BIOINFORMATICS ORIGINAL PAPER

Vol. 21 no. 10 2005, pages 2447–2455
doi:10.1093/bioinformatics/bti342

Genetics and population analysis

Mapping genome–genome epistasis: a high-dimensional model

Yuehua Cui and Rongling Wu*

Department of Statistics, University of Florida, Gainesville, FL 32611, USA

Received on December 20, 2004; revised on January 25, 2005; accepted on February 17, 2005
Advance Access publication February 22, 2005

ABSTRACT

Motivation: The proper development of any organ or tissue requires the coordinated expression of its underlying genes that can be located on different genomes present in an organism. For instance, each step in the development of seed for a higher plant is the consequence of gene interactions from the maternal, embryo and endosperm genomes.

Results: We present a multivariate statistical model for mapping quantitative trait loci (QTL) by incorporating two important aspects of seed development in plants—QTL interactions derived from different genomes, the maternal, embryo and endosperm, and genetic correlations among phenotypic traits expressed in different genome-specific tissues. This model, which has a high dimensionality, is constructed within the maximum-likelihood context based on a finite mixture model. The implementation of the expectation–maximization algorithm allows for the efficient estimation of QTL positions, their action and interaction effects and pleiotropic effects. The application of this high-dimensional model to a real rice dataset has validated its usefulness.

Conclusions: Our model was derived for self-pollinated plants, but it can be extended to cross-pollinated plants and to animals. With the burgeoning of genetic and genomic data, this high-dimensional model will have many implications for agricultural and evolutionary genetic research.

Availability: A package of software will be provided from the corresponding author upon request.

Contact: rwu@stat.ufl.edu

INTRODUCTION

Studies of genome-wide scans for quantitative trait loci (QTL) that determine phenotypic traits have received considerable attention in the past 15 years (Lander and Botstein, 1989; Zeng, 1994; Wu *et al*., 2002a). The aim of these studies was to understand the genetic architecture of quantitative variation for complex traits of agricultural, evolutionary and biomedical interest (reviewed in Mackay, 2001). The genetic principle behind these studies is the occurrence of recombination events between genetic loci when gametes are formed and transmitted from parents to offspring. Although statistical methods for QTL mapping were proposed originally on the basis of a bivariate approach that associates one gene with one trait, considerable attempts have been made to develop multivariate approaches for mapping multiple interacting QTL (reviewed in Carlborg and Haley, 2004) and multiple correlated phenotypic traits (Jiang and Zeng, 1995; Korol *et al*., 1995; Knott and Haley, 2000). In this work, we address problems associated with intergenomic epistasis and multiple correlated traits that have not previously been addressed.

First, the genetic interaction between different genes or epistasis provides important fuel for creating novel quantitative genetic variation when an organism is forced to adapt to a new environment (Whitlock *et al*., 1995). The conventional concept of epistasis implies the effect of an allele at one gene affected by another allele at another gene on the same genome or individual (Falconer and Mackay, 1996). However, there is also another type of epistasis that occurs between different genes each from a different genome or individual (Wolf, 2000; Wolf *et al*., 1998). Such genome–genome or individual–individual epistasis has been believed to be an important force in maintaining genetic variation in fluctuating environments (Wolf *et al*., 2002) and to help select optimal life history strategies (Wolf, 2003). An excellent example of genome–genome epistasis is the coordinated regulation among the maternal, embryo and endosperm tissues in a developing seed (Walbot and Evans, 2003). The genetic mapping of genome–genome epistasis based on molecular markers is in its infant stage. Cui *et al*. (2004) recently published a series of statistical models for detecting epistatic effects on embryo- or endosperm-specific traits between different QTL derived from the maternal, embryo and endosperm genomes in seed plants. These models take into account the genetic and developmental mechanisms for seed development and can be of greater significance in the study of genetic control of seed traits aimed at improving grain production in crops with the aid of molecular biotechnologies.

Second, correlations between different biological traits are ubiquitous, with the pattern and degree of trait correlations thought to be the consequence of natural selection and evolution (Scheiner, 1993). Traditional correlation analysis deals with different traits from the same individual. But it is common for two different traits each from a different individual to be correlated. For example, maternal preferences for oviposition sites affect the survival rate and development of offspring in birds (Lloyd and Martin, 2004). In plants, the level of hormones released by endosperm is thought to guide embryo development (Chaudhury *et al*., 2001). Genetic mapping approaches for multiple traits capitalize on the information about interrelationships among different traits measured and, therefore, can affect the statistical power of QTL detection. Although a joint analysis of many traits does not necessarily lead to a higher power of detection due to an increased number of parameters being estimated, it has been shown that the statistical power to detect a QTL can be increased by including a few correlated traits. Such an increase in power has been demonstrated using regression methods (Knott and Haley, 2000),

*To whom correspondence should be addressed.

a maximum-likelihood method (Korol *et al.*, 1995; Jiang and Zeng, 1995), and variance component models (Almasy *et al.*, 1997). It is particularly favorable to utilize the correlated information when mapping QTL for low heritability traits that are correlated to a trait of higher heritability. Lund *et al.* (2003) documented several advantages of multitrait QTL mapping over a single trait analysis.

With the burgeoning recognition of the importance of genome–genome epistasis and genetic correlations between individual-specific traits, it is appealing to develop a multivariate statistical model for mapping QTL interactions that affect multiple correlated traits expressed on different individuals or genomes. This motivation stimulates us to develop a high-dimensional model for estimating and testing the gene action and interaction effects on individual-specific traits between the QTL from different genomes. This high-dimensional model was derived from a mixture-based likelihood model and implemented with the expectation(E)–maximization(M) algorithm (Dempster *et al.*, 1977) for Monte Carlo simulations under different sampling strategies to investigate the statistical behavior of our multivariate model. The successful detection of interactive QTL in an example, for rice validates the usefulness of this model.

EXPERIMENTAL DESIGN

Our model will be developed for a simple backcross, but can be extended to an F_2 or other designs. Consider two homozygous inbred lines which are crossed to generate the heterozygous F_1. Crossing the F_1 to one of the two parents (say the homozygous recessive) leads to two different genotypes at each locus in the backcross. The progeny of the backcross can be obtained through self-pollination for autogamous species, such as rice and soybean or through outcrossing pollination for allogamous species, such as maize and animals.

The backcross is genotyped for a set of molecular markers to construct a genetic linkage map. As shown in Wu *et al.* (2002a), genotyping the diploid progeny of the backcross with the same set of markers can increase the power to map the QTL that are expressed in the progeny generation, such as the embryo and endosperm of the seed. Here, we suppose that the markers from both the backcross and its diploid progeny are available to characterize interactions between multiple QTL from different genomes. For animals, a genome–genome interaction may occur as a maternal–offspring interaction. For plants, the progeny (seeds) develop within the maternal sporophyte tissue after double fertilization of the gametophyte; hence there are potentially extensive genome–genome interactions. Double fertilization forms the diploid embryo by fusing the haploid egg with one of the sperm cells and the triploid endosperm by fusing the maternal homodiploid central cell with a second sperm cell (Chaudhury *et al.*, 2001). Proper seed development requires the coordinated expression of the maternal, embryo and endosperm tissues (van Hengel *et al.*, 1998; Opsahl-Ferstad *et al.*, 1997). There has been a wealth of evidence for the genetic control of different genes from these three genomes over seed development (Chaudhury *et al.*, 2001; Evans and Kermicle, 2001; Dilkes *et al.*, 2002; Walbot and Evans, 2003). Therefore, for plants, the genome–genome interaction should include three possible types, maternal–embryo, maternal–endosperm and embryo–endosperm. For this reason, the models to characterize genome–genome interactions developed for plant systems will also cover those for animal systems.

If the backcross is assumed to be at generation t, then its progeny obtained through outcrossing pollination is viewed as generation $t+1$. Let $\mathbf{P}_t$, $\mathbf{Q}_{t+1}$ and $\mathbf{Q}'_{t+1}$ be three different QTL from the maternal (generation t), embryo (generation $t+1$) and endosperm (generation $t+1$) genomes, respectively. In generation t, there are two QTL genotypes at $\mathbf{P}_t$, expressed as P_tp_t and p_tp_t, whereas, in generation $t+1$, there are three QTL genotypes at $\mathbf{Q}_{t+1}$ in the embryo, expressed as $Q_{t+1}Q_{t+1}$, $Q_{t+1}q_{t+1}$ and $q_{t+1}q_{t+1}$, and four QTL genotypes at $\mathbf{Q}'_{t+1}$ in the endosperm, expressed as $Q'_{t+1}Q'_{t+1}Q'_{t+1}$, $Q'_{t+1}Q'_{t+1}q'_{t+1}$, $Q'_{t+1}q'_{t+1}q'_{t+1}$ and $q'_{t+1}q'_{t+1}q'_{t+1}$.

The maternal–embryo interaction model

The two QTL from the maternal ($\mathbf{P}_t$) and embryo genomes ($\mathbf{Q}_{t+1}$) form six across-generation QTL genotypes. Their genotypic values for a quantitative trait, denoted by $\mu_{j_tj_{t+1}}$, where j_tj_{t+1} stands for the genome-specific QTL genotypes in terms of different numbers of capital QTL alleles, are assigned as follows:

$$\begin{array}{cc} & \begin{array}{ccc} \qquad Q_{t+1}Q_{t+1} \qquad & \qquad Q_{t+1}q_{t+1} \qquad & \qquad q_{t+1}q_{t+1} \qquad \end{array} \\ \begin{array}{c} P_tp_t \\ \\ p_tp_t \end{array} & \begin{bmatrix} \mu_{12}=\mu+\frac{1}{2}a_t+a_{t+1}+\frac{1}{2}I & \mu_{11}=\mu+\frac{1}{2}a_t+d_{t+1}+\frac{1}{2}J & \mu_{10}=\mu+\frac{1}{2}a_t-a_{t+1}-\frac{1}{2}I \\ \\ \mu_{02}=\mu-\frac{1}{2}a_t+a_{t+1}-\frac{1}{2}I & \mu_{01}=\mu-\frac{1}{2}a_t+d_{t+1}-\frac{1}{2}J & \mu_{00}=\mu-\frac{1}{2}a_t-a_{t+1}+\frac{1}{2}I \end{bmatrix} \end{array}, \qquad (1)$$

where μ is the overall mean, a_t and a_{t+1} are the additive effects of the maternal $\mathbf{P}_t$ and embryo $\mathbf{Q}_{t+1}$, respectively, d_{t+1} is the dominant effect of embryo $\mathbf{Q}_{t+1}$, and I and J are the across-generation maternal-additive × embryo-additive and maternal-additive × embryo-dominant effects between the two QTL, respectively.

We treat the genetic map location of the QTL as missing data, to be inferred from known markers by the EM algorithm. The marker information provided differently by the backcross and its offspring will be combined for our mapping model. Assume that the maternal $\mathbf{P}_t$ is bracketed by two flanking markers, $\mathbf{M}_t^1$ and $\mathbf{M}_t^2$, genotyped from the backcross, and the offspring $\mathbf{Q}_{t+1}$ is bracketed by two flanking markers, $\mathbf{N}_{t+1}^1$ and $\mathbf{N}_{t+1}^2$, genotyped from the offspring. Let r, r_1 and r_2 be the recombination fractions between the two maternal markers, marker $\mathbf{M}_t^1$ and maternal $\mathbf{Q}_t$, and maternal $\mathbf{Q}_t$ and marker $\mathbf{M}_t^2$, respectively. The corresponding recombination fractions are denoted as s, s_1 and s_2 for the offspring markers and QTL. The conditional probabilities of maternal QTL genotypes given maternal marker interval, $\mathbf{M}_t^1-\mathbf{M}_t^2$, in the backcross can be expressed in terms of r, r_1 and r_2. Depending on the pollination type, we can also derive the conditional probabilities of embryo QTL genotypes in terms of s, s_1 and s_2, given the across-generation marker interval $(\mathbf{N}_t^1-\mathbf{N}_t^2)/(\mathbf{N}_{t+1}^1-\mathbf{N}_{t+1}^2)$. Wu *et al.* (2002) and Cui *et al.* (2004) provided such conditional probabilities for self-pollinated plants. Similar procedures can be used to derive these conditional probabilities for cross-pollinated plants.

If different from the offspring interval $\mathbf{M}_t^1$ – $\mathbf{M}_t^2$ is different from the marker interval $\mathbf{N}_{t+1}^1$ – $\mathbf{N}_{t+1}^2$, the conditional probabilities of across-generation QTL genotypes given across-generation marker genotypes can be calculated as the product of QTL-specific conditional probabilities. If these two markers are the same, i.e. the maternal and offspring QTL are located on the same interval, then the conditional probabilities of across-generation QTL genotypes should be derived independently (Cui *et al.*, 2004). These conditional probabilities will be used for the test and estimation of the positions of the two interacting QTL.

The maternal–endosperm interaction model

Across-generation QTL genotypes for the maternal ($\mathbf{P}_t$) and endosperm ($\mathbf{Q}'_{t+1}$) genomes include eight combinations between two maternal genotypes and four endosperm genotypes. The genotypic values of the maternal–endosperm QTL genotypes, $\mu_{j_t j_{t+1}}$, can be assigned as follows:

$$\begin{array}{cc} & \begin{array}{llll} \quad Q'_{t+1}Q'_{t+1}Q'_{t+1} & \quad Q'_{t+1}Q'_{t+1}q'_{t+1} & \quad Q'_{t+1}q'_{t+1}q'_{t+1} & \quad q'_{t+1}q'_{t+1}q'_{t+1} \end{array} \\ \begin{array}{c} P_t p_t \\ \\ \\ \\ \\ p_t p_t \\ \\ \\ \end{array} & \left[\begin{array}{llll} \mu_{13}=\mu & \mu_{12}=\mu & \mu_{10}=\mu & \mu_{11}=\mu \\ +\frac{1}{2}a_t+\frac{3}{2}a'_{t+1} & +\frac{1}{2}a_t+\frac{1}{2}a'_{t+1}+d'_{(t+1)1} & +\frac{1}{2}a_t-\frac{1}{2}a'_{t+1}+d'_{(t+1)2} & +\frac{1}{2}a_t-\frac{3}{2}a'_{t+1} \\ -\frac{3}{4}I' & -\frac{1}{4}I'-\frac{1}{2}J'_1 & +\frac{1}{4}I'-\frac{1}{2}J'_2 & +\frac{3}{4}I \\ & & & \\ \mu_{03}=\mu & \mu_{02}=\mu & \mu_{00}=\mu & \mu_{1}=\mu \\ -\frac{1}{2}a_t+\frac{3}{2}a'_{t+1} & -\frac{1}{2}a_t+\frac{1}{2}a'_{t+1}+d'_{(t+1)1} & -\frac{1}{2}a_t-\frac{1}{2}a'_{t+1}+d'_{(t+1)2} & -\frac{1}{2}a_t-\frac{3}{2}a'_{t+1} \\ -\frac{3}{4}I' & -\frac{1}{4}I'-\frac{1}{2}J'_1 & +\frac{1}{4}I'-\frac{1}{2}J'_2 & +\frac{3}{4}I \end{array}\right] \end{array}, \qquad (2)$$

where a'_{t+1} is the additive effect at endosperm $\mathbf{Q}'_{t+1}$, $d'_{(t+1)1}$ and $d'_{(t+1)2}$ are the dominance effects due to the intra-locus interaction between QQ and q and between Q and qq at $\mathbf{Q}'_{t+1}$, respectively, I' is the cross-generation maternal-additive × endosperm-additive epistatic effect, and J'_1 and J'_2 are the across-generation maternal-additive × endosperm-dominant epistatic effects for $d_{(t+1)1}$ and $d_{(t+1)2}$, respectively.

Assume that a pair of flanking markers $\mathbf{N}'^1_{t+1}$ and $\mathbf{N}'^2_{t+1}$ are used to map the endosperm $\mathbf{Q}'_{t+1}$. Let s', s'_1 and s'_2 be the recombination fractions between the two markers, marker $\mathbf{N}'^1_{t+1}$ and the QTL, and the QTL and marker $\mathbf{N}'^2_{t+1}$, respectively. The conditional probabilities of endosperm QTL genotypes given the across-generation maternal–embryo marker genotypes can be derived in terms of s', s'_1 and s'_2, depending on the type of pollination. These conditional probabilities for self-pollinated plants have been derived by some groups. The conditional probabilities for cross-pollinated plants can be similarly derived.

The embryo–endosperm interaction model

For the embryo ($\mathbf{Q}_{t+1}$) and endosperm ($\mathbf{Q}'_{t+1}$) QTL at the same generation $t+1$, we have 12 joint QTL genotypes whose values, $\mu_{j_t j_{t+1}}$, are expressed as

$$\begin{array}{cc} & \begin{array}{llll} \quad Q'_{t+1}Q'_{t+1}Q'_{t+1} & \quad Q'_{t+1}Q'_{t+1}q'_{t+1} & \quad Q'_{t+1}q'_{t+1}q'_{t+1} & \quad q'_{t+1}q'_{t+1}q'_{t+1} \end{array} \\ \begin{array}{c} Q_{t+1}Q_{t+1} \\ \\ \\ \\ Q_{t+1}q_{t+1} \\ \\ \\ \\ q_{t+1}q_{t+1} \\ \\ \\ \end{array} & \left[\begin{array}{llll} \mu_{23}=\mu & \mu_{22}=\mu & \mu_{21}=\mu & \mu_{20}=\mu \\ +\frac{1}{2}a_{t+1}+\frac{3}{2}a'_{t+1} & +\frac{1}{2}a_t+\frac{1}{2}a_{t+1}+d_{(t+1)1} & +\frac{1}{2}a_t-\frac{1}{2}a_{t+1}+d_{(t+1)2} & +\frac{1}{2}a_t-\frac{3}{2}a_{t+1} \\ +\frac{3}{4}\mathcal{I} & +\frac{1}{4}\mathcal{I}+\frac{1}{2}\mathcal{J}_1 & -\frac{1}{4}\mathcal{I}+\frac{1}{2}\mathcal{J}_2 & -\frac{3}{4}\mathcal{I} \\ & & & \\ \mu_{13}=\mu & \mu_{12}=\mu & \mu_{11}=\mu & \mu_{10}=\mu \\ +\frac{1}{2}a_t+\frac{3}{2}a_{t+1} & +\frac{1}{2}a_t+\frac{1}{2}a_{t+1}+d_{(t+1)1} & +\frac{1}{2}a_t-\frac{1}{2}a_{t+1}+d_{(t+1)2} & +\frac{1}{2}a_t-\frac{3}{2}a_{t+1} \\ +\frac{3}{4}\mathcal{I} & +\frac{1}{4}\mathcal{I}+\frac{1}{2}\mathcal{J}_1 & -\frac{1}{4}\mathcal{I}+\frac{1}{2}\mathcal{J}_2 & -\frac{3}{4}\mathcal{I} \\ & & & \\ \mu_{03}=\mu & \mu_{02}=\mu & \mu_{01}=\mu & \mu_{00}=\mu \\ +\frac{1}{2}a_t+\frac{3}{2}a_{t+1} & +\frac{1}{2}a_t+\frac{1}{2}a_{t+1}+d_{(t+1)1} & +\frac{1}{2}a_t-\frac{1}{2}a_{t+1}+d_{(t+1)2} & +\frac{1}{2}a_t-\frac{3}{2}a_{t+1} \\ +\frac{3}{4}\mathcal{I} & +\frac{1}{4}\mathcal{I}+\frac{1}{2}\mathcal{J}_1 & -\frac{1}{4}\mathcal{I}+\frac{1}{2}\mathcal{J}_2 & -\frac{3}{4}\mathcal{I} \end{array}\right] \end{array}, \qquad (3)$$

where $\mathcal{I}$ is the embryo-additive × endosperm-additive and embryo-dominant × endosperm-additive epistatic effect between embryo $\mathbf{Q}_{t+1}$ and endosperm $\mathbf{Q}'_{t+1}$, $\mathcal{J}_1$ and $\mathcal{J}_2$ are the embryo-additive × endosperm-additive epistatic effect for $d_{(t+1)1}$ and $d_{(t+1)2}$, respectively, $\mathcal{K}$ is the embryo-dominant × endosperm-dominant epistatic effect, and $\mathcal{J}_1$ and $\mathcal{J}_2$ are the embryo-dominant × endosperm-dominant epistatic effects for $d_{(t+1)1}$ and $d_{(t+1)2}$, respectively.

Similarly, the conditional probabilities of embryo–endosperm QTL genotypes given across-generation marker genotypes can be derived separately for two different cases in which the two QTL are located in the same interval or in different intervals. Such derivations will be different for self- and cross-pollinated systems.

As shown in Cui *et al*. (2004), the genetic effect parameter vectors $\mathbf{h}_1 = (\mu, a_t, a_{t+1}, d_{t+1}, I, J)$ for the maternal–embryo interaction model, $\mathbf{h}_2 = (\mu, a_t, a'_{t+1}, d'_{(t+1)1}, d'_{(t+1)2}, I', J'_1, J'_2)$ for the maternal–endosperm interaction model and $\mathbf{h}_3 = (\mu, a_{t+1}, d_{t+1}, a'_{t+1}, d'_{(t+1)1}, d'_{(t+1)2}, \mathcal{I}, \mathcal{J}_1, \mathcal{J}_2, \mathcal{K}, \mathcal{L}_1, \mathcal{L}_2)$ for the embryo–endosperm interaction model can be estimated from the corresponding genotypic values, $\mu_{j_t j_{t+1}}$, by solving a group of regular linear equations as contained in matrices (1)–(3). As can be seen below, we derive a closed-form solution for the EM algorithm to obtain the maximum-likelihood estimates (MLEs) of the genotypic values. Thus, the MLEs of the genetic effect parameters can be estimated accordingly.

STATISTICAL METHOD

Statistical model for multiple traits

Let us suppose there are three quantitative traits, one expressed in the maternal tissue (denoted by x), the second expressed in the embryo tissue (denoted by y) and the third expressed in the endosperm tissue (denoted by z). The three QTL from different genomes, $\mathbf{P}_t$, $\mathbf{Q}_{t+1}$ and $\mathbf{Q}'_{t+1}$, interact through coordinated pathways to affect each of these three traits. The statistical models for the phenotypic values of the three traits affected by the hypothetical epistatic QTL are formulated for each of the three types of genome–genome interactions.

For the maternal–embryo interaction model, the bivariate phenotypes (x_i, y_i) for seed i in the backcross population in terms of genotypic values, can be expressed, as

$$\begin{aligned} x_i &= \sum_{j_t=0}^{1} \sum_{j_{t+1}=0}^{2} m^x_{j_t j_{t+1}} \xi_{i j_t j_{t+1}} + e^x_i, \\ y_i &= \sum_{j_t=0}^{1} \sum_{j_{t+1}=0}^{2} m^y_{j_t j_{t+1}} \xi_{i j_t j_{t+1}} + e^y_i, \end{aligned} \tag{4}$$

where $\xi_{i j_t j_{t+1}}$ is the indicator variable defined as 1 if seed i carries the maternal–embryo QTL genotype $j_t j_{t+1}$ and 0 otherwise; $m^x_{j_t j_{t+1}}$ and $m^y_{j_t j_{t+1}}$ are the values of QTL genotype $j_t j_{t+1}$ for two traits x and y, respectively, and e^x_i and e^y_i are the residual errors that follow a bivariate normal distribution with means zero and covariance matrix

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma^2_x & \sigma_{xy} \\ \sigma_{yx} & \sigma^2_y \end{pmatrix}.$$

Note that we use the superscript or subscript x and y to distinguish between the two traits in genotypic values, genetic effects and residual effects and variances.

Equation (4) can be written, in matrix notation, as

$$\mathbf{u}_i = \sum_{j_t=0}^{1} \sum_{j_{t+1}=0}^{2} \mathbf{m}_{j_t j_{t+1}} \xi_{i j_t j_{t+1}} + \mathbf{e}_i, \tag{5}$$

where $\mathbf{u}_i = (x_i, y_i)$ is the vector for the phenotypic values of maternal and embryo traits for seed i, $\mathbf{m}_{j_t j_{t+1}} = (m^x_{j_t j_{t+1}}, m^y_{j_t j_{t+1}})$ is the vector for the genotypic values of a joint maternal–embryo QTL genotype and $\mathbf{e}_i = (e^x_i, e^y_i)$ is the vector for the residual effects of seed i.

For self-pollinated plants, the maternal parent receives no genes from other sources to generate its progeny. Thus, the gene segregation in the progeny would not lead to the variation of the maternal trait. To reflect this characteristic, the maternal–embryo interaction that occurs across generations should be modeled with the constraints

$$m^x_{12} = m^x_{11} = m^x_{10} \quad \text{and} \quad m^x_{02} = m^x_{01} = m^x_{00}, \tag{6}$$

which imply that embryo QTL $\mathbf{Q}_{t+1}$ has no genetic effect on trait x, i.e. $a^x_{t+1} = d^x_{t+1} = I^x = 0$ [see Matrix (1)].

Similarly, we can formulate a statistical model for the maternal–endosperm interaction, except for four triploid QTL genotypes at $\mathbf{Q}'_{t+1}$. But the embryo–endosperm interaction model will be different because such an interaction occurs within the same generation in which embryo (y) and endosperm traits (z) are also affected by a QTL from the opposite genome. The bivariate model for phenotypic traits (y, z) can be expressed as

$$\mathbf{w}_i = \sum_{j_{t+1}=0}^{2} \sum_{j'_{t+1}=0}^{3} \mathbf{m}_{j_{t+1} j'_{t+1}} \zeta_{i j_{t+1} j'_{t+1}} + \epsilon_i, \tag{7}$$

where $\mathbf{w}_i = (y_i, z_i)$ is the vector for the phenotypic values of embryo and endosperm traits for seed i, $\zeta_{i j_{t+1} j'_{t+1}}$ is the indicator for the embryo–endosperm QTL genotype, $\mathbf{m}_{j_{t+1} j'_{t+1}} = \left(m^x_{j_{t+1} j'_{t+1}}, m^y_{j_{t+1} j'_{t+1}}\right)$ is the vector for the genotypic values of a joint embryo–endosperm QTL genotype and $\epsilon_i = (e^y_i, e^z_i)$ is the vector for the residual effects of seed i.

Bivariate mixture model

Finite mixture models are a type of density model that comprises a number of component functions, usually Gaussian. These component functions are combined to provide a multimodal density. Gaussian mixture models can be employed to model genotypic segregation of specific genetic factors that determine quantitative traits. According to mixture models, each observation is assumed to have arisen from one of a known or unknown number of components (QTL genotypes), each component being modeled by a multivariate normal distribution density. Under the maternal–embryo epistasis model, the bivariate likelihood function of phenotypic traits ($\mathbf{u}$) and marker data ($\mathcal{M}$) based on mixture models is expressed as

$$L(\varpi, \mathbf{m}, \mathbf{\Sigma} | \mathbf{u}, \mathcal{M}) = \prod_{i=1}^{n} \left[\sum_{j_t=0}^{1} \sum_{j_{t+1}=0}^{2} \varpi_{j_t j_{t+1} | i} f_{j_t j_{t+1}}(\mathbf{u}_i; \mathbf{m}_{j_t j_{t+1}}, \mathbf{\Sigma}) \right], \tag{8}$$

where $\varpi = \{\varpi_{j_t j_{t+1} | i}\}$ is the vector for the conditional (or prior) probability of maternal–embryo QTL genotype $j_t j_{t+1}$ given a particular across-generation marker genotype for seed i and $\mathbf{m} = \{\mathbf{m}_{j_t j_{t+1}}\}$ is the vector of genotypic means for two traits that follow a bivariate normal distribution $N(\mathbf{m}_{j_t j_{t+1}}, \mathbf{\Sigma})$.

With the knowledge about conditional probabilities and genotypic values, we can construct similar mixture-based likelihood functions for the maternal–endosperm and embryo–endosperm interaction models. We provide a procedure for estimating the parameters contained in the likelihood functions.

The EM algorithm

Conditional probabilities are a function of the recombination fractions between QTL and their flanking markers and therefore can provide the information about QTL locations. Mean vectors and the covariance matrix are quantitative genetic parameters associated with the genetic effects of QTL. Let $\mathbf{\Omega} = (\varpi, \mathbf{m}, \mathbf{\Sigma})$ denote the unknown parameters. We implement the EM algorithm to obtain the MLE of $\mathbf{\Omega}$. The log-likelihood function of Equation (8) for the maternal–embryo interaction model is given by

$$\log L(\mathbf{\Omega}) = \sum_{i=1}^{n} \log \left[\sum_{j_t=0}^{1} \sum_{j_{t+1}=0}^{2} \varpi_{j_t j_{t+1} | i} f_{j_t j_{t+1}}(\mathbf{u}_i; \mathbf{m}_{j_t j_{t+1}}, \mathbf{\Sigma}) \right] \tag{9}$$

with a derivative for an unknown Ω_λ,

$$\begin{aligned}
&\frac{\partial}{\partial \Omega_\lambda} \log L(\Omega) \\
&= \sum_{i=1}^{n} \sum_{j_t=0}^{1} \sum_{j_{t+1}=0}^{2} \frac{\varpi_{j_t j_{t+1}|i} \frac{\partial}{\partial \Omega_\lambda} f_{j_t j_{t+1}}(\mathbf{u}_i; \mathbf{m}_{j_t j_{t+1}}, \mathbf{\Sigma})}{\sum_{j_t=0}^{1} \sum_{j_{t+1}=0}^{2} \varpi_{j_t j_{t+1}|i} f_j(\mathbf{u}_i; \mathbf{m}_{j_t j_{t+1}}, \mathbf{\Sigma})} \\
&= \sum_{i=1}^{n} \sum_{j_t=0}^{1} \sum_{j_{t+1}=0}^{2} \frac{\varpi_{j_t j_{t+1}|i} f_j(\mathbf{u}_i; \mathbf{m}_{j_t j_{t+1}}, \mathbf{\Sigma})}{\sum_{j_t=0}^{1} \sum_{j_{t+1}=0}^{2} \varpi_{j_t j_{t+1}|i} f_{j_t j_{t+1}}(\mathbf{u}_i; \mathbf{m}_{j_t j_{t+1}}, \mathbf{\Sigma})} \\
&\quad \times \frac{\partial}{\partial \Omega_\lambda} \log f_{j_t j_{t+1}}(\mathbf{u}_i; \mathbf{m}_{j_t j_{t+1}}, \mathbf{\Sigma}) \\
&= \sum_{i=1}^{n} \sum_{j_t=0}^{1} \sum_{j_{t+1}=0}^{2} \Pi_{ij} \frac{\partial}{\partial \Omega_\lambda} \log f_{j_t j_{t+1}}(\mathbf{u}_i; \mathbf{m}_{j_t j_{t+1}}, \mathbf{\Sigma}),
\end{aligned}$$

where we define

$$\Pi_{j_t j_{t+1}|i} = \frac{\varpi_{j_t j_{t+1}|i} f_{j_t j_{t+1}}(\mathbf{u}_i; \mathbf{m}_{j_t j_{t+1}}, \mathbf{\Sigma})}{\sum_{j=1}^{4} \varpi_{j_t j_{t+1}|i} f_{j_t j_{t+1}}(\mathbf{u}_i; \mathbf{m}_{j_t j_{t+1}}, \mathbf{\Sigma})}, \qquad (10)$$

which could be thought of as a posterior probability that seed i has joint maternal–embryo QTL genotype $j_t j_{t+1}$. We then implement the EM algorithm with the expanded parameter set $\{\mathbf{\Omega}, \mathbf{\Pi}\}$, where $\mathbf{\Pi} = \{\Pi_{j_t j_{t+1}|i}\}$. Conditional on $\mathbf{\Pi}$, we solve for the zeros of $(\partial/\partial\Omega_\lambda)\log L(\mathbf{\Omega})$ to get our estimates of $\mathbf{\Omega}$.

In the E-step, the prior conditional probabilities of the QTL genotypes given the marker genotypes and the normal distribution function are used to calculate the $\Pi_{j_t j_{t+1}|i}$ matrix. In the M-step, the calculated posterior probabilities are used to solve the unknown parameters using

$$\widehat{\mathbf{m}}_{j_t j_{t+1}} = \frac{\sum_{i=1}^{n} \Pi_{j_t j_{t+1}|i} \mathbf{u}_i}{\sum_{i=1}^{n} \Pi_{j_t j_{t+1}|i}}, \qquad (11)$$

$$\widehat{\mathbf{\Sigma}} = \frac{1}{n} \left[\sum_{i=1}^{n} \sum_{j_t=0}^{1} \sum_{j_{t+1}=0}^{2} \Pi_{j_t j_{t+1}|i} (\mathbf{y}_i - \widehat{\mathbf{m}}_{j_t j_{t+1}})(\mathbf{u}_i - \widehat{\mathbf{m}}_{j_t j_{t+1}})^{\mathrm{T}} \right]. \qquad (12)$$

Using sample parameters as initial values, we iterate the E and M steps between Equations (10) and (12) until the specified convergence criteria are satisfied. The values at convergence are regarded as the MLEs. The MLEs of the genotypic values $\mathbf{m}$ can be used to solve the MLEs of the genetic effects $\mathbf{h}$.

In the procedure described above for the EM algorithm, we treated the positions of QTL as known parameters, although their MLEs can also be obtained through iterative steps. We can use a grid approach to estimate the QTL positions. By hypothesizing a pair of embryo and endosperm QTL every 2 cM at marker intervals, we can draw the landscape of log-likelihood test statistics throughout the entire genome. The positions corresponding to the peak of the landscape across a linkage group are the MLEs of the QTL positions.

The MLEs of the QTL positions and effects under the maternal–endosperm and embryo–endosperm epistasis models can be similarly derived. The QTL effects are specified differently among these three models, depending on the dosage of QTL alleles (Table 1). As like in general QTL mapping models, the proportion of the total variance explained by each QTL from a different genome can be calculated for each trait.

Hypothesis testing

A number of statistical hypothesis tests can be performed for the underlying parameters of interest. The presence of the QTL from different genomes with joint effects on two quantitative traits expressed in different tissues can be tested by a log-likelihood ratio (LLR) test statistic calculated under the full model (assuming that there are such QTL) and the reduced model (assuming that there is no QTL). The LLR is asymptotically χ^2-distributed with the degrees of freedom that are equivalent to the number of unknown parameters estimated. For a mixture model like ours here, this may be violated due to some regularity problem (McLachlan and Peel, 2000). The critical threshold value for declaring the existence of the testing QTL is empirically calculated on the basis of permutation tests (Churchill and Doerge, 1994).

Table 1. The MLEs of the additive genetic effects of the embryo (a_{t+1}) and endosperm (a'_{t+1}) QTL and their additive × additive epistatic interaction effect ($\mathcal{I}$) on gel consistency in the endosperm measured for two different years in a backcross derived from two inbred lines in rice[a]

Trait measured in two years	$\widehat{\mu}$	$\widehat{a}_{t+1}$	$\widehat{a}'_{t+1}$	$\widehat{\mathcal{I}}$	$\widehat{\sigma}^2$	$\widehat{\rho}$
1999	36.62	17.54	−0.11	−2.06	42.87	0.2034
2000	44.51	13.95	1.08	1.00	30.06	
LLR for testing year-dependent difference		47.2	22.6	7.8		
P-value		6.42×10^{-12}	1.98×10^{-6}	0.0052		

[a] The residual variances (σ^2) and residual correlation (ρ) are estimated between gel consistency measured in 1999 and 2000.

After the existence of QTL from different genomes is tested, we can test the additive and dominant QTL effect from a particular genome and additive × additive, additive × dominant, dominant × additive and dominant × dominant epistatic effects derived from two different genomes. Our model allows for testing the effects of specific QTL on individual traits, although, for our experimental design, different genome–genome interaction models characterize different types of genetic effects. All these effect-specific tests are performed by implementing the EM algorithm and the critical value for declaring significance can be obtained empirically through simulation studies.

A WORKED EXAMPLE

The newly developed model was used to analyze published data on the endosperm in rice (Tan *et al.*, 1999). The F_1 heterozygote between two rice inbred lines, ZS97 and MH63, was self-crossed for 9 generations to produce 241 recombinant inbred lines (RILs) for high-resolution genetic mapping of genes influencing endosperm traits. These RILs that are homozygous for the alternative alleles were genotyped for 221 polymorphic markers distributed throughout the genome to construct a molecular linkage map composed of 12 rice chromosomes. These RILs as the female parent were backcrossed toward one original inbred line, ZS97, as the male parent to generate a backcross population containing 241 plants. All the backcross plants were evaluated for gel consistency in their endosperm tissues in two successive years (1999 and 2000) to determine any major QTL segregating in this material.

Because of the nature of this pedigree, we make some modifications to our general embryo–endosperm model to identify interacting QTL on embryo and endosperm tissues. First, the conditional probabilities that suit this pedigree are derived to predict the embryo–endosperm QTL genotypes based on the markers collected in the embryo. Second, in this design, the number of embryo–endosperm QTL genotypes is reduced to 4 and, thus, the genetic effects that can be estimated are the additive effects of embryo $\mathbf{Q}_{t+1}$ (a_{t+1}) and endosperm $\mathbf{Q}'_{t+1}$ (a'_{t+1}) and additive × additive epistatic effect ($\mathcal{I}$) between these two QTL. Third, our model was originally developed to analyze the phenotypes expressed in the embryo and endosperm, but the data for this design were collected from the endosperm in two different years. According to Falconer (1952), the same trait measured in different years can be viewed as different traits.

The phenotypic correlation between endosperm gel consistency measured in two different years is 0.68, suggesting that some common genetic basis is shared over years. A genome-wide scan was performed to detect the existence and distribution of interacting QTL throughout the entire genome. Significant joint genetic effects were detected between two QTL on chromosomes 6 and 8. The maximum LLR value throughout the genome is 270.9, markedly larger than the genome-wide critical threshold 30.5, empirically obtained from permutation tests at the 0.005 significance level. One of the detected significant QTL is located at 12.0 cM from the first marker on chromosome 6 of the embryo genome, whereas the second QTL is located at 29.4 cM from the first marker on chromosome 8 of the endosperm genome. The embryo QTL is located at a candidate gene, *Waxy*, that is associated with a critical step of amylose biosynthesis (Okagaki and Wessler, 1988), which well validates our model.

We estimated the additive effect, a_{t+1}, of the embryo QTL, the additive effect, a'_{t+1}, of the endosperm QTL and their epistatic effects, $\mathcal{I}$, on gel consistency in two different years (Table 1). Further hypotheses were performed for the significance tests of the additive and epistatic genetic effects. The LLRs for testing the significance of these effect parameters suggest that the additive effect of the embryo QTL is highly significant, whereas the additive effect of the endosperm QTL and the additive × additive effect between the two QTL are significant, but at lower levels.

In this example, we can use our model to test how genetic effects are expressed differently from year to year. If the genetic effect of a QTL is year-dependent, then this QTL is thought to display a significant genotype × year interaction. Figure 1 illustrates the unparallel changes of the four joint embryo–endosperm QTL genotypes across different years for gel consistency in the endosperm. The LLR test for the year-dependent non-parallel response suggests that there are significant QTL × year interactions ($P < 0.0001$). Further tests indicate that the additive effects of the QTL from the two genomes are expressed differently between the two years studied ($P = 6.42 \times 10^{-12}$ for the embryo QTL and $P = 1.98 \times 10^{-6}$ for the endosperm QTL; Table 1). The additive × additive epistatic effect between the embryo and endosperm QTL is also different between the two years ($P = 0.005$). These pieces of information obtained from data analyses by our model are fundamental to the design of crop breeding aimed at improving high-quality starch in rice.

MONTE CARLO SIMULATION

We carried out a series of simulation studies to examine the statistical properties of our genome–genome models by focusing on the epistatic model from the embryo and endosperm genomes. A similar statistical behavior should be held for the other two epistatic models, maternal–embryo and maternal–endosperm. Our simulation studies aim to examine the model performance under different situations when heritability, sample size and QTL location change. Five equidistant markers are simulated from the embryo population and are ordered as $\mathcal{M}_1$–$\mathcal{M}_5$ on a linkage group with the length of 80 cM. The Haldane map function was used to convert the map distance into the recombination fraction. For simplicity, we use two traits to achieve our goals. Three different combinations of heritability between two traits (0.1, 0.1), (0.1, 0.4) and (0.4, 0.4) and two different sample sizes (200, 400) were used.

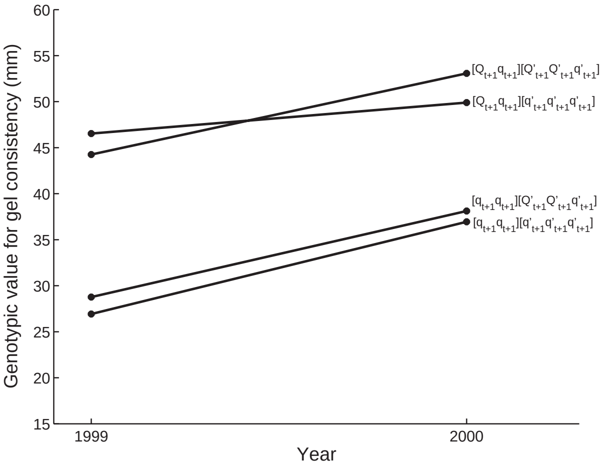

Fig. 1. Four joint genotypic values at the embryo ($\mathbf{Q}_{t+1}$) and endosperm ($\mathbf{Q}'_{t+1}$) QTL for endosperm-specific gel consistency (mm) measured for rice in two different years. Data from Tan *et al.* (1999).

Suppose there are two different putative QTL on the embryo and endosperm genomes. Both the embryo ($\mathbf{Q}_{t+1}$) and endosperm ($\mathbf{Q}'_{t+1}$) QTL are assumed to pleiotropically affect two traits, one expressed in the embryo (y) and the other expressed in the endosperm (z). The two QTL could be either linked together and located on the same marker interval or located on different marker intervals. The phenotypic values for each seed were simulated according to a bivariate normal distribution with different joint QTL genotypic values, determined by effect parameters, the overall (μ), additive effect of $\mathbf{Q}_{t+1}$ (a_{t+1}), additive effect of $\mathbf{Q}'_{t+1}$ (a'_{t+1}), the additive × additive epistatic effect ($\mathcal{I}$) between the two QTL for each trait, y and z, and residual variances (σ^2) and correlation (ρ).

Tables 2 and 3 give the hypothesized values and MLEs of the QTL effect parameters for each trait, as well as the square roots of the mean squared errors used to evaluate the precision and accuracy of the parameter estimation, under different simulation schemes. In general, our model can provide reasonable estimates of the parameters with estimation precision increasing with increased heritability levels and sampling sizes. The QTL position estimates when located in the same interval (Table 3) were not as good as when they were located at different intervals (Table 2). But this problem can be avoided if it is possible to increase the density of mapped markers to reduce the probability that two QTL are located in the same interval.

Our model has an excellent capacity to detecst epistatically interacting embryo and endosperm QTL effects. In all cases of different sample sizes and heritabilities, the maximum values of the LLR landscapes from 100 simulation replicates are all beyond the critical thresholds at the $\alpha = 0.001$ level determined from 1000 permutation tests for the simulated data. Furthermore, there is reasonable estimation precision for the additive × additive genetic effects even when the heritability is at a modest level.

DISCUSSION

We have proposed a general statistical framework for simultaneously mapping multiple correlated traits expressed in different

Table 2. The MLEs of the QTL position and effect parameters exerted by an embryo QTL and an endosperm QTL each on different intervals for a backcross of size 400 under different heritability combinations and residual variances estimated from 100 simulation replicates

True parameter	$\rho = 0.1$			$\rho = 0.6$		
	0.1, 0.1	0.1, 0.4	0.4, 0.4	0.1, 0.1	0.1, 0.4	0.4, 0.4
$\tau_{t+1} = 8$	7.16	7.68	8.16	7.04	7.72	7.52
	(4.7729)	(3.1653)	(2.4120)	(5.1787)	(3.0878)	(2.5421)
$\tau'_{t+1} = 48$	48.16	48.12	48.04	48.52	48.24	47.72
	(4.0201)	(2.7561)	(2.6968)	(4.5302)	(2.8426)	(2.6360)
$\mu^y = 10$	9.9959	9.9959	10.0028	9.9943	9.9957	9.9987
	(0.0698)	(0.0565)	(0.0276)	(0.0683)	(0.0564)	(0.0293)
$a^y_{t+1} = 0.5$	0.5019	0.5003	0.4906	0.4959	0.5020	0.4973
	(0.1650)	(0.1428)	(0.0572)	(0.1559)	(0.1391)	(0.0667)
$a'^y_{t+1} = 0.5$	0.4940	0.5059	0.5062	0.4921	0.5049	0.5047
	(0.1673)	(0.1545)	(0.0518)	(0.1591)	(0.1525)	(0.0659)
$\mathcal{I}^y = 0.3$	0.3118	0.2884	0.2937	0.3334	0.2936	0.3088
	(0.2980)	(0.2966)	(0.1173)	(0.3097)	(0.2952)	(0.1115)
$\mu^z = 11$	10.9983	10.9989	11.0022	10.9953	10.9981	11.0000
	(0.0891)	(0.0311)	(0.0328)	(0.0882)	(0.0299)	(0.0378)
$a^z_{t+1} = 0.6$	0.5818	0.5953	0.5971	0.5946	0.5980	0.5907
	(0.1983)	(0.0844)	(0.0718)	(0.1759)	(0.0850)	(0.0746)
$a'^z_{t+1} = 0.6$	0.6364	0.6028	0.5991	0.6089	0.6031	0.6163
	(0.2035)	(0.0783)	(0.0623)	(0.1856)	(0.0811)	(0.0830)
$\mathcal{I}^z = 0.4$	0.4388	0.3902	0.4028	0.3853	0.3891	0.4102
	(0.3462)	(0.1303)	(0.1419)	(0.3658)	(0.1344)	(0.1372)
σ_y^2	1.1712	1.1783	0.1943	1.1545	1.1774	0.1957
	(0.0884)	(0.0831)	(0.0153)	(0.0869)	(0.0820)	(0.0160)
σ_z^2	1.7145	0.2819	0.2791	1.6967	0.2850	0.2876
	(0.1473)	(0.0226)	(0.0249)	(0.1114)	(0.0236)	(0.0262)
ρ_{yz}	0.0933	0.1060	0.0941	0.5990	0.6055	0.5998
	(0.0506)	(0.0538)	(0.0511)	(0.0347)	(0.0345)	(0.0333)

The squared roots of the mean square errors of the MLEs are given in parentheses.
The locations (τ_{t+1} and τ'_{t+1}) of the two QTL are described by the map distances (in cM) from the first marker of the linkage group (80 cM long). The hypothesized σ_y^2 value is 1.1756 for $H^2 = 0.1$ and 0.1959 for $H^2 = 0.4$. The hypothesized σ_z^2 value is 1.71 for $H^2 = 0.1$ and 0.285 for $H^2 = 0.4$.

genome-specific tissues. Different from previous multitrait QTL mapping (Jiang and Zeng, 1995; Korol *et al*., 1995; Knott and Haley, 2000; Evans, 2002; Lund *et al*., 2003), our model framework implements interactions between multiple QTL located on different genomes. It has been well recognized that the coordinated expression of genes from different genomes is essential for the proper development of organs. For example, in higher plants, support and nourishment of embryo and endosperm tissues by the maternal tissue is fundamental to proper seed development (Chaudhury *et al*., 2001; Evans and Kermicle, 2001; Dilkes *et al*., 2002; Walbot and Evans, 2003).

The current literature has well established the belief that multiple correlated traits can add information to each other and, therefore, multitrait linkage analysis can give rise to more precise inferences about the position and effects of pleiotropic QTL affecting multiple traits, as compared to single-trait analyses (Jiang and Zeng, 1995; Korol *et al*., 1995; Knott and Haley, 2000; Evans, 2002; Wu *et al*., 2002c; Lund *et al*., 2003). Somewhat equivalent to the role of repeated measurements, information from correlated traits can reduce the effect of error variance, thus making it easier (more powerful) to detect QTL. Not only is the power of QTL detection increased, but also the estimation of the QTL map position is more precise. The model proposed in this paper deals with a different type of trait correlation that occurs between different individuals connected through coherent pathways. The best example is the impact of the growth vigor of a plant on its seed development by supplying adequate nutrients. In light of the consideration of the coordinated expression of traits owing to genes and development, our model, which can be viewed as ‘high-dimensional’, should be able to produce results that are closer to biological realism than those without such a solid developmental basis of phenotypic traits.

The statistical behavior of our high-dimensional model has been carefully investigated through computer simulation. The model has been found to provide reasonable power and estimation of interactive QTL from the embryo and endosperm genomes in a range of trait heritabilities and sample sizes. Nevertheless, the best validation for our model may be the successful detection of significant QTL that exert considerable effects on an endosperm trait measured in two consecutive years. These two annual measurements can be viewed as two different traits (Falconer, 1952). Previous approaches for endosperm mapping are purely based on the triploid inheritance of the endosperm (Wu *et al*., 2002a,b; Xu *et al*., 2003; Kao, 2004). Our model has the power to identify interactive QTL from the embryo and endosperm genomes. Using our high-dimensional model, both the embryo and endosperm genomes were detected to harbor QTL for gel consistency in rice, with the embryo QTL located almost at the same position as the *Waxy* gene on the short arm of chromosome 6 (Terada *et al*., 2002). The *Waxy* gene is known to influence a

Table 3. The MLEs of the QTL position and effect parameters exerted by an embryo QTL and an endosperm QTL on the same interval for a backcross of size 400 under different heritability combinations and residual variances estimated from 100 simulation replicates

True parameter	$\rho = 0.1$			$\rho = 0.6$		
	0.1, 0.1	0.1, 0.4	0.4, 0.4	0.1, 0.1	0.1, 0.4	0.4, 0.4
$\tau_{t+1} = 8$	7.96	7.56	7.12	9.88	6.80	7.40
	(5.1010)	(4.2732)	(3.7271)	(5.5818)	(4.6867)	(4.3107)
$\tau'_{t+1} = 16$	36.68	25.00	21.84	43.48	29.12	27.28
	(31.7777)	(18.6664)	(16.9318)	(37.1637)	(25.5374)	(22.6347)
$\mu^y = 10$	10.0120	10.0239	10.0113	10.0647	10.0371	10.0295
	(0.1817)	(0.1796)	(0.0777)	(0.1593)	(0.1522)	(0.0855)
$a^y_{t+1} = 0.5$	0.6763	0.5809	0.5245	0.7946	0.5766	0.6115
	(0.4938)	(0.4352)	(0.2479)	(0.5565)	(0.3981)	(0.3063)
$a'^y_{t+1} = 0.5$	0.3175	0.4189	0.4756	0.2199	0.4414	0.3825
	(0.5002)	(0.4294)	(0.2376)	(0.5676)	(0.3994)	(0.3159)
$\mathcal{I}^y = 0.3$	0.2873	0.2722	0.2486	0.0340	0.1529	0.1643
	(0.8087)	(0.7277)	(0.3203)	(0.7375)	(0.6813)	(0.3629)
$\mu^z = 11$	11.0539	11.0403	11.0269	11.0819	11.0417	11.0481
	(0.2347)	(0.1070)	(0.0957)	(0.2681)	(0.0905)	(0.1067)
$a^z_{t+1} = 0.6$	0.8667	0.7386	0.6237	0.9166	0.7071	0.7550
	(0.6054)	(0.3748)	(0.3233)	(0.7003)	(0.3634)	(0.3824)
$a'^z_{t+1} = 0.6$	0.3220	0.4737	0.5807	0.3047	0.4958	0.4420
	(0.6102)	(0.3686)	(0.3200)	(0.6991)	(0.3658)	(0.3954)
$\mathcal{I}^z = 0.4$	0.1343	0.2613	0.2952	0.0677	0.2200	0.2004
	(1.0798)	(0.4213)	(0.3821)	(1.1530)	(0.3931)	(0.4624)
σ^2_y	1.1525	1.1506	0.1921	1.1390	1.1468	0.1938
	(0.1069)	(0.0928)	(0.0164)	(0.1025)	(0.0914)	(0.0152)
σ^2_z	1.6621	0.2815	0.2798	1.6622	0.2771	0.2803
	(0.1473)	(0.0208)	(0.0224)	(0.1298)	(0.0239)	(0.0218)
ρ_{yz}	0.1014	0.0992	0.1002	0.6007	0.5994	0.6024
	(0.0445)	(0.0513)	(0.0504)	(0.0345)	(0.0318)	(0.0353)

The squared roots of the mean square errors of the MLEs are given in the parentheses.
The locations (τ_{t+1} and τ'_{t+1}) of the two QTL are described by the map distances (in cM) from the first marker of the linkage group (80 cM long). The hypothesized σ^2_y value is 1.1756 for $H^2 = 0.1$ and 0.1959 for $H^2 = 0.4$. The hypothesized σ^2_z value is 1.71 for $H^2 = 0.1$ and 0.285 for $H^2 = 0.4$.

major step in amylose synthesis in the endosperm for many grasses including maize and rice. Our bivariate mapping model also has the power to discern how genetic effects of the embryo and endosperm QTL are different across years. Whereas the embryo QTL triggers a large effect on gel consistency, a significant additive effect × interaction year of the endosperm QTL suggests that this QTL can modify the endosperm trait to make seed development better adapted to a year-to-year environmental change. Beyond traditional single trait mapping, our high- dimensional mapping model can detect the interaction for gel consistency between the additive × additive epistatic effect and year of interaction. Further functional analysis of these detected embryo and endosperm QTL will accelerate their usefulness to improve the quality and quantity of rice grains.

The derivations of our model were based on the plant system that undergoes self-pollinated reproduction. This model can be extended in several aspects. First, by incorporating unique segregation patterns of genes in the mixture-based likelihood function, this model can be modified to map genome–genome interactive QTL for cross-pollinated systems. Such a modified model will also be useful for animals in which birth weight is influenced by the uterine environment through the coordinated expression of the maternal and offspring QTL. Second, a mature cereal plant contains three sets of genomes, the maternal, embryo and endosperm. The current model allows for the modeling of interactions between any two sets of genomes. It is crucial to extend it to consider the triple-genome interactions among these three organs. With this triple interaction model, we can understand better the network of gene expression and regulation during seed development.

ACKNOWLEDGEMENTS

We thank the two anonymous referees for their constructive comments on this manuscript. This work is supported by an Outstanding Young Investigator Award of the National Natural Science Foundation of China (30128017), a University of Florida Research Opportunity Fund (02050259) and a University of South Florida Biodefense grant (7222061-12) to R.W. We thank Drs Jianguo Wu and Chunhai Shi at Zhejiang University for providing molecular marker and phenotypic data. The publication of this manuscript was approved as Journal Series No. R-10584 by the Florida Agricultural Experiment Station.

REFERENCES

Almasy,L. *et al.* (1997) Bivariate quantitative trait linkage analysis: pleiotropy versus co-incident linkages. *Genet. Epidemiol.*, **14**, 953–958.

Chaudhury,A.M. *et al.* (2001) Control of early seed development. *Ann. Rev. Cell Dev. Biol.*, **17**, 677–699.

Churchill,G.A. and Doerge,R.W. (1994) Empirical threshold values for quantitative trait mapping. *Genetics*, **138**, 963–971.

Cui,Y.H. *et al*. (2004) Mapping quantitative trait locus interactions from the maternal and offspring genomes. *Genetics*, **167**, 1017–1026.

Dempster,A.P. *et al*. (1977) Maximum likelihood from incomplete data via EM algorithm. *J. Roy. Stat. Soc. B*, **39**, 1–38.

Dilkes,B.P. *et al*. (2002) Genetic analyses of endoreduplication in *Zea mays* endosperm: evidence of sporophytic and zygotic maternal control. *Genetics*, **160**, 1163–1177.

Evans,D.M. (2002) The power of multivariate quantitative-trait loci linkage analysis is influenced by the correlation between the variables. *Am. J. Hum. Genet.*, **70**, 1599–1602.

Evans,M.M.S. and Kermicle,J.L. (2001) Interaction between maternal effect and zygotic effect mutations during maize seed development. *Genetics*, **159**, 303–315.

Falconer,D.S. (1952) The problem of environment and selection. *Am. Nat.*, **86**, 293–298.

Falconer,D.S. and Mackay,T.F.C. (1996) *Introduction to Quantitative Genetics*, edn. 4. Longmans Green, Harlow, Essex, UK.

Jiang,C. and Zeng,Z.-B. (1995) Multiple trait analysis of genetic mapping of quantitative trait loci. *Genetics*, **140**, 1111–1127.

Kao,C.-H. (2004) Multiple-interval mapping for quantitative trait loci controlling endosperm traits. *Genetics*, **167**, 1987–2002.

Knott,S.A. and Haley,C.S. (2000) Multitrait least squares for quantitative trait loci detection. *Genetics*, **156**, 899–911.

Korol,A.B. *et al*. (1995) Interval mapping of quantitative trait loci employing correlated trait complexes. *Theor. Appl. Genet.*, **92**, 998–1002.

Lander,E.S. and Botstein,D. (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics*, **121**, 185–199.

Lloyd,D.J. and Martin,T.E. (2004) Nest-site preference and maternal effects on offspring growth. *Behavioral Ecology*, **15**, 816–823.

Lund,M.S. *et al*. (2003) Multitrait fine mapping of quantitative trait loci using combined linkage disequilibria and linkage analysis. *Genetics*, **163**, 405–410.

Mackay,T.F.C. (2001) Quantitative trait loci in Drosophila. *Nat. Rev. Genet.*, **2**, 11–20.

McLachlan,G.J. and Peel,D. (2000) *Finite Mixture Models*. Wiley, New York.

Okagaki,R.J. and Wessler,S.R. (1988) Comparison of non-mutant and mutant *waxy* genes in rice and maize. *Genetics*, **120**, 1137–1143.

Opsahl-Ferstad,H.G. *et al*. (1997) *ZmEsr*, a novel endosperm-specific gene expressed in a restricted region around the maize embryo. *Plant J.*, **12**, 235–246.

Scheiner,S.M. (1993) Genetics and evolution of phenotypic plasticity. *Ann. Rev. Ecol. Syst.*, **24**, 25–68.

Tan,Y.F. *et al*. (1999) The three important traits for cooking and eating quality of rice grains are controlled by a single locus in an elite rice hybrid, Shanyou 63. *Theor. Appl. Genet.*, **99**, 642–648.

Terada,R. *et al*. (2002) Efficient gene targeting by homologous recombination in rice. *Nat. Biotech.*, **20**, 1030–1034.

van Hengel,A.J. *et al*. (1998) Expression pattern of the carrot EP3 endochitinase genes in suspension cultures and in developing seeds. *Plant Phys.*, **117**, 43–53.

Walbot,W. and Evans,N.M.S. (2003) Unique features of the plant life cycle and their consequences. *Nat. Rev. Genet.*, **4**, 369–379.

Whitlock,M.C. *et al*. (1995) Multiple fitness peaks and epistasis. *Ann. Rev. Ecol. Syst.*, **26**, 601– 629.

Wolf,J.B. (2000) Gene interactions from maternal effects. *Evolution,* **54**, 1882–1898.

Wolf,J.B. (2003) Genetic architecture and evolutionary constraint when the environment contains genes. *Proc. Natl Acad. Sci. USA*, **100**, 4655–4660.

Wolf,J.B. *et al*. (1998) Evolutionary consequences of indirect genetic effects. *Trends Ecol. Evol.*, **13**, 64–69.

Wolf,J.B. *et al*. (2002) Contribution of maternal effect QTL to genetic architecture of early growth in mice. *Heredity*, **89**, 300–310.

Wu,R.L. *et al*. (2002a) Statistical methods for dissecting triploid endosperm traits using molecular markers: an autogamous model. *Genetics*, **162**, 875–892.

Wu,R.L. *et al*. (2002b) An improved genetic model generates high-resolution mapping of QTL for protein quality in maize endosperm. *Proc. Natl Acad. Sci. USA*, **99**, 11281–11286.

Wu,R.L. *et al*. (2002c) A statistical model for the genetic origin of allometric scaling laws in biology. *J. Theor. Biol.*, **217**, 275–287.

Xu,C. *et al*. (2003) Mapping quantitative trait loci underlying triploid endosperm traits. *Heredity*, **90**, 228– 235.

Zeng,Z.-B. (1994) Precision mapping of quantitative trait loci. *Genetics*, **136**, 1457–1468.