_{F_{j'}}^{(1)}) \\ &= p^i (1 - p^i) \sum_{j'} \delta_{j'}^i E(x_{F_j}^{(1)} x_{F_{j'}}^{(1)}) + \sum_{i' \neq i} (-p^i p^{i'}) \sum_{j'} \delta_{i',j'}^{i'} E(x_{F_j}^{(1)} x_{F_{j'}}^{(1)}) \\ &= p^i \sum_{j'} \delta_{j'}^i E(x_{F_j}^{(1)} x_{F_{j'}}^{(1)}) - \sum_{i'} p^i p^{i'} \sum_{j'} \delta_{i',j'}^{i'} E(x_{F_j}^{(1)} x_{F_{j'}}^{(1)}) \\ &= p^i (p_j \delta_j^i + \sum_{j'} \delta_{j'}^i p_{j'}) - p^i \sum_{j'} E(x_{F_j}^{(1)} x_{F_{j'}}^{(1)}) \sum_{i'} p^{i'} \delta_{i',j'}^{i'} \\ &= p^i p_j \delta_j^i \text{ for } i, j = 1, 2, \dots, n_1 . \end{aligned}$$

On the other hand,

$$\begin{aligned} E[(G - \mu)x_{M_i}^{(1)} x_{F_j}^{(1)}] &= E[G(z_{M_i}^{(1)} - p^i)(z_{F_j}^{(1)} - p_j)] \\ &= E[Gz_{M_i}^{(1)} z_{F_j}^{(1)} - p^i E(Gz_{F_j}^{(1)}) - p_j E(Gz_{M_i}^{(1)}) + p^i p_j E(G)] \\ &= E[E(Gz_{M_i}^{(1)} z_{F_j}^{(1)} | z_{M_i}^{(1)}, z_{F_j}^{(1)})] - p^i E[E(Gz_{F_j}^{(1)} | z_{F_j}^{(1)})] - p_j E[E(Gz_{M_i}^{(1)} | z_{M_i}^{(1)})] + p^i p_j G_{..} \\ &= p^i p_j (G_{.j} - G_{..} - G_{.i} + G_{..}) \end{aligned}$$

Therefore,  $\delta_j^i = G_{.j} - G_{..} - G_{.i} + G_{..}$  for  $i, j = 1, 2, \dots, n_1$ . Similar results can be derived for other dominance terms at locus 2.

- Additive x additive effects: Note that

$$\begin{aligned} E[(G - \mu)x_{M_i}^{(1)} x_{M_k}^{(2)}] &= \sum_{i',k'} (\alpha^{i'} \beta^{k'}) E(x_{M_i}^{(1)} x_{M_{i'}}^{(1)}) E(x_{M_k}^{(2)} x_{M_{k'}}^{(2)}) \\ &= \sum_k (\alpha^i \beta^k) E(x_{M_i}^{(1)2}) E(x_{M_k}^{(2)} x_{M_k}^{(2)}) + \sum_{i' \neq i} \sum_{k'} (\alpha^{i'} \beta^{k'}) (-p^i p^{i'}) E(x_{M_k}^{(2)} x_{M_{k'}}^{(2)}) \\ &= p^i \sum_{k'} (\alpha^i \beta^{k'}) E(x_{M_k}^{(2)} x_{M_{k'}}^{(2)}) - p^i \sum_{i'} \sum_{k'} (\alpha^{i'} \beta^{k'}) p^{i'} E(x_{M_k}^{(2)} x_{M_{k'}}^{(2)}) \\ &= p^i q^k (\alpha^i \beta^k) \end{aligned}$$

and

$$E[(G - \mu)x_{M_i}^{(1)} x_{M_k}^{(2)}] = E[Gx_{M_i}^{(1)} x_{M_k}^{(2)}] = E[G(z_{M_i}^{(1)} - p^i)(z_{M_k}^{(2)} - q^k)] = p^i q^k (G_{.ik} - G_{..} - G_{.i} + G_{..})$$

We have  $(\alpha^i \beta^k) = G_{.ik} - G_{..} - G_{.i} + G_{..}$  for  $i = 1, 2, \dots, n_1$  and  $k = 1, 2, \dots, n_2$ . Similar results can be derived for other additive by additive terms.

- Additive x dominance effects: Note that

$$\begin{aligned} E[(G - \mu)x_{M_i}^{(1)} x_{M_k}^{(2)} x_{F_l}^{(2)}] &= E[x_{M_i}^{(1)} x_{M_k}^{(2)} x_{F_l}^{(2)} \sum_{i',k',l'} (\alpha^{i'} \gamma^{k'l'}) x_{M_{i'}}^{(1)} x_{M_{k'}}^{(2)} x_{F_{l'}}^{(2)}] \\ &= \sum_{k',l'} (\alpha^i \gamma_{k'l'}^i) E(x_{M_i}^{(1)2}) E(x_{M_k}^{(2)} x_{M_{k'}}^{(2)}) E(x_{F_l}^{(2)} x_{F_{l'}}^{(2)}) = p^i \sum_{k',l'} (\alpha^i \gamma_{k'l'}^i) E(x_{M_k}^{(2)} x_{M_{k'}}^{(2)}) E(x_{F_l}^{(2)} x_{F_{l'}}^{(2)}) = p^i q^k q_l (\alpha^i \gamma_{k'l'}^i) \end{aligned}$$

and

$$E[(G - \mu)x_{M_i}^{(1)}x_{M_k}^{(2)}x_{F_j}^{(2)}] = E[G(z_{M_i}^{(1)} - p^i)(z_{M_k}^{(2)} - q^k)(z_{F_j}^{(2)} - q_l)]$$

$$= p^i q^k q_l (G_{..}^{ik} - G_{..}^{ik} - G_{.i}^{.k} - G_{.i}^{.k} + G_{..}^{.i} + G_{..}^{.k} + G_{.i}^{.j} - G_{..}^{.j})$$

For  $i = 1, 2, \dots, n_1$  and  $k, l = 1, 2, \dots, n_2$ , we have

$$(\alpha^i \gamma_l^k) = G_{.i}^{.k} - G_{..}^{.k} - G_{.i}^{.j} - G_{.i}^{.k} + G_{..}^{.i} + G_{..}^{.k} + G_{.i}^{.j} - G_{..}^{.j}$$

Similar results can be derived for dominance by additive terms.

• Dominance × dominance effects: Note that

$$E[(G - \mu)x_{M_i}^{(1)}x_{F_j}^{(1)}x_{M_k}^{(2)}x_{F_j}^{(2)}] = E[x_{M_i}^{(1)}x_{F_j}^{(1)}x_{M_k}^{(2)}x_{F_j}^{(2)} \sum_{i',j',k',l'} (\delta_{i',j',k',l'}^i \gamma_{i',j',k',l'}^k) x_{M_i'}^{(1)}x_{F_j'}^{(1)}x_{M_k'}^{(2)}x_{F_j'}^{(2)}]$$

$$= \sum_{i',j'} E(x_{M_i}^{(1)}x_{M_k'}^{(1)})E(x_{F_j}^{(1)}x_{F_j'}^{(1)}) \sum_{k',l'} (\delta_{i',j',k',l'}^i \gamma_{i',j',k',l'}^k) E(x_{M_k}^{(2)}x_{M_k'}^{(2)})E(x_{F_j}^{(2)}x_{F_j'}^{(2)})$$

$$= \sum_{i',j'} E(x_{M_i}^{(1)}x_{M_i'}^{(1)})E(x_{F_j}^{(1)}x_{F_j'}^{(1)}) (\delta_{i',j',k',l'}^i \gamma_{i',j',k',l'}^k) q^k q_l = (\delta_{i',j',k',l'}^i \gamma_{i',j',k',l'}^k) p^i p_j q^k q_l$$

On the other hand

$$E[(G - \mu)x_{M_i}^{(1)}x_{F_j}^{(1)}x_{M_k}^{(2)}x_{F_j}^{(2)}] = E[G(z_{M_i}^{(1)} - p^i)(x_{F_j}^{(1)} - p_j)(z_{M_k}^{(2)} - q^k)(z_{F_j}^{(2)} - q_l)]$$

$$= p^i p_j q^k q_l (G_{.ij}^{.k} - G_{.ij}^{.k} - G_{.i}^{.k} - G_{.i}^{.k} + G_{.i}^{.j} + G_{.i}^{.k} + G_{.j}^{.i} - G_{..}^{.j} - G_{..}^{.k} - G_{.j}^{.i} - G_{.j}^{.k} + G_{..}^{.i} + G_{..}^{.k})$$

For  $i, j = 1, 2, \dots, n_1$  and  $k, l = 1, 2, \dots, n_2$ , we have

$$(\delta_{i',j',k',l'}^i \gamma_{i',j',k',l'}^k) = G_{.ij}^{.k} - G_{.ij}^{.k} - G_{.i}^{.k} - G_{.i}^{.k} + G_{.i}^{.j} + G_{.i}^{.k} + G_{.j}^{.i} + G_{.j}^{.k} + G_{.i}^{.l} - G_{..}^{.l} - G_{..}^{.k} - G_{.j}^{.i} - G_{.j}^{.k} + G_{..}^{.i} + G_{..}^{.k}$$

**B. Partition of genotypic variance in linkage equilibrium**

Here we show that the genotypic variance  $V_G$  of model (7) has the orthogonal partition (4) under Hardy-Weinberg, linkage and genotypic equilibria. First, note that the index variables  $x_{M_i}^{(1)}, x_{F_j}^{(1)}, x_{M_k}^{(2)}, x_{F_j}^{(2)}$  have expectation zero. Second, the assumption of Hardy-Weinberg, linkage and genotypic equilibria mean that all alleles in different gametes and loci are independent so that, for example,

$$\text{Cov}(x_{M_i}^{(r)}, x_{F_j}^{(r)}) = E(x_{M_i}^{(r)}x_{F_j}^{(r)}) = E(x_{M_i}^{(r)})E(x_{F_j}^{(r)}) = 0$$

Thus the additive and dominance effects within a locus are orthogonal to each other because

$$\text{Cov}(x_{M_k}^{(r)}, x_{M_i}^{(r)}, x_{F_j}^{(r)}) = E(x_{M_k}^{(r)}x_{M_i}^{(r)}x_{F_j}^{(r)}) = E(x_{M_k}^{(r)})E(x_{M_i}^{(r)})E(x_{F_j}^{(r)}) = 0$$

for any  $1 \leq i, j, k \leq n_1$  and locus  $r = 1$  or  $2$ . Similarly, the epistatic effects between loci are orthogonal to additive and dominance effects and also to other epistatic effects.

Therefore, the total genotypic variance  $V_G$  can be partitioned as

$$V_G = V_{A_1} + V_{A_2} + V_{D_1} + V_{D_2} + V_{A_1A_2} + V_{A_1D_2} + V_{D_1A_2} + V_{D_1D_2}$$

with each component analyzed below.

• The additive variance: For locus 1,

$$V_{A_1} = \text{Var}(\sum_{i=1}^{n_1} \alpha^i x_{M_i}^{(1)} + \sum_{j=1}^{n_1} \alpha_j x_{F_j}^{(1)}) = \text{Var}(\sum_{i=1}^{n_1} \alpha^i x_{M_i}^{(1)}) + \text{Var}(\sum_{j=1}^{n_1} \alpha_j x_{F_j}^{(1)})$$

$$= E[(\sum_{i=1}^{n_1} \alpha^i x_{M_i}^{(1)})^2] + E[(\sum_{j=1}^{n_1} \alpha_j x_{F_j}^{(1)})^2]$$

$$= E[(\sum_{i=1}^{n_1} \alpha^i z_{M_i}^{(1)} - \sum_{i=1}^{n_1} \alpha^i p^i)^2] + E[(\sum_{j=1}^{n_1} \alpha_j x_{F_j}^{(1)} - \sum_{j=1}^{n_1} \alpha_j p_j)^2]$$

$$= E[(\sum_{i=1}^{n_1} \alpha^i z_{M_i}^{(1)})^2] + E[(\sum_{j=1}^{n_1} \alpha_j z_{F_j}^{(1)})^2]$$

$$= \sum_{i=1}^{n_1} (\alpha^i)^2 E[(z_{M_i}^{(1)})^2] + \sum_{j=1}^{n_1} (\alpha_j)^2 E[(z_{F_j}^{(1)})^2]$$

$$= \sum_{i=1}^{n_1} (\alpha^i)^2 p^i + \sum_{j=1}^{n_1} (\alpha_j)^2 p_j$$

as  $\sum_{i=1}^{n_1} \alpha^i p^i = 0$  by the constrain condition (2). Similarly,

$$\text{for locus 2, we have } V_{A_2} = \sum_{k=1}^{n_2} (\beta^k)^2 q^k + \sum_{j=1}^{n_2} (\beta_j)^2 q_l .$$

• The dominance variance: For locus 1,

$$V_{D_1} = \text{Var}(\sum_{i,j} \delta_{ij}^i x_{M_i}^{(1)}x_{F_j}^{(1)}) = E[(\sum_{i,j} \delta_{ij}^i (z_{M_i}^{(1)} - p^i)(z_{F_j}^{(1)} - p_j))^2]$$

$$= E[(\sum_{i,j} \delta_{ij}^i z_{M_i}^{(1)}z_{F_j}^{(1)} - \sum_{i,j} p_j \delta_{ij}^i z_{M_i}^{(1)} - \sum_{i,j} p^i \delta_{ij}^i z_{F_j}^{(1)} + \sum_{i,j} p^i p_j \delta_{ij}^i)^2]$$

$$= E[(\sum_{i,j} \delta_{ij}^i z_{M_i}^{(1)}z_{F_j}^{(1)})^2] = \sum_{i,j} (\delta_{ij}^i)^2 E(z_{M_i}^{(1)2})E(z_{F_j}^{(1)2}) = \sum_{i,j} (\delta_{ij}^i)^2 p^i p_j$$

$$\text{Similarly, for locus 2, we have } V_{D_2} = \sum_{k,l} (\gamma_{kl}^k)^2 q^k q_l .$$

• The additive × additive variance:

$$V_{A_1A_2} = \text{Var}(\sum_{i,k} (\alpha^i \beta^k) x_{M_i}^{(1)}x_{M_k}^{(2)}) + \text{Var}(\sum_{i,l} (\alpha^i \beta_l) x_{M_i}^{(1)}x_{F_l}^{(2)})$$

$$+ \text{Var}(\sum_{j,k} (\alpha_j \beta^k) x_{F_j}^{(1)}x_{M_k}^{(2)}) + \text{Var}(\sum_{j,l} (\alpha_j \beta_l) x_{F_j}^{(1)}x_{F_l}^{(2)})$$

$$= \sum_{i,k} p^i q^k (\alpha^i \beta^k)^2 + \sum_{j,l} p_j q_l (\alpha_j \beta_l)^2 + \sum_{i,l} p^i q_l (\alpha^i \beta_l)^2 + \sum_{j,k} p_j q^k (\alpha_j \beta^k)^2$$

• The additive × dominance variance:



$$V_{A_1D_2} = \text{Var}\left(\sum_{i,k,l} (\alpha^i \gamma_l^k) x_{M_i}^{(1)} x_{M_k}^{(2)} x_{F_j}^{(2)}\right) + \text{Var}\left(\sum_{j,k,l} (\alpha_j \gamma_l^k) x_{F_j}^{(1)} x_{M_k}^{(2)} x_{F_j}^{(2)}\right)$$

$$= \sum_{i,k,l} p^i q^k q_l (\alpha^i \gamma_l^k)^2 + \sum_{j,k,l} p_j q^k q_l (\alpha_j \gamma_l^k)^2$$

Similarly, the dominance × additive variance is

$$V_{D_1A_2} = \sum_{i,j,k} p^i p_j q^k (\delta_j^i \beta^k)^2 + \sum_{i,j,l} p^i p_j q_l (\delta_j^i \beta_l)^2$$

- The dominance × dominance variance:

$$V_{D_1D_2} = \text{Var}\left(\sum_{i,j,k,l} (\delta_j^i \gamma_l^k) x_{M_i}^{(1)} x_{F_j}^{(1)} x_{M_k}^{(2)} x_{F_l}^{(2)}\right)$$

$$= E\left[\left(\sum_{i,j,k,l} (\delta_j^i \gamma_l^k) z_{M_i}^{(1)} z_{F_j}^{(1)} z_{M_k}^{(2)} z_{F_l}^{(2)}\right)^2\right] = \sum_{i,j,k,l} p^i p_j q^k q_l (\delta_j^i \gamma_l^k)^2$$

**C. Partition of genotypic variance in linkage disequilibrium**  
 We present the partition of the genotypic variance based on model (7) under the assumption of Hardy-Weinberg equilibrium but linkage disequilibrium.

- The additive variance:

$$V_{A_1} = \sum_{i=1}^{n_1} p^i (\alpha^i)^2 + \sum_{j=1}^{n_1} p_j (\alpha_j)^2, \quad V_{A_2} = \sum_{k=1}^{n_2} q^k (\beta^k)^2 + \sum_{j=1}^{n_2} q_l (\beta_l)^2$$

- The dominance variance:

$$V_{D_1} = \sum_{i,j} (\delta_j^i)^2 p^i p_j, \quad V_{D_2} = \sum_{k,l} (\gamma_l^k)^2 q^k q_l$$

- The additive × additive variance:

$$V_{A_1A_2} = \sum_{i,k} (p^i q^k + D_{ik}) (\alpha^i \beta^k)^2 - \left[ \sum_{i,k} (\alpha^i \beta^k) D_{ik} \right]^2 + \sum_{j,l} (p_j q_l + D_{jl}) (\alpha_j \beta_l)^2$$

$$- \left[ \sum_{j,l} (\alpha_j \beta_l) D_{jl} \right]^2 + \sum_{i,l} p^i q_l (\alpha^i \beta_l)^2 + \sum_{j,k} p_j q^k (\alpha_j \beta^k)^2 + 2 \sum_{i,j,k,l} (\alpha^i \beta^k) (\alpha_j \beta_l) D_{ij} D_{kl}$$

$$+ 2 \sum_{i,j,k,l} (\alpha^i \beta_l) (\alpha_j \beta^k) D_{ij} D_{kl} + 2 \left( \sum_{i,k} (\alpha^i \beta^k) D_{ik} \right) \left( \sum_{j,l} (\alpha_j \beta_l) D_{jl} \right)$$

- The additive × dominance variance:

$$V_{A_1D_2} = \sum_{i,k,l} q_l (p^i q^k + D_{ik}) (\alpha^i \gamma_l^k)^2 + \sum_{j,k,l} q^k (p_j q_l + D_{jl}) (\alpha_j \gamma_l^k)^2 + 2 \sum_{i,j,k,l} (\alpha^i \gamma_l^k) (\alpha_j \gamma_l^k) D_{ij} D_{kl}$$

- The dominance × additive variance:

$$V_{D_1A_2} = \sum_{i,j,k} p_j (p^i q^k + D_{ik}) (\delta_j^i \beta^k)^2 + \sum_{i,j,l} p^i (p_j q_l + D_{jl}) (\delta_j^i \beta_l)^2 + 2 \sum_{i,j,k,l} (\delta_j^i \beta^k) (\delta_j^i \beta_l) D_{ij} D_{kl}$$

- The dominance × dominance variance:

$$V_{D_1D_2} = \sum_{i,j,k,l} (\delta_j^i \gamma_l^k)^2 (D_{ik} + p^i q^k) (D_{jl} + p_j q_l)$$

- The covariances related to additive and dominance effects:

$$\text{Cov}(A_1, A_2) = \sum_{i,k} (\alpha^i \beta^k) D_{ik} + \sum_{j,l} (\alpha_j \beta_l) D_{jl}$$

$$\text{Cov}(D_1, D_2) = \sum_{i,j,k,l} (\delta_j^i) (\gamma_l^k) D_{ij} D_{kl}$$

$$\text{Cov}(A_1, D_1) = \text{Cov}(A_1, D_2) = \text{Cov}(A_2, D_1) = \text{Cov}(A_2, D_2) = 0$$

- The covariances related to additive × additive effects:

$$\text{Cov}(A_1, A_1A_2) = \sum_{i,k} (\alpha^i) (\alpha^i \beta^k) D_{ik} + \sum_{j,l} (\alpha_j) (\alpha_j \beta_l) D_{jl}$$

$$\text{Cov}(A_2, A_1A_2) = \sum_{i,k} (\beta^k) (\alpha^i \beta^k) D_{ik} + \sum_{j,l} (\beta_l) (\alpha_j \beta_l) D_{jl}$$

$$\text{Cov}(D_1, A_1A_2) = \sum_{i,j,k} p_j (\delta_j^i) (\alpha_j \beta^k) D_{ik} + \sum_{i,j,l} p^i (\delta_j^i) (\alpha^i \beta_l) D_{jl}$$

$$\text{Cov}(D_2, A_1A_2) = \sum_{i,k,l} q_l (\alpha^i \beta_l) (\gamma_l^k) D_{ik} + \sum_{j,k,l} q^k (\alpha_j \beta^k) (\gamma_l^k) D_{jl}$$

- The covariances related to additive × dominance and dominance × additive effects:

$$\text{Cov}(A_1, A_1D_2) = \sum_{i,j,k,l} [( \alpha^i ) (\alpha_j \gamma_l^k) + ( \alpha_j ) (\alpha^i \gamma_l^k) ] D_{ij} D_{kl}$$

$$\text{Cov}(A_2, A_1D_2) = \sum_{i,k,l} q_l (\beta_l) (\alpha^i \gamma_l^k) D_{ik} + \sum_{j,k,l} q^k (\beta^k) (\alpha_j \gamma_l^k) D_{jl}$$

$$\text{Cov}(A_1, D_1A_2) = \sum_{i,j,k} p_j (\alpha_j) (\delta_j^i \beta^k) D_{ik} + \sum_{i,j,l} p^i (\alpha^i) (\delta_j^i \beta_l) D_{jl}$$

$$\text{Cov}(A_2, D_1A_2) = \sum_{i,j,k,l} [( \beta^k ) (\delta_j^i \beta_l) + ( \beta_l ) (\delta_j^i \beta^k) ] D_{ij} D_{kl}$$

$$\text{Cov}(D_1, A_1D_2) = \sum_{i,j,k,l} (\delta_j^i) [( \alpha^i \gamma_l^k ) + ( \alpha_j \gamma_l^k ) ] D_{ij} D_{kl}$$

$$\text{Cov}(D_2, A_1D_2) = \sum_{i,k,l} q_l (\gamma_l^k) (\alpha^i \gamma_l^k) D_{ik} + \sum_{j,k,l} q^k (\gamma_l^k) (\alpha_j \gamma_l^k) D_{jl}$$

$$\text{Cov}(D_1, D_1A_2) = \sum_{i,j,k} p_j (\delta_j^i) (\delta_j^i \beta^k) D_{ik} + \sum_{i,j,l} p^i (\delta_j^i) (\delta_j^i \beta_l) D_{jl}$$

$$\text{Cov}(D_2, D_1A_2) = \sum_{i,j,k,l} (\gamma_l^k) [ (\delta_j^i \beta^k) + (\delta_j^i \beta_l) ] D_{ij} D_{kl}$$

$$\text{Cov}(A_1A_2, A_1D_2) = \sum_{i,k,l} q_l (\alpha^i \beta_l) (\alpha^i \gamma_l^k) D_{ik} + \sum_{j,k,l} q^k (\alpha_j \beta^k) (\alpha_j \gamma_l^k) D_{jl}$$

$$+ \sum_{i,j,k,l} [ (\alpha^i \beta^k) (\alpha_j \gamma_l^k) + (\alpha^i \beta_l) (\alpha_j \gamma_l^k) + (\alpha_j \beta_l) (\alpha^i \gamma_l^k) + (\alpha_j \beta^k) (\alpha^i \gamma_l^k) ] D_{ij} D_{kl}$$

$$\text{Cov}(A_1, A_2, D_1A_2) = \sum_{i,j,k} p_j (\alpha_j \beta^k) (\delta_j^i \beta^k) D_{ik} + \sum_{i,j,l} p^i (\alpha^i \beta_l) (\delta_j^i \beta_l) D_{jl}$$

$$+ \sum_{i,j,k,l} [ (\alpha^i \beta^k) (\delta_j^i \beta_l) + (\alpha_j \beta^k) (\delta_j^i \beta_l) + (\alpha^i \beta_l) (\delta_j^i \beta^k) + (\alpha_j \beta_l) (\delta_j^i \beta^k) ] D_{ij} D_{kl}$$

$$\text{Cov}(A_1D_2, D_1A_2) = \sum_{i,j,k,l} [ (\alpha^i \gamma_l^k) (\delta_j^i \beta^k) (D_{ik} + p^i q^k) D_{jl} + (\alpha^i \gamma_l^k) (\delta_j^i \beta_l) D_{ij} D_{kl}$$

$$+ (\alpha_j \gamma_l^k) (\delta_j^i \beta^k) D_{ij} D_{kl} + (\alpha_j \gamma_l^k) (\delta_j^i \beta_l) (D_{ij} + p_j q_l) D_{kl} ]$$

- The covariances related to dominance × dominance effects:

$$\text{Cov}(A_1, D_1D_2) = \sum_{i,j,k,l} (\delta_j^i \gamma_l^k) (\alpha^i + \alpha_j) D_{ij} D_{kl}$$

$$\text{Cov}(A_2, D_1D_2) = \sum_{i,j,k,l} (\delta_j^i \gamma_l^k) (\beta^k + \beta_l) D_{ij} D_{kl}$$

$$\text{Cov}(D_1, D_1D_2) = \sum_{i,j,k,l} (\delta_j^i \gamma_l^k) (\delta_j^i) D_{ij} D_{kl}$$

$$\text{Cov}(D_2, D_1D_2) = \sum_{i,j,k,l} (\delta_j^i \gamma_l^k) (\gamma_l^k) D_{ij} D_{kl}$$

$$\text{Cov}(A_1A_2, D_1D_2) = - \left[ \sum_{i,k} (\alpha^i \beta^k) D_{ik} + \sum_{j,l} (\alpha_j \beta_l) D_{jl} \right] \left[ \sum_{i,j,k,l} (\delta_j^i \gamma_l^k) D_{ij} D_{kl} \right] + \sum_{i,j,k,l} (\delta_j^i \gamma_l^k) \{ (\alpha^i \beta^k) + (\alpha^i \beta_l) + (\alpha_j \beta^k) + (\alpha_j \beta_l) \} D_{ij} D_{kl} + (\alpha^i \beta^k) p^i q^k D_{ij} + (\alpha_j \beta_l) p_j q_l D_{ij}$$

$$\text{Cov}(A_1D_2, D_1D_2) = \sum_{i,j,k,l} (\delta_j^i \gamma_l^k) [ (\alpha^i \gamma_l^k) (D_{ik} + p^i q^k) D_{jl} + (\alpha_j \gamma_l^k) (D_{ij} + p_j q_l) D_{kl} ]$$

$$\text{Cov}(D_1A_2, D_1D_2) = \sum_{i,j,k,l} (\delta_j^i \gamma_l^k) [ (\delta_j^i \beta^k) (D_{ik} + p^i q^k) D_{jl} + (\delta_j^i \beta_l) (D_{ij} + p_j q_l) D_{kl} ]$$

**D. Partition of genotypic variance in F<sub>2</sub> population with linkage disequilibrium**

Detail of each component of equation (19) for F<sub>2</sub> population is presented here.

- The additive variance:

$$V_A = \frac{1}{2} \sum_{r=1}^m a_r^2 + 2 \sum_{r \neq s} a_r a_s D_{rs}$$

- The dominance variance:

$$V_D = \frac{1}{4} \sum_{r=1}^m d_r^2 + 4 \sum_{r \neq s} d_r d_s D_{rs}^2$$

- The additive × additive variance:

$$V_{AA} = \frac{1}{4} \sum_{r < s} (aa)_{rs}^2 + \sum_{r \neq s \neq s'} (aa)_{rs} (aa)_{rs'} D_{ss'} \\ + \frac{1}{2} \sum_{r \neq s \neq r' \neq s'} (aa)_{rs} (aa)_{r's'} (D_{rr'ss'} + D_{rs'D_{r's}} + D_{rr'D_{ss'}} - D_{rs} D_{r's'})$$

- The additive × dominance variance:

$$V_{AD} = \frac{1}{8} \sum_{r \neq s} (ad)_{rs}^2 + \frac{1}{2} \sum_{r \neq s} (ad)_{rs} (ad)_{sr} D_{rs} + 2 \sum_{r \neq s \neq s'} (ad)_{rs} [(ad)_{rs'} D_{ss'}^2 + (ad)_{ss'} D_{ss'} D_{rs'} \\ + (ad)_{s'r} D_{rs} D_{ss'} + \frac{1}{4} (ad)_{s's} D_{rs'}^2] + 8 \sum_{r \neq s \neq r' \neq s'} (ad)_{rs} (ad)_{r's'} D_{rsr's'} D_{ss'}$$

- The dominance × dominance variance:

$$V_{DD} = \sum_{r < s} (dd)_{rs}^2 \left( \frac{1}{16} - 16 D_{rs}^4 \right) + \sum_{r \neq s \neq s'} (dd)_{rs} (dd)_{rs'} (D_{ss'}^2 - 16 D_{rs}^2 D_{rs'}^2) \\ + 4 \sum_{r \neq s \neq r' \neq s'} (dd)_{rs} (dd)_{r's'} (D_{rsr's'}^2 - D_{rs}^2 D_{r's'}^2)$$

- The covariances: Cov(A, D) = Cov(A, AA) = Cov(A, DD) = Cov(D, AD) = Cov(D, DD) = Cov(AA, AD) = Cov(AD, DD) = 0, and

$$\text{Cov}(A, AD) = - \sum_{r \neq s} a_r D_{rs} [4(ad)_{rs} D_{rs} + (ad)_{sr}] - 4 \sum_{r \neq s \neq r'} a_r (ad)_{rs} D_{rs} D_{r's'}$$

$$\text{Cov}(D, AA) = - \sum_{r < s} (d_r + d_s) (aa)_{rs} D_{rs} - 4 \sum_{r \neq s \neq r'} d_r (aa)_{rs} D_{rr'} D_{r's'}$$

$$\text{Cov}(AA, DD) = \sum_{r < s} (aa)_{rs} (dd)_{rs} \left( \frac{1}{2} D_{rs} - 8 D_{rs}^3 \right) + \sum_{r \neq s \neq s'} (aa)_{rs} (dd)_{rs'} (2 D_{rs'} D_{ss'} - 8 D_{rs} D_{rs'}^2) \\ + 2 \sum_{r \neq s \neq r' \neq s'} (aa)_{rs} (dd)_{r's'} (D_{rsr's'} D_{r's'} - D_{rs} D_{r's'}^2)$$

where  $(aa)_{sr} = (aa)_{rs}$  and  $(dd)_{sr} = (dd)_{rs}$  for  $r < s$ .

In this presentation, we utilized the assumption of no crossing-over interference which results in the third order linkage disequilibrium of three loci being zero, *i.e.*  $D_{rss'} = 0$ . It may be instructive to show this result.

$$D_{rst} = E(x_r x_s x_t) = E((z_r - p_r)(z_s - p_s)(z_t - p_t)) \\ = E(z_r z_s z_t - p_r z_s z_t - z_r p_s z_t - z_r z_s p_t + p_r p_s z_t + p_r z_s p_t + z_r p_s p_t - p_r p_s p_t) \\ = E(z_r z_s z_t - \frac{1}{2} z_s z_t - \frac{1}{2} z_r z_t - \frac{1}{2} z_r z_s + \frac{1}{4})$$

Let  $r_1$  be the recombination frequency between loci  $r$  and  $s$ ,  $r_2$  be that between  $s$  and  $t$ , and  $r_{12}$  be that between  $r$  and  $t$ . Under the assumption of no crossing-over interference, for loci  $r, s, t$  in this order, we have  $E(z_r z_s z_t) = \frac{1}{2} (1 - r_1)(1 - r_2)$ ,  $E(z_r z_s) = \frac{1}{2} (1 - r_1)$ ,  $E(z_s z_t) = \frac{1}{2} (1 - r_2)$ , and  $E(z_r z_t) = \frac{1}{2} (1 - r_{12})$ . Since the assumption of no crossing-over interference implies  $r_{12} = r_1 + r_2 - 2r_1 r_2$ , thus  $D_{rst} = 0$ .

**E. Partition of genotypic variance for the general two-allele model with linkage disequilibrium**

Detail of each component of equation (19) for the general two-allele model is presented here.

- The additive variance:

$$V_A = 2 \sum_{r=1}^m p_r (1 - p_r) a_r^2 + 2 \sum_{r \neq s} a_r a_s D_{rs}$$

- The dominance variance:

$$V_D = \sum_{r=1}^m p_r^2 (1 - p_r)^2 d_r^2 + \sum_{r \neq s} d_r d_s D_{rs}^2$$

- The additive × additive variance:

$$V_{AA} = 2 \sum_{r < s} (aa)_{rs}^2 [(1 - 2p_r)(1 - 2p_s) D_{rs} + 2p_r (1 - p_r) p_s (1 - p_s)] \\ + 2 \sum_{r \neq s \neq s'} (aa)_{rs} (aa)_{rs'} [(1 - 2p_r) D_{rss'} + 2p_r (1 - p_r) D_{ss'}] \\ + \frac{1}{2} \sum_{r \neq s \neq r' \neq s'} (aa)_{rs} (aa)_{r's'} (D_{rr'ss'} + D_{rs'D_{r's}} + D_{rr'D_{ss'}} - D_{rs} D_{r's'})$$

- The additive × dominance variance:

$$V_{AD} = 2 \sum_{r \neq s} (ad)_{rs}^2 [(1 - 2p_s)^2 D_{rs}^2 + (1 - 2p_r)(1 - 2p_s) p_s (1 - p_s) D_{rs} + p_r (1 - p_r) p_s^2 (1 - p_s)^2] \\ + 2 \sum_{r \neq s} (ad)_{rs} (ad)_{sr} D_{rs} [2(1 - 2p_r)(1 - 2p_s) D_{rs} + p_r (1 - p_r) p_s (1 - p_s)] \\ + 2 \sum_{r \neq s \neq s'} (ad)_{rs} [(ad)_{rs'} [D_{rs'}^2 + (1 - 2p_r) D_{rss'} D_{ss'} + p_r (1 - p_r) D_{ss'}^2] \\ + (ad)_{ss'} [2(1 - 2p_s) D_{rs'} D_{ss'} + p_s (1 - p_s) D_{ss'} D_{rs'}] \\ + (ad)_{s'r} [2(1 - 2p_r) D_{rss'} D_{rs} + p_r (1 - p_r) D_{rs} D_{ss'}] \\ + (ad)_{s's} [(1 - 2p_s)^2 D_{rs} D_{ss'} + (1 - 2p_s) p_s (1 - p_s) D_{rss'} + p_s^2 (1 - p_s)^2 D_{rs}]] \\ + 2 \sum_{r \neq s \neq r' \neq s'} (ad)_{rs} (ad)_{r's'} (D_{rsr's'} D_{ss'} + D_{rss'} D_{sr's'})$$

- The dominance × dominance variance:

$$V_{DD} = \sum_{r < s} (dd)_{rs}^2 \left\{ [(1-2p_r)(1-2p_s)D_{rs} + p_r(1-p_r)p_s(1-p_s)]^2 - D_{rs}^4 \right\} + \sum_{r \neq s \neq s'} (dd)_{rs} (dd)_{rs'} \left\{ [(1-2p_r)D_{rss'} + p_r(1-p_r)D_{ss'}]^2 - D_{rs}^2 D_{rs'}^2 \right\} + \frac{1}{4} \sum_{r \neq s \neq r' \neq s'} (dd)_{rs} (dd)_{r's'} (D_{rss'r'}^2 - D_{rs}^2 D_{r's'}^2)$$

- The covariances:

$$\begin{aligned} \text{Cov}(A, D) &= 0 \\ \text{Cov}(A, AA) &= 2 \sum_{r < s} (aa)_{rs} D_{rs} [(1-2p_r)a_r + (1-2p_s)a_s] + \sum_{r \neq s \neq r'} a_r (aa)_{rs} D_{r's} \\ \text{Cov}(A, AD) &= 2 \sum_{r \neq s} a_r D_{rs} [(ad)_{rs} D_{rs} + p_r(1-p_r)(ad)_{sr}] + 2 \sum_{r \neq s \neq r'} a_r (ad)_{rs} D_{rs} D_{r's} \\ \text{Cov}(A, DD) &= 2 \sum_{r < s} (dd)_{rs} D_{rs}^2 [(1-2p_r)a_r + (1-2p_s)a_s] + 2 \sum_{r \neq s \neq r'} a_r (dd)_{rs} D_{rss'} D_{rs} \\ \text{Cov}(D, AA) &= 2 \sum_{r < s} (aa)_{rs} D_{rs} [p_r(1-p_r)d_r + p_s(1-p_s)d_s] + 2 \sum_{r \neq s \neq r'} d_r (aa)_{rs} D_{rr'} D_{r's} \\ \text{Cov}(D, AD) &= 2 \sum_{r \neq s} d_r D_{rs} (1-2p_r) [D_{rs}(ad)_{rs} + p_r(1-p_r)(ad)_{sr}] + 2 \sum_{r \neq s \neq r'} d_r (ad)_{rs} D_{rs} D_{r's} D_{r's} \\ \text{Cov}(D, DD) &= \sum_{r < s} (dd)_{rs} D_{rs}^2 [(1-2p_r)^2 d_r + (1-2p_s)^2 d_s] + 2 \sum_{r \neq s \neq r'} d_r (dd)_{rs} D_{rs}^2 D_{rs} \\ \text{Cov}(AA, AD) &= 2 \sum_{r < s} (aa)_{rs} D_{rs} \{ (ad)_{rs} [(1-2p_r)p_r(1-p_s) + 2(1-2p_s)D_{rs}] \\ &\quad + (ad)_{sr} [(1-2p_s)p_r(1-p_r) + 2(1-2p_r)D_{rs}] \} \\ &\quad + 2 \sum_{r \neq s \neq s'} (aa)_{rs} \{ (ad)_{ss'} [(1-2p_r)D_{ss'} D_{rs'} + 2D_{rss'} D_{ss'}] \\ &\quad + (ad)_{s's} [2(1-2p_s)D_{rs} D_{ss'} + p_s(1-p_s)D_{rss'}] \} \\ &\quad + \sum_{r \neq s \neq r' \neq s'} (aa)_{rs} (ad)_{r's'} (D_{rss'} D_{r's} + D_{rr'} D_{ss'} + D_{r'ss'} D_{rs'}) \\ \text{Cov}(AA, DD) &= 2 \sum_{r < s} (aa)_{rs} (dd)_{rs} [2(1-2p_r)(1-2p_s)D_{rs}^2 + p_r(1-p_r)p_s(1-p_s)D_{rs} - D_{rs}^3] \\ &\quad + 2 \sum_{r \neq s \neq s'} (aa)_{rs} (dd)_{rs'} [2(1-2p_r)D_{rss'} D_{rs'} + p_r(1-p_r)D_{rs'} D_{ss'} - D_{rs} D_{rs'}^2] \\ &\quad + \frac{1}{2} \sum_{r \neq s \neq r' \neq s'} (aa)_{rs} (dd)_{r's'} (D_{rss'} D_{r's} + D_{rr'} D_{ss'} - D_{rs} D_{r's}^2) \\ \text{Cov}(AD, DD) &= 2 \sum_{r \neq s} (ad)_{rs} (dd)_{rs} D_{rs} (1-2p_s) [(1-2p_r)(1-2p_s)D_{rs} + p_r(1-p_r)p_s(1-p_s)] \\ &\quad + 2 \sum_{r \neq s \neq s'} (ad)_{rs} \{ (dd)_{rs'} D_{rss'} [(1-2p_r)D_{rss'} + p_r(1-p_r)D_{ss'}] \\ &\quad + (dd)_{ss'} D_{ss'} (1-2p_s) [(1-2p_s)D_{rss'} + p_s(1-p_s)D_{rs'}] \} \\ &\quad + \sum_{r \neq s \neq r' \neq s'} (ad)_{rs} (dd)_{r's'} D_{rss'} D_{r's} \end{aligned}$$

where  $(aa)_{sr} = (aa)_{rs}$  and  $(dd)_{sr} = (dd)_{rs}$  for  $r < s$ .

The result in Appendix D for the  $F_2$  model with  $p_r = 1/2$  for  $r = 1, \dots, m$  and also assuming  $D_{rss'} = 0$  is a special case of the results presented here. There is a difference, by a factor -2, on the specification of  $v$  variable for dominance effect for the  $F_2$  model and the general two-allele model, which carries over to the comparison of results in Appendix D and E.

### Acknowledgements

We are grateful to Bill Hill and Bruce Weir for very helpful discussions and comments and to Wei Zou, Sheng Feng, Sheng-Mao Cheng and Jessica Maia for checking some results. This work was partially supported by NIH GM45344 and by the National Research Initiative of the USDA Cooperative State Research, Education and Extension Service, grant number #2005-00754.

### References

1. Yule GU: **Mendel's laws and their probable relations to intra-racial heredity.** *New Phytologist* 1902, **1**:193-207. 222-238
2. Yule GU: **On the theory of inheritance of quantitative compound characters on the basis of Mendel's laws – A preliminary note.** In *Int Conf Genet 3rd, Rep* Edited by: Wilks W. Royal Horticultural Society, London; 1906:140-142.
3. Pearson K: **Mathematical contributions to the theory of evolution XII. On a generalized theory of alternative inheritance, with special reference to Mendel's laws.** *R Soc Philos Trans A* 1904, **203**:53-86.
4. Provine WB: *The Origins of Theoretical Population Genetics* Chicago: University of Chicago Press; 1971.
5. Hill WG: **Quantitative Genetics, Part I: Explanation and Analysis of Continuous Variation.** In *Benchmark Papers in Genetics Volume 15.* New York: Van Nostrand Reinhold Company; 1984.
6. Fisher RA: **The correlation between relatives on the supposition of Mendelian inheritance.** *Trans Roy Soc* 1918, **52**:399-433.
7. Cockerham C: **An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present.** *Genetics* 1954, **39**:859-882 [[http://www.genetics.org/cgi/reprint/3859?maxow=&HITS=10&hits=10&RESULTFORMAT=&author\\_larchid=1140647473818\\_2969&FIRSTIN\\_DEX=30&journalde=genetics](http://www.genetics.org/cgi/reprint/3859?maxow=&HITS=10&hits=10&RESULTFORMAT=&author_larchid=1140647473818_2969&FIRSTIN_DEX=30&journalde=genetics)].
8. Kempthorne O: **The correlation between relatives in a random mating population.** *Proc Royal Soc B* 1954, **143**:103-113.
9. Kempthorne O: *An Introduction to Genetic Statistics* New York: John Wiley & Sons; 1957.
10. Zeng ZB, Wang T, Zou W: **Modeling quantitative trait loci and interpretation of models.** *Genetics* 2005, **169**:1711-1725.
11. Cheverud JM, Routman EJ: **Epistasis and its contribution to genetic variance components.** *Genetics* 1995, **139**:1455-1461.
12. Cheverud JM: **Detecting epistasis among quantitative trait loci.** In *Epistasis and the Evolutionary Process* Edited by: Wolf JB, Brodie EDI, Wade MJ. Oxford Univ. Press, Oxford, UK; 2000:58-81.
13. Cockerham C: **Effects of linkage on the covariances between relatives.** *Genetics* 1956, **41**:138-141 [[http://www.genetics.org/cgi/reprint/48?maxtoshow=&HITS=10&hits=10&RESULTFORMAT=&a=&searchid=1140647473818\\_2969&FIRSTIN\\_DEX=30&journalcode=genetics](http://www.genetics.org/cgi/reprint/48?maxtoshow=&HITS=10&hits=10&RESULTFORMAT=&a=&searchid=1140647473818_2969&FIRSTIN_DEX=30&journalcode=genetics)].
14. Weir BS: *Genetic Data Analysis II* Sunderland, Massachusetts: Sinauer Associates, Inc; 1996.
15. Slatkin M: **On treating the chromosome as the unit of selection.** *Genetics* 1972, **72**:157-168.
16. Weir BS, Cockerham CC: **Mixed self and random mating at two loci.** *Genet Res* 1973, **21**:247-262.
17. Weir BS, Cockerham CC: **Two-locus theory in quantitative genetics.** In *Proceedings of the international conference on quantitative genetics* Edited by: Pollack E, Kempthorne O, Bailey TB. Iowa State University Press; 1977:247-269.
18. Kao CH, Zeng ZB: **Modeling epistasis of quantitative trait loci using Cockerham's model.** *Genetics* 2002, **160**:1243-1261.
19. Kao CH, Zeng ZB, Teasdale RD: **Multiple interval mapping for quantitative trait loci.** *Genetics* 1999, **152**:1203-1216.
20. Zeng ZB, Liu J, Stam LF, Kao CH, Mercer JM, Laurie CC: **Genetic architecture of a morphological shape difference between two Drosophila species.** *Genetics* 2000, **154**:299-310.
21. Zeng ZB, Kao CH, Basten CJ: **Estimating the genetic architecture of quantitative traits.** *Genet Res* 1999, **74**:279-289.
22. Weber K, Eisman R, Higgins S, Kulk L, Patty A, Sparks J, Zeng ZB: **An analysis of polygenes affecting wing shape on chromosome three in Drosophila melanogaster.** *Genetics* 1999, **153**:773-786.
23. Weber K, Eisman R, Higgins S, Morey L, Patty A, Tausek M, Zeng ZB: **An analysis of polygenes affecting wing shape on chromosome 2 in Drosophila melanogaster.** *Genetics* 2001, **159**:1045-1057.
24. Lander ES, Botstein D: **Mapping Mendelian Factors Underlying Quantitative Traits Using RFLP Linkage Maps.** *Genetics* 1989, **121**:185-199.
25. Zeng ZB: **Precision mapping of quantitative trait loci.** *Genetics* 1994, **136**:1457-1468.