A bivariate variance components model for mapping iQTLs underlying endosperm traits

Gengxin Li[1], Cen Wu[1], Cintia Coelho[2], Rongling Wu[3,4], Brian A. Larkins[2], Yuehua Cui[1]

[1]Department of Statistics and Probability, Michigan State University, East Lansing, MI 48824, USA, [2]Department of Plant Sciences, University of Arizona, Tucson, AZ 85721, USA, [3]Center for Statistical Genetics, Pennsylvania State University, Hershey, PA 17033, USA, [4]Center for Computational Biology, Beijing Forestry University, Beijing, People's Republic of China

TABLE OF CONTENTS

1. ABSTRACT

Genomic imprinting plays a pivotal role in early stage development in plants. Linkage analysis has been proven to be useful in mapping imprinted quantitative trait loci (iQTLs) underlying imprinting phenotypic traits in natural populations or experimental crosses. For correlated traits, studies have shown that multivariate genetic linkage analysis can improve QTL mapping power and precision, especially when a QTL has a pleiotropic effect on several traits. In addition, the joint analysis of multiple traits can test a number of biologically interesting hypotheses, such as pleiotropic effects vs close linkage. Motivated by a triploid maize endosperm dataset, we extended the variance components linkage analysis model incorporating imprinting effect proposed by Li and Cui (2010) to a bivariate trait modeling framework, aimed to improve the mapping precision and to identify pleiotropic imprinting effects. We proposed to partition the genetic variance of a QTL into sex-specific allelic variance components, to model and test the imprinting effect of an iQTL on two traits. Both simulation studies and real data analysis show the power and utility of the method.

2. INTRODUCTION

With the availability of linkage map and molecular markers in many species, coupled with the development of statistical and computational methods, enormous progress has been made in the identification of novel genes or quantitative trait loci (QTL) underlying various complex traits of interest (e.g. 1). Recent advances in biotechnology have enabled the generation of high throughput genome-wide dense single nucleotide polymorphism (SNP) data. Even though there have been large successes in genome-wide association studies (GWAS), GWAS still cannot substitute QTL mapping due to high false positive and false negative rates compared to linkage mapping (2). In linkage mapping, when multiple correlated traits are available, a number of studies have shown that jointly modeling correlated traits can significantly improve the power and mapping precision to detect QTLs (3-5). For a gene with a pleiotropic effect, often it might code for a product which has a signaling function on various targets. This is practically important as the gene might belong to a signaling pathway and could be a potential target for further functional validation. This

makes linkage mapping with multiple traits particularly attractive as it can test the pleiotropic effect of a QTL (6).

Genomic imprinting is an epigenetic phenomenon in which the expression of the same alleles could be different, owning to their parental origin (7). It plays a critical role in early stage development in many species, which makes it practically important to identify imprinted genes underlying various traits of interest (8). In addition to Mendelian traits, genomic imprinting has been routinely considered in genetic mapping, with the aim to identify genes with sex-specific expression due to epigenetic modification. So far, various statistical attempts have been made to map imprinted QTLs (iQTLs) underlying complex traits (e.g. 9-16).

In current applications, statistical methods for iQTL mapping are predominantly focused on single trait analysis, termed univariate iQTL mapping. In real application, it is frequent to observe multivariate traits which are potentially controlled by imprinted genes. For example, the percentage of endoreduplication and mean ploidy level in maize endosperm are two highly correlated traits (17). The two traits describe the level of endoreduplication in endosperm, which is thought to be genetically controlled by imprinted genes (18). We have previously developed a variance components statistical framework for mapping iQTLs underlying the single endosperm trait (see 16). Considering the advantage of joint analysis of multiple traits in QTL mapping, in this study we propose a multivariate variance components model in mapping iQTLs underlying multivariate imprinted traits, and further determine if there is a pleiotropic iQTL effect (6, 19).

Our work is based on a published endosperm mapping data set, and the biological application makes the study particularly attractive (17). Endosperm is a triploid tissue resulting from a unique double fertilization process in angiosperms. As a result, the endosperm genome carries two copies of chromosomes inherited from female parent and one copy from male parent. Maize endosperm cells undergoing endoreduplication are generally larger than other cells, which consequently results in larger fruits or seeds and is beneficial to human beings (20). It is thus particularly important from the breeding point of view to identify which genes control the endoreduplication process and where they are located in the genome. To our best knowledge, no study has been conducted to map iQTLs underlying the imprinting process with multivariate traits.

Variance components models have been shown to be powerful tools in multi-trait linkage analysis for an outbred or human population (e.g., 3, 4). Due to the special inbreeding structure and unique genetic make-up of the endosperm genome, the current multi-trait linkage analysis methods cannot be directly applied to the endosperm genome. We have previously shown that the variance components model can be applied to an inbreeding population to identify imprinted QTLs underlying a univariate endosperm trait (16). As an extension to our previous method, in this work we propose a bivariate iQTL

mapping method to target iQTLs with potential pleiotropic effects. This study will fill the gap in genetic mapping iQTL underlying multiple endosperm traits by considering the imprinting property of a QTL.

3. STATISTICAL METHOD

3.1. The genetic model and parent-specific allelic sharing

We consider a backcross design initiated with two inbreeding parental lines with a large contrast in the phenotype of interest. Denote the genotype of two parental lines as AA, aa, the F1 lines (Aa) as the maternal parents to backcross with both parental lines to generate two backcross segregation populations. In terms of the endosperm genotype, the backcross offspring is denoted as $A_m A_m A_f$, $a_m a_m A_f$ and $a_m a_m a_f$, where the subscript m or f implies that the corresponding allele is inherited from the maternal or paternal parent, respectively. Similarly, we can use the two parental lines as the maternal parent and backcross with the F1 line to generate two different backcross populations which contain the same sets of genotype as the other two crosses. For a detailed description of the backcross design, readers are referred to table 1 in Li and Cui (16). Note that the endosperm genotypes resulting from the backcross design contain two identical gene copies from the maternal parent and are different from a regular diploid mapping population.

Consider two phenotypic traits of interest. Let $\boldsymbol{y}_{1_k} = (y_{1_1},...,y_{1_{n_k}})^T$ and $\boldsymbol{y}_{2_k} = (y_{2_1},...,y_{2_{n_k}})^T$ be two vectors of observed trait values for trait 1 and 2 in the kth backcross family, where n_k is the number of observations in family $k(=1,...,K)$. We assume a multivariate normality for the joint distribution of $\boldsymbol{y}_{1_k}$ and $\boldsymbol{y}_{2_k}$. Denote the genotype-specific cytoplasmic maternal effects as (β_{1_k}, β_{2_k}), additive genetic effects at a QTL as (a_{1_k}, a_{2_k}), polygenic additive effects as (g_{1_k}, g_{2_k}), and random environmental effects as (e_{1_k}, e_{2_k}) for the two traits in a bivariate model. To consider the imprinting effect of a QTL, the additive genetic effects are further partitioned into parent-of-origin effects due to the maternal alleles with respect to each trait (denoted as $a_{1_{mk}}$, $a_{2_{mk}}$), and effects due to the paternal allele (denoted as $a_{1_{fk}}$, $a_{2_{fk}}$). The genetic model underlying two endosperm traits can be expressed as:

$$(\boldsymbol{y}_{1_k}, \boldsymbol{y}_{1_k}) = (\beta_{1_k}, \beta_{2_k}) + 2(a_{1_{mk}}, a_{2_{mk}}) + (a_{1_{fk}}, a_{2_{fk}})$$
$$+ (g_{1_k}, g_{2_k}) + (e_{1_k}, e_{2_k})$$

Equation (1)

where the coefficient of the maternal allele effect is set as 2 due to the two identical copies.

For the proposed backcross design, there are a total of three possible maternal genotypes, denoted as AA, Aa and aa. Thus (β_{1_k} and β_{2_k}) denote mean parameters of two traits with respect to three maternal genotypes, i.e., $\beta_{1_k} = (\mu_{11_k}, \mu_{12_k}, \mu_{13_k})^T$, $\beta_{2_k} = (\mu_{21_k}, \mu_{22_k}, \mu_{23_k})^T$. The random effects corresponding to trait $j(=1,2)$ are $a_{j_{mk}}$, $a_{j_{fk}}$, g_{j_k} and e_{j_k} which are normally distributed, i.e., $a_{j_{mk}} \sim N(0, \prod_{m|k} \sigma^2_{m_j})$, $a_{j_{fk}} \sim N(0, \prod_{f|k} \sigma^2_{f_j})$, $g_{j_k} \sim N(0, \Phi_k \sigma^2_{g_j})$ and $e_{j_k} \sim N(0, I_k \sigma^2_{e_j})$, where $\sigma^2_{m_j}$ and $\sigma^2_{f_j}$ are the additive genetic variances at a QTL for the maternal and paternal alleles respectively; $\prod_{m|k}$ and $\prod_{f|k}$ are identical-by-descent (IBD) sharing matrices that are derived from the maternal and paternal alleles among sib-pairs, respectively; $\sigma^2_{g_j}$ and $\sigma^2_{e_j}$ are the additive polygenic and residual variances, respectively; Φ_k is the expected proportion of alleles shared IBD; and I_k is the identity matrix. The above model is similar to a bivariate variance components model described in Almasy et al. (4), except that here we incorporate the parent-specific allelic effects. For two correlated traits, the covariances of random effects are expressed as $\text{Cov}(a_{1_{mk}}, a_{2_{mk}}) = \prod_{m|k} \sigma_{m_{12}}$ and $\text{Cov}(a_{1_{fk}}, a_{2_{fk}}) = \prod_{f|k} \sigma_{f_{12}}$ for the additive genetic effects at a QTL; $\text{Cov}(g_{1_k}, g_{2_k}) = \Phi_k \sigma_{g_{12}}$ for the polygenic effects; and $\text{Cov}(e_{1_k}, e_{2_k}) = I_k \sigma_{e_{12}}$ for the residual effects.

The above variance components model is built upon the basis of IBD sharing at a QTL. For a triploid inbreeding population, a unique decomposition of parent-specific allele sharing is illustrated in figure 1 of Li and Cui (16). Following the definition given in Li and Cui (16), the phenotypic variance-covariance corresponding to trait j (=1, 2) in family k can be expressed as:
$$\Sigma_{kj} = \prod_{m|k} \sigma^2_{m_j} + \prod_{m/f|k} \sigma^2_{mf_j} + \prod_{f|k} \sigma^2_{f_j}$$
$$+ \Phi_k \sigma^2_{g_j} + I_k \sigma^2_{e_j}$$
, where $\prod_{m/f|k}$ is the IBD sharing matrix due to cross-sharing of allele derived from different parents and $\sigma^2_{mf_j}$ is the variance

term due to allele cross-sharing for an inbreeding population (16). Note that $\sigma^2_{mf_j} = 0$ for an non-inbreeding population, but it could be non-zero for a partially inbreeding population (21). The covariance of two phenotypic traits is expressed as

$$\Sigma_{k12} = \prod_{m|k} \sigma_{m_{12}} + \prod_{m/f|k} \sigma_{mf_{12}} + \prod_{f|k} \sigma_{f_{12}}$$
$$+ \Phi_k \sigma_{g_{12}} + I_k \sigma_{e_{12}}$$

With the above notation, the phenotypic variance-covariance matrix of two phenotypic traits within the kth backcross family can be expressed as

$$\Sigma_k = \begin{pmatrix} \Sigma_{k1} & \Sigma_{k12} \\ & \Sigma_{k2} \end{pmatrix} \quad \text{Equation (2)}$$

The IBD sharing probability mentioned above is calculated assuming that a QTL is located at a marker position. Unless markers are dense enough, a QTL can be anywhere in the genome bracketed by two flanking markers. Here we assume a QTL can be anywhere in the genome and calculate the IBD sharing probability based on the recombination information between a putative QTL and two flanking markers. In a genome-wide linkage scan, we search a QTL every 1 or $2cM$ throughout the entire genome and the conditional probability of a QTL conditional on two flanking markers can be calculated (see 22). These conditional probabilities are then considered when calculating the IBD probability of a putative QTL at a given genome position (see 16 for more details).

3.2. Parameter estimation

Let $\mathbf{y}_k = (y_{1_k}, y_{2_k})^T$. Assuming multivariate normality of $\mathbf{y}_k$ and independence between different families, the log-likelihood function for K families can be expressed as

$$\ell(\Omega) = \sum_{k=1}^{K} \log\{f(\mathbf{y}_k; \beta, \Sigma_k)\} \quad \text{Equation (3)}$$

where $\Omega = (\beta, \theta)$ and $\beta = (\mu_{11}, \mu_{12}, \mu_{13}, \mu_{21}, \mu_{22}, \mu_{23})$ contains the genotype-specific maternal effects, and $\theta = (\sigma^2_{m_1}, \sigma_{m_{12}}, \sigma^2_{m_2}, \sigma^2_{f_1}, \sigma_{f_{12}}, \sigma^2_{f2}, \sigma^2_{mf_1}, \sigma_{mf_1},$ contains $\sigma^2_{mf2}, \sigma^2_{g_1}, \sigma_{g_{12}}, \sigma^2_{g_2}, \sigma^2_{e_1}, \sigma_{e_{12}}, \sigma^2_{e_2})$ different random variance components. The parameters can be estimated with either maximum likelihood (ML) method or the restricted maximum likelihood (REML) method. In our previous investigation, we did a comprehensive comparison of the two methods in estimating variance components based on a diploid mapping population (15). The results indicated that the ML method is faster than the REML method, but the REML method gives less biased results, which is consistent with the work of Corbeil and Searle (23). The less biased results make the REML estimation method more attractive. In the following, we briefly outline the REML estimation procedure and

more details about the REML algorithm can be found in the Appendix.

All data in K families form one big vector denoted as $\mathbf{y}$ with dimension $N \times 1$ $(N = 2\sum_{k=1}^{K} n_k)$. The vector $\mathbf{y}$ can be further partitioned into three vectors $\mathbf{y} = (y_1, y_2, y_3)^T$, where phenotype y_1 corresponding to families with maternal genotypes AA, y_2 corresponding to families with maternal genotypes Aa and y_3 corresponding to families with maternal genotypes aa in different backcross families. Similarly, β can be expressed as $\beta = \{\beta_1, \beta_2, \beta_3\} = \left\{ \begin{pmatrix} \mu_{11} \\ \mu_{21} \end{pmatrix}, \begin{pmatrix} \mu_{12} \\ \mu_{22} \end{pmatrix}, \begin{pmatrix} \mu_{13} \\ \mu_{23} \end{pmatrix} \right\}$. With this partition, the REML log-likelihood function can be expressed as

$$\ell^*(\Omega) = \sum_{r=1}^{3} \log[f(y_r | \Omega)]$$
$$\propto -\frac{1}{2}\sum_{r=1}^{3}\left\{ \log|\Sigma_r| + \log(|X_r' \Sigma_r^{-1} X_r|) + y_r' P_r y_r \right\}$$

Equation (4)

where $y_1 \sim N\left(\begin{pmatrix} \mu_{11} 1_{\sum_{k=1}^{l_1} n_k} \\ \mu_{21} 1_{\sum_{k=1}^{l_1} n_k} \end{pmatrix}, \Sigma_1 = \begin{pmatrix} \Sigma_1 & & \\ & \cdots & \\ & & \Sigma_{l_1} \end{pmatrix} \right)$;

$y_2 \sim N\left(\begin{pmatrix} \mu_{12} 1_{\sum_{k=l_1+1}^{l_1+l_2} n_k} \\ \mu_{22} 1_{\sum_{k=l_1+1}^{l_1+l_2} n_k} \end{pmatrix}, \Sigma_1 = \begin{pmatrix} \Sigma_{l_1+1} & & \\ & \cdots & \\ & & \Sigma_{l_1+l_2} \end{pmatrix} \right)$;

$y_3 \sim N\left(\begin{pmatrix} \mu_{13} 1_{\sum_{k=l_1+l_2+1}^{K} n_k} \\ \mu_{23} 1_{\sum_{k=l_1+l_2+1}^{K} n_k} \end{pmatrix}, \Sigma_1 = \begin{pmatrix} \Sigma_{l_1+l_2+1} & & \\ & \cdots & \\ & & \Sigma_{K} \end{pmatrix} \right)$;

$l_1 + l_2 + l_3 = K$; l_r $(r = 1, 2, 3)$ denotes the number of families generated from the backcross with maternal genotype AA $(r = 1)$, Aa $(r = 2)$ and aa $(r = 3)$; Σ_r's are block diagonal matrices; and $P_r = \Sigma_r^{-1} - \Sigma_r^{-1} X_r (X_r' \Sigma_r^{-1} X_r)^{-1} X_r' \Sigma_r^{-1}$. Then the Fisher scoring algorithm can be derived for parameter estimation. The details are given in the Appendix.

3.3. Hypothesis testing

We propose to search for iQTLs across the genome by assuming a putative QTL every 1 or $2cM$ by partitioning the whole linkage map into small intervals. At each putative QTL position, we test if there is a significant QTL effect on the bivariate traits by formulating the following hypotheses

$$\begin{cases} H_0: \sigma_{m_1}^2 = \sigma_{m_2}^2 = \sigma_{m_{12}} = \sigma_{f_1}^2 = \sigma_{f_2}^2 = \sigma_{f_{12}} \\ \quad = \sigma_{mf_1}^2 = \sigma_{mf_2}^2 = \sigma_{mf_{12}} = 0 \\ H_1: H_0 \text{ is not true} \end{cases}$$

Hypothesis (5)

The significance of the above test is assessed through the likelihood ratio test (LRT). Let $\tilde{\Omega}$ and $\hat{\Omega}$ be estimates of the unknown parameters under H_0 and H_1, respectively, then the likelihood ratio statistic is evaluated by

$$LR = -2[\log L(\tilde{\Omega}|y) - \log L(\hat{\Omega}|y)]$$

Equation (6)

which, under the null hypothesis, is distributed as a mixture chi-square distribution with the form

$$\frac{\binom{6}{0}}{2^6}\chi_9^2 : \frac{\binom{6}{1}}{2^6}\chi_7^2 : \frac{\frac{1}{5}\binom{6}{2}}{2^6}\chi_6^2 : \frac{\frac{4}{5}\binom{6}{2}}{2^6}\chi_5^2 : \frac{\frac{3}{5}\binom{6}{3}}{2^6}\chi_4^2 :$$ (24). Due to

$$\frac{\frac{2}{5}\binom{6}{3} + \frac{1}{5}\binom{6}{4}}{2^6}\chi_3^2 : \frac{\frac{4}{5}\binom{6}{4}}{2^6}\chi_2^2 : \frac{\binom{6}{5}}{2^6}\chi_1^2 : \frac{\binom{6}{6}}{2^6}\chi_0^2$$

correlations of tests across the genome, permutations can be applied to determine the genome-wide significance threshold.

Once a QTL is identified at a genomic position, its imprinting property for both phenotypic traits is assessed by the following imprinting hypothesis

$$\begin{cases} H_0: \sigma_{f_1}^2 = \sigma_{m_1}^2, \sigma_{f_2}^2 = \sigma_{m_2}^2, \sigma_{f_{12}} = \sigma_{m_{12}} \\ H_1: H_0 \text{ is not true} \end{cases}$$

Hypothesis (7)

Again, the likelihood ratio test is applied and the test statistic (denoted as LR_{imp}) asymptotically follows a chi-square distribution with 3 degrees of freedom. If the null is rejected at the tested QTL position, the QTL is declared as an iQTL. We can further assess whether the imprinting effect is due to complete silence of the maternal allele by testing

$$\begin{cases} H_0: \sigma_{m_1}^2 = \sigma_{m_2}^2 = \sigma_{m_{12}} = 0 \\ H_1: H_0 \text{ is not true} \end{cases}$$

Hypothesis (8)

or due to complete silence of the paternal allele by testing

$$\begin{cases} H_0: \sigma_{f_1}^2 = \sigma_{f_2}^2 = \sigma_{f_{12}} = 0 \\ H_1: H_0 \text{ is not true} \end{cases}$$

The likelihood ratio test statistic (denoted as LR_{imp_m} and LR_{imp_f}) corresponding to the above two tests follows a mixture chi-square distribution with

$$\frac{\pi - \cos^{-1}\rho_{12}}{2\pi}\chi_3^2 : \frac{1}{2}\chi_1^2 : \frac{\cos^{-1}\rho_{12}}{2\pi}\chi_0^2 \quad (25).$$

We can also test whether a QTL controls trait 1 by testing

$$\begin{cases} H_0 : \sigma_{m_1}^2 = \sigma_{f_1}^2 = \sigma_{mf_1} = 0 \\ H_1 : H_0 \text{ is not true} \end{cases}$$

Hypothesis (9) or controls trait 2 by testing

$$\begin{cases} H_0 : \sigma_{m_2}^2 = \sigma_{f_2}^2 = \sigma_{mf_2} = 0 \\ H_1 : H_0 \text{ is not true} \end{cases}$$

Hypothesis (10)

The likelihood ratio statistic corresponding to the above tests is denoted as LR_{pleio}^j ($j = 1, 2$) which under the null asymptotically follows a mixture chi-square distribution with the form

$$\frac{1}{4\pi}[2\pi - \cos^{-1}\rho_{12} - \cos^{-1}\rho_{13} - \cos^{-1}\rho_{23}]\chi_3^2 :$$

$$\frac{1}{4\pi}[3\pi - \cos^{-1}\rho_{12|3} - \cos^{-1}\rho_{13|2} - \cos^{-1}\rho_{23|1}]\chi_2^2 :$$

$$\frac{1}{4\pi}[\cos^{-1}\rho_{12} + \cos^{-1}\rho_{13} + \cos^{-1}\rho_{23}]\chi_1^2 :$$

$$[\frac{1}{2} - \frac{1}{4\pi}[3\pi - \cos^{-1}\rho_{12|3} - \cos^{-1}\rho_{13|2} - \cos^{-1}\rho_{23|1}]\chi_0^2$$

where ρ_{rs} refers to the correlation between the variance terms r and s ($r, s = 1, 2, 3$) which is calculated from the Fisher information matrix, and the conditional correlation is defined as $\rho_{rs|t} = \dfrac{(\rho_{rs} - \rho_{rt}\rho_{st})}{\left(1 - \rho_{rt}^2\right)^{1/2}\left(1 - \rho_{st}^2\right)^{1/2}}$. The detailed derivation can be found in Li and Cui (25).

Rejection of the null of the above two tests indicates the pleiotropic effect (i.e., one gene acts on two traits). But if two genes are closely linked at the detected (i)QTL (i.e., close linkage), the pleiotropic effect might be a false positive due to close linkage. Thus, it is essential to distinguish a pleiotropic effect vs close linkage. This is exactly the relative advantage of the multi-trait linkage analysis. To further distinguish close linkage against pleiotropic effect, we develop the following two tests

$$\begin{cases} H_0 : \rho_{m_{12}} = \rho_{f_{12}} = \rho_{mf_{12}} = 1 \\ H_1 : H_0 \text{ is not true} \end{cases}$$

Hypothesis (11) for testing pleiotropic effect and

$$\begin{cases} H_0 : \rho_{m_{12}} = \rho_{f_{12}} = \rho_{mf_{12}} = 0 \\ H_1 : H_0 \text{ is not true} \end{cases}$$

Hypothesis (12)

for testing close linkage, where ρ.'s are genetic correlation measures for different variance components between two traits. The null hypothesis in test Hypothesis (11) indicates that the additive effects for two traits are perfectly correlated and two traits are possibly controlled by a single gene. On the contrary, the null hypothesis in test Hypothesis (12) indicates two closely linkage genes at one (i)QTL location. The likelihood ratio test is denoted by LR_p for test Hypothesis (11) and LR_{co-in} for test Hypothesis (12). The null distribution of LR_p has a mixture chi-square distribution (since 1 is a boundary point for correlation ρ) with the form

$$\frac{1}{4\pi}[2\pi - \cos^{-1}\rho_{12} - \cos^{-1}\rho_{13} - \cos^{-1}\rho_{23}]\chi_3^2 :$$

$$\frac{1}{4\pi}[3\pi - \cos^{-1}\rho_{12|3} - \cos^{-1}\rho_{13|2} - \cos^{-1}\rho_{23|1}]\chi_2^2 : \quad , \text{ while}$$

$$\frac{1}{4\pi}[\cos^{-1}\rho_{12} + \cos^{-1}\rho_{13} + \cos^{-1}\rho_{23}]\chi_1^2 :$$

$$[\frac{1}{2} - \frac{1}{4\pi}[3\pi - \cos^{-1}\rho_{12|3} - \cos^{-1}\rho_{13|2} - \cos^{-1}\rho_{23|1}]\chi_0^2$$

the null distribution of LR_{co-in} follows a regular chi-square distribution with 3 degrees of freedom, i.e., $LR_{co-in} \sim \chi_3^2$ since 0 is not a boundary point. Note that the assessment of pleiotropic effects vs close linkage only occurs at the genomic location where there is an (i)QTL being identified by the overall genetic test.

4. SIMULATION

4.1. Simulation design

We designed a simulation study to evaluate the performance of the joint analysis as well as the effect of different genetic designs on testing power and parameter estimation. Six equally-spaced markers (M_1-M_6) were simulated for one linkage group assuming a backcross design. This linkage group covers a length of 100cM. Haldane map function was used to convert map distance to recombination rate. Assume there was one QTL located at 48cM away from the first marker which had effects on two phenotypic traits. Phenotypic values of two traits were generated from a multivariate normal distribution with variance-covariance specified in Equation (2). Parameter values used for simulation are given in Table 2-3.

As described in Li and Cui (16), different combinations of family and offspring size could influence testing power and parameter estimation. To mimic the real data, we fixed the total sample size as 400 and considered two designs, i.e., 4 families with 100 offsprings each (denoted as 4×100) and 20 families with 20 offsprings each (denoted as 20×20). In each simulation scenario, the IBD value of any two siblings was calculated at every 2cM

Table 1. The power, QTL position and variance components parameter estimates based on 100 simulation replicates for data simulated assuming a Mendelian QTL with no imprinting effect under the 4×100 design. True values used for simulation studies are indicated by <>. The standard errors of the parameter estimates are given in parentheses

Trait	Position	$\sigma^2_{m_1}$	$\sigma^2_{f_1}$	$\sigma^2_{mf_1}$	$\sigma^2_{m_2}$	$\sigma^2_{f_2}$	$\sigma^2_{mf_2}$	$\sigma^2_{g_1}$	$\sigma^2_{g_2}$	$\sigma^2_{e_1}$	$\sigma^2_{e_2}$	Power	Type I Error
	<48cM>	<0.75?>	<0.75>	<0.5>	<0.375>	<0.375>	<0.25>	<1.5>	<0.75>	<3>	<1.5>		
T1+T2*	47.82	0.56	0.55	0.80	0.44	0.41	0.52	1.91	0.97	3.04	1.50	0.76	0.06
(4×100)	(10.04)	(0.27)	(0.35)	(1.17)	(0.18)	(0.21)	(0.56)	(1.28)	(0.57)	(0.51)	(0.23)	-	-
T1	44.66	0.30	0.50	0.87	-	-	-	2.04	-	2.88	-	0.70	0.10
	(18.28)	(0.45)	(2.32)	(1.49)	-	-	-	(1.23)	-	(0.61)	-	-	-
T2	46.78	-	-	-	0.20	0.35	0.65	-	0.95	-	1.50	0.60	0.12
	(15.02)	-	-	-	(0.32)	(1.01)	(1.65)	-	(0.56)	-	(0.25)	-	-
T1+T2	47.36	0.72	0.71	0.71	0.44	0.45	0.43	1.56	0.78	3.04	1.50	0.78	0.03
(20×20)	(10.36)	(0.42)	(0.59)	(0.66)	(0.25)	(0.31)	(0.31)	(0.77)	(0.34)	(0.39)	(0.17)	-	-

*T1 and T2 refer to trait 1 and trait 2. The true QTL is located at 48cM. Power is calculated based on Hypothesis (5) for testing the overall QTL effect. Type I error is calculated based on Hypothesis (7) for testing the imprinting effect.
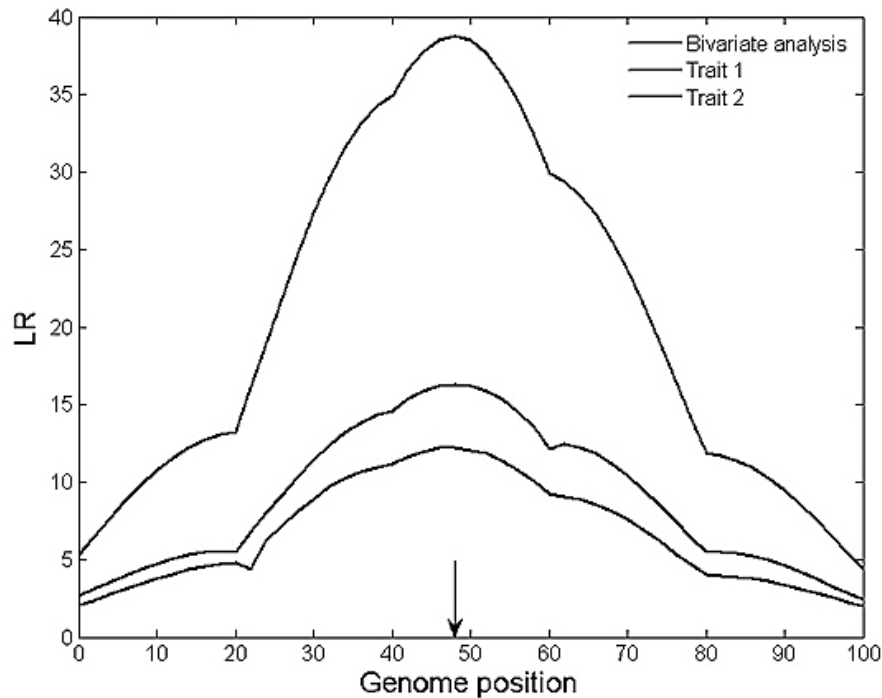

Figure 1. The LR profile plot averaged over 100 replicates. The solid, dotted and dash lines correspond to LR profiles with the analysis considering both traits, trait 1 and trait 2, respectively. The true QTL position is indicated by the arrow sign (48cM from marker 1).

along the linkage group using the approach described in Li and Cui (16). The REML method was adopted to estimate unknown parameters of interests, and 100 simulation replicates were recorded.

4.2. Simulation results

Figure 1 plots the likelihood ratio statistic averaged over 100 simulation replicates across the entire linkage group obtained with the proposed multivariate analysis and single trait analysis. The bivariate analysis shows consistently high LR values across the linkage group with a clear peak corresponding to the simulated QTL position, which implies the potential power gain in QTL identification with the joint analysis. This is further confirmed by the power analysis in Table 1 (82% power for

the bivariate analysis vs 70% and 60% for each single trait analysis) with data simulated assuming no imprinting effect (i.e., $\sigma^2_{m_1} = \sigma^2_{f_1}$ and $\sigma^2_{m_2} = \sigma^2_{f_2}$).

The parameter estimation, power of QTL identification and type I error of the imprinting test are summarized in Table 1. Overall, the joint analysis gives larger statistical power (82%), smaller type I error rate for the imprinting test (6%) and less biased variance components estimation compared to the results obtained with the single trait analysis. We also noticed that the joint analysis gives less bias and smaller standard error in QTL location estimation compared to the single trait analysis. In addition, the joint analysis greatly improved the false

Table 2. The power, REMLs of the QTL position and effect parameters estimated based on 100 simulation replicates for a QTL showing different imprinting effects under the 20×20 sampling design. True values used for simulation studies are indicated by <>. The standard errors of the parameters are given in parentheses

Position	$\sigma_{m_1}^2$	$\sigma_{f_1}^2$	$\sigma_{m_2}^2$	$\sigma_{f_2}^2$	σ_{mf_1}	$\sigma_{mf_{12}}$	$\sigma_{mf_2}^2$	$\sigma_{g_1}^2$	$\sigma_{g_{12}}$	$\sigma_{g_2}^2$	$\sigma_{e_1}^2$	$\sigma_{e_{12}}$	$\sigma_{e_2}^2$	Power[1]	Power[2]	ipower[1]
<48cM>					<0.5>	<0.3>	<0.25>	<1.5>	<0.9>	<0.75>	<3>	<1.8>	<1.5>			
	<1.5>	<0>	<0.75>	<0>												
46.46	1.16	0.33	0.73	0.29	0.68	0.30	0.50	2.07	1.26	1.04	2.90	1.74	1.46	0.87	0.87	0.34
(10.59)	(0.61)	(0.29)	(0.32)	(0.22)	(0.61)	(0.50)	(0.32)	(0.83)	(0.52)	(0.39)	(0.38)	(0.24)	(0.17)			
	<0>	<1.5>	<0>	<0.75>												
50.52	0.17	1.10	0.14	0.65	0.66	0.36	0.40	1.46	0.87	0.71	3.08	1.85	1.54	0.65	0.69	0.43
(27.39)	(0.18)	(0.97)	(0.14)	(0.52)	(0.59)	(0.43)	(0.30)	(0.63)	(0.39)	(0.30)	(0.32)	(0.22)	(0.17)			
	<1>	<0.5>	<0.5>	<0.25>												
46.10	0.92	0.69	0.54	0.49	0.74	0.36	0.48	1.63	0.98	0.83	3.02	1.81	1.50	0.78	0.82	0.13
(12.52)	(0.44)	(0.69)	(0.29)	(0.31)	(0.66)	(0.47)	(0.32)	(0.80)	(0.51)	(0.37)	(0.40)	(0.26)	(0.19)			
	<0.5>	<1>	<0.25>	<0.5>												
48.34	0.56	0.95	0.32	0.55	0.86	0.41	0.45	1.59	0.96	0.79	2.95	1.78	1.50	0.68	0.68	0.25
(18.83)	(0.42)	(0.82)	(0.24)	(0.42)	(0.77)	(0.54)	(0.41)	(0.68)	(0.42)	(0.32)	(0.34)	(0.22)	(0.17)			

Power[1] and Power[2] refer to the overall QTL detection powers calculated based on the empirical and the theoretical test; iPower[1] refers to the imprinting power for testing $H_0: \sigma_{f_1}^2 = \sigma_{m_1}^2, \sigma_{f_2}^2 = \sigma_{m_2}^2, \sigma_{f_{12}} = \sigma_{m_{12}}$. The imprinting test is only conducted at the position where Hypothesis (5) is rejected.

Table 3. The estimated parameters for the two endosperm traits: mean ploidy level (Ploid) and percent of the endoreduplicated nuclei (Endo)

Ch	QTL Position	$\sigma_{m_1}^2$	$\sigma_{m_{12}}$	$\sigma_{m_2}^2$	$\sigma_{f_1}^2$	$\sigma_{f_{12}}$	$\sigma_{f_2}^2$	$\sigma_{mf_1}^2$	$\sigma_{mf_{12}}$	$\sigma_{mf_2}^2$	p_{imp}	p_{imp_m}	p_{imp_f}	p_p	p_{co-in}
2[*]	bnlg	0.19	0.31	0.96	0.04	-0.16	0.76	0.09	-0.29	1.97	0.15	-	-	0.88	0.95
4[*]	umc+5.62cM	0.70	1.23	2.48	0.32	0.59	1.10	0.01	0.06	1.30	0.02	0.56	≈ 0	0.03	0.09
6[*]	bnlg+23.49cM	0.30	0.60	2.10	0.43	1.11	2.84	0.12	-0.51	4.97	0.04	0.03	0.88	≈ 0	0.09
7	dupssr	0.27	0.38	0.52	0.28	0.50	1.05	0.01	-0.05	0.20	0.99	-	-	0.88	0.73
9[*]	umc+10cM	0.31	0.21	0.19	0.30	0.42	1.73	0.32	-0.03	≈ 0	0.26	-	-	0.88	0.64
10[*]	MMC+8cM	0.38	0.62	1.07	0.32	0.68	1.43	0.01	-0.08	0.66	0.16	-	-	0.07	0.01

Note: Six QTLs for joint traits Ploi and Endo are located at chromosome 2, 4, 6, 7, 9 and 10. QTLs showing significance at the genome-wide level are indicated by *. p_{imp}, p_{imp_m}, p_{imp_f}, p_p and p_{co-in} are the p-values for testing imprinting effect ($H_0: \sigma_{m_1}^2 = \sigma_{f_1}^2, \sigma_{m_2}^2 = \sigma_{f_2}^2, \sigma_{m_{12}} = \sigma_{f_{12}}$), complete maternal imprinting ($H_0: \sigma_{m_1}^2 = \sigma_{m_2}^2 = \sigma_{m_{12}} = 0$), complete paternal ($H_0: \sigma_{f_1}^2 = \sigma_{f_2}^2 = \sigma_{f_{12}} = 0$), pleiotropic effect ($H_0: \rho_{m_{12}} = \rho_{f_{12}} = \rho_{mf_{12}} = 1$) and co-incidence linkage ($H_0: \rho_{m_{12}} = \rho_{f_{12}} = \rho_{mf_{12}} = 0$), respectively.

positive control with the 4×100 design. The simulation results indicate that the joint mapping by incorporating bivariate phenotypic information greatly improves the efficiency of QTL detection.

The above simulation results were based on a 4×100 design. Our previous investigation showed that this design is less powerful for single trait analysis compared to the 20×20 design when the sample size is fixed (16). Thus, we did additional simulation under the 20×20 design. The results showed improved power for QTL detection and also improved precision for parameter estimation (Table 2). In real application, our recommendation is to adopt a balanced design with a reasonably large family size.

Focusing on the 20×20 design, we did additional simulation to evaluate the performance of imprinting analysis for the joint model under different imprinting mechanisms. The results are tabulated in Table 3. Four imprinting mechanisms were considered: complete paternal imprinting (i.e., $\sigma_{m_1}^2 = 1.5, \sigma_{m_2}^2 = 0.75$ and $\sigma_{f_1}^2 = \sigma_{f_2}^2 = 0$), complete maternal imprinting (i.e., $\sigma_{m_1}^2 = \sigma_{m2}^2 = 0$ and $\sigma_{f_1}^2 = 1.5, \sigma_{f_2}^2 = 0.75$), partial paternal imprinting (i.e., $\sigma_{m_1}^2 = 1, \sigma_{m_2}^2 = 0.5$ and $\sigma_{f_1}^2 = 0.5, \sigma_{f_2}^2 = 0.25$), and partial maternal imprinting (i.e., $\sigma_{m_1}^2 = 0.5, \sigma_{m_2}^2 = 0.25$ and $\sigma_{f_1}^2 = 1, \sigma_{f_2}^2 = 0.5$).

The results showed that the complete maternal imprinting case has the lowest power (65%). This is due to the fact that a backcross endosperm genome carries two maternal copies. When there is no genetic effect for the maternal copies, we expect the power to be lower. We also realized the overall relatively low imprinting power. By increasing the sample (e.g., double the total size to 800), we did find largely improved imprinting power (data not shown). The simulation study implies that large samples are always desirable to obtain reasonably large imprinting power, which is feasible for an experimental cross, especially for plants. For the proposed backcross design study, more than 1000 samples are needed in order to achieve good imprinting power, preferably with a 40×25 design.
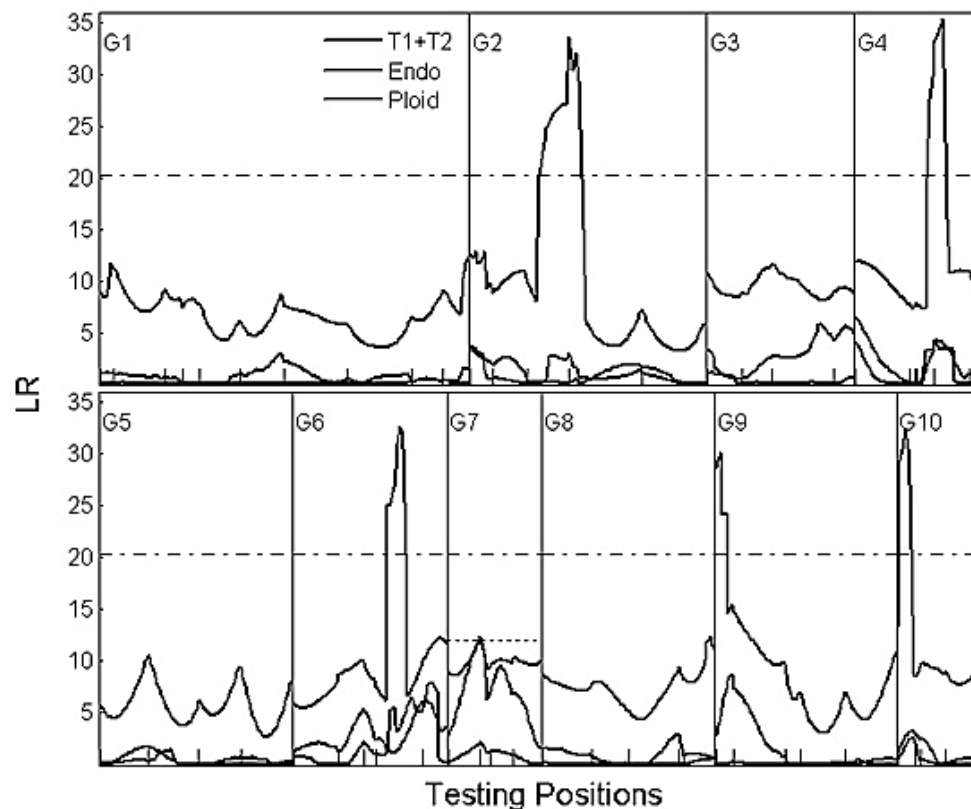
Figure 2. The profile of the log-likelihood ratios (LR) for testing the existence of QTLs underlying the two endosperm traits across the 10 maize linkage groups ($G_1,...,G_{10}$).

5. REAL DATA ANALYSIS

Endosperm is created from the fusion of two polar nuclei and a sperm cell, resulting in a triploid tissue with two identical chromosomes inherited from the maternal parent and one chromosome from the paternal parent. The endosperm supplies nutrition to the embryo during seed germination, and serves as a major source of food for human beings (26). The function of the endosperm is far more complicated than we understand, and is beyond simple nutrient delivery to the embryo. As mentioned in the introduction section, maize endosperm cells undergoing endoreduplication are generally larger than other cells (20). Recent empirical study revealed that endoreduplication in maize endosperm may be associated with genes showing parent-of-origin effects (18).

A study was recently launched to identify parent-of-origin effects associated with endoreduplication in maize endosperm (17). Two inbred lines (Sg18 and Mo17) were used to generate the F_1 population. Four backcross segregation populations were then generated following a reciprocal backcross design, i.e., Sg18 $\times$ F_1, Mo17 $\times$ F_1, $F_1 \times$ Sg18, and $F_1 \times$ Mo17, with each one containing roughly around 90 offspring. Two endosperm traits, % of endoreduplication (denoted as Endo) and mean ploidy level (denoted as Ploid) were measured for each of the backcross

offspring, and 10 linkage groups were constructed from the observed marker data. More details about the description of the data can be found in Coelho *et al.* (17). The data set was previously analyzed with a single trait variance components model and significant imprinting genes were identified (16).

The sample correlation between the two traits is 0.812 after combining the four backcross phenotype data together, which indicates the potential benefit of a joint analysis by using the proposed bivariate variance components model. To enhance the power of iQTL mapping and further identify the pleiotropic iQTL effect, we applied our newly derived method to this data set. Figure 2 shows the LR profile plot across the ten linkage groups. In addition to the results based on the joint analysis (solid curve), the single trait analysis results are also plotted (dotted line for trait Endo and dashed line for trait Ploid). The 5% genome-wide significance threshold corresponding to the joint analysis (horizontal dotted lines) was determined by permutations based on repeatedly shuffling the relationship between marker genotypes and phenotypes (27). The horizontal dotted line in G7 indicates the 5% chromosome-wide permutation threshold. Note that the single trait analysis was done without correcting for the neighborhood QTL effect, hence the LR curves obtained in Li and Cui are different from the ones presented in this paper (16).

As shown in the figure, five QTLs were detected at the 5% genome-wide significance level and are located in linkage groups G2, G4, G6, G9 and G10. One QTL in G7 only passes the chromosome-wide threshold and is a suggestive QTL. Compared to the single trait analysis results, more QTLs were detected by the joint analysis. Table 3 lists the estimates of the QTL position, variance components, and p-values for testing overall QTL effect and various imprinting and pleiotropic effects. With the joint analysis, the identified QTLs in G4 and G6 are imprinted with imprinting p-values 0.02, and 0.04 respectively. Further tests show that the QTL in G4 shows paternal expression ($p_{imp_f} \approx 0$), and the QTL on G6 shows maternal expression ($p_{imp_m} = 0.03$). The other QTLs do not show an imprinting effect. We further evaluated the pleiotropic vs close linkage effect of the six QTLs. The results indicated that two iQTLs in G4 and G6 show significant co-incident linkage ($p_{co-in} > 0.05$ and $p_p < 0.05$), while the QTL in G10 shows a pleiotropic effect ($p_{co-in} < 0.05$ and $p_p > 0.05$). The p-values for the other QTLs are all larger than 0.05, hence no conclusion about pleiotropic or close linkage effect can be made. This might be due to the issue of genetic design and small sample sizes. The LR profile plot in Figure 2 indicates the power of the joint analysis over the single trait analysis. In addition to the increased power for QTL identification, we were also able to test the pleiotropic effect of (i)QTLs. The (i)QTLs shown pleiotropic effects should be paid special attention for follow-up functional validation.

6. DISCUSSION

A number of studies have shown that for correlated traits, multivariate approaches for genetic linkage analysis can increase the power and precision to identify genetic effects, especially when a QTL has a pleiotropic effect on several traits (5, 6). Considering the importance of imprinted genes in endosperm development and the relative merit of multi-trait analysis, we developed a bivariate variance components model based on a reciprocal backcross design to identify iQTLs while incorporating the special genetic make-up of the triploid inbreeding population. Simulation studies showed the performance of the method under different sampling designs with finite sample size. Comparing the results of joint analysis with those of single trait analysis, the joint analysis greatly improves the performance in QTL position estimation, testing power, and type I error rate.

We applied the joint model to a real data set with two endosperm traits, e.g., % of endoreduplication and mean ploidy level. Six QTLs were detected on G2, G4, G6, G7, G9, and G10 across the maize endosperm genome. Among the six QTLs, five showed genome-wide significance, two are iQTLs with maternal imprinting (on G4) and paternal imprinting (on G6). Compared with the single trait analysis, more QTLs were mapped in the joint analysis. In maize, several paternal imprinting genes have been identified. For example, there

is the r gene in the regulation of anthocyanin, the seed storage protein regulatory gene *dsrl*, the *MEA* gene in seed development and some α - *tubulin* genes (28-31). Study has shown that endoreduplication shows a maternally controlled parent-of-origin effect (18). Given that no specific gene has been reported to control endoreduplication, the identified iQTLs could serve to locate potential candidate genes for further functional validation.

In addition to mapping several QTLs, we also identified significant pleiotropic effects. In maize, some vital genes displaying pleiotropic effects have been reported. For example, maize *zfl* regulatory genes have pleiotropic effects on structure traits in branching and inflorescence formation (32). The *tb1* gene and its intergenic sequences illustrates the pleiotropic effects on maize morphology (33). A maize gene *GLOSSY1* (GL1) expresses its effect on trichome size and cutin structure during epidermis development (34). Given the high correlation between the two endoreduplication traits, the identification of genes with pleiotropic effects is practically important. Further functional verification is needed to confirm the findings of this investigation.

In the simulation study, the results indicate low power for imprinting detection with the 4×100 design. This result is consistent with the findings we found earlier for a single trait analysis (16). When we changed the design to 20×20 with a fixed total sample size, improved results were observed (Table 3). Even with the extremely unbalanced design (4×100), simulations also show reasonable false imprinting detection rate (6%). The real data setting is quite similar to the simulation design, thus the detected imprinting effects should have little chance to be false positives. Overall, the simulation studies provide practical guidance to real experimental designs: try to maintain a balanced sampling design and avoid extremely small families with large offspring and extremely large families with very small offspring.

The current method was derived for single iQTL mapping. Extension to multiple iQTL mapping is in fact not straightforward. Further investigation is needed in this context. On the other hand, even though the method was developed for experimental crosses, extension to human genetic mapping studies is straightforward. The only modification is the IBD sharing probability of sibpairs, where the calculation should take the family structure into account (e.g., 9). In fact, the IBD calculation is simpler in humans, because the cross sharing probability reduces to zero for a natural population with random mating. We hope our method will shed light on human genetic mapping studies to identify imprinting effects with variance components models while considering multiple traits.

7. ACKNOWLEDGEMENT

8. APPENDIX

Derivation of Fisher scoring algorithm for REML estimation

Define the IBD sharing matrices corresponding to the three phenotypic vectors as

$$\Pi_{m|r}^{(1)} = \begin{pmatrix} \Pi_{m|r} & 0 \\ 0 & 0 \end{pmatrix}$$

with dimension

$$\sum_{k=1}^{l_1} n_k \times \sum_{k=1}^{l_1} n_k \text{ for matrix 0.}$$

$$\Pi_{m|r}^{(2)} = \begin{pmatrix} 0 & 0 \\ 0 & \Pi_{m|r} \end{pmatrix}$$

with dimension

$$\sum_{k=l_1+1}^{l_1+l_2} n_k \times \sum_{k=l_1+1}^{l_1+l_2} n_k \text{ for matrix 0.}$$

$$\Pi_{m|r}^{(3)} = \begin{pmatrix} 0 & \Pi_{m|r} \\ \Pi_{m|r} & 0 \end{pmatrix}$$

with dimension

$$\sum_{k=l_1+l_2+1}^{K} n_k \times \sum_{k=l_1+l_2+1}^{K} n_k \text{ for matrix 0.}$$

The IBD sharing matrices for $\Pi_{f|r}$, $\Pi_{mf|r}$, Φ_r, and I_r can be similarly defined and are denoted as $\Pi_{f|r}^s$, $\Pi_{m|r}^s$, Φ_r^s and I_r^s ($s, r = 1, 2, 3$; $l_1 + l_2 + l_3 = K$). Let $u = \left\{\dfrac{\partial \ell^*}{\partial \Omega}\right\}$ be the score vector of the first derivative of the log-likelihood function in (4) with respective to each variance component, i.e.,

$$\frac{\partial \ell^*}{\partial \sigma_{m_1}^2} = -\frac{1}{2}\sum_{r=1}^{3}\left(tr(\hat{P}_r \Pi_{m|r}^{(1)} - y_r^T \hat{P}_r \Pi_{m|r}^{(1)} \hat{P}_r y_r\right),$$

$$\frac{\partial \ell^*}{\partial \sigma_{m_2}^2} = -\frac{1}{2}\sum_{r=1}^{3}\left(tr(\hat{P}_r \Pi_{m|r}^{(2)} - y_r^T \hat{P}_r \Pi_{m|r}^{(2)} \hat{P}_r y_r\right),$$

$$\frac{\partial \ell^*}{\partial \sigma_{m_{12}}} = -\frac{1}{2}\sum_{r=1}^{3}\left(tr(\hat{P}_r \Pi_{m|r}^{(3)} - y_r^T \hat{P}_r \Pi_{m|r}^{(3)} \hat{P}_r y_r\right),$$

$$\frac{\partial \ell^*}{\partial \sigma_{f_1}^2} = -\frac{1}{2}\sum_{r=1}^{3}\left(tr(\hat{P}_r \Pi_{f|r}^{(1)} - y_r^T \hat{P}_r \Pi_{f|r}^{(1)} \hat{P}_r y_r\right),$$

$$\frac{\partial \ell^*}{\partial \sigma_{f_2}^2} = -\frac{1}{2}\sum_{r=1}^{3}\left(tr(\hat{P}_r \Pi_{f|r}^{(2)} - y_r^T \hat{P}_r \Pi_{f|r}^{(2)} \hat{P}_r y_r\right),$$

$$\frac{\partial \ell^*}{\partial \sigma_{f_{12}}} = -\frac{1}{2}\sum_{r=1}^{3}\left(tr(\hat{P}_r \Pi_{f|r}^{(3)} - y_r^T \hat{P}_r \Pi_{f|r}^{(3)} \hat{P}_r y_r\right),$$

$$\frac{\partial \ell^*}{\partial \sigma_{mf_1}^2} = -\frac{1}{2}\sum_{r=1}^{3}\left(tr(\hat{P}_r \Pi_{mf|r}^{(1)} - y_r^T \hat{P}_r \Pi_{mf|r}^{(1)} \hat{P}_r y_r\right),$$

$$\frac{\partial \ell^*}{\partial \sigma_{mf_2}^2} = -\frac{1}{2}\sum_{r=1}^{3}\left(tr(\hat{P}_r \Pi_{mf|r}^{(2)} - y_r^T \hat{P}_r \Pi_{mf|r}^{(2)} \hat{P}_r y_r\right),$$

$$\frac{\partial \ell^*}{\partial \sigma_{mf_{12}}} = -\frac{1}{2}\sum_{r=1}^{3}\left(tr(\hat{P}_r \Pi_{mf|r}^{(3)} - y_r^T \hat{P}_r \Pi_{mf|r}^{(3)} \hat{P}_r y_r\right),$$

$$\frac{\partial \ell^*}{\partial \sigma_{g_1}^2} = -\frac{1}{2}\sum_{r=1}^{3}\left(tr(\hat{P}_r \Phi_r^{(1)} - y_r^T \hat{P}_r \Phi_r^{(1)} \hat{P}_r y_r\right),$$

$$\frac{\partial \ell^*}{\partial \sigma_{g_2}^2} = -\frac{1}{2}\sum_{r=1}^{3}\left(tr(\hat{P}_r \Phi_r^{(2)} - y_r^T \hat{P}_r \Phi_r^{(2)} \hat{P}_r y_r\right),$$

$$\frac{\partial \ell^*}{\partial \sigma_{g_{12}}} = -\frac{1}{2}\sum_{r=1}^{3}\left(tr(\hat{P}_r \Phi_r^{(3)} - y_r^T \hat{P}_r \Phi_r^{(3)} \hat{P}_r y_r\right),$$

$$\frac{\partial \ell^*}{\partial \sigma_{e_1}^2} = -\frac{1}{2}\sum_{r=1}^{3}\left(tr(\hat{P}_r I_r^{(1)} - y_r^T \hat{P}_r I_r^{(1)} \hat{P}_r y_r\right),$$

$$\frac{\partial \ell^*}{\partial \sigma_{e_2}^2} = -\frac{1}{2}\sum_{r=1}^{3}\left(tr(\hat{P}_r I_r^{(2)} - y_r^T \hat{P}_r I_r^{(2)} \hat{P}_r y_r\right),$$

$$\frac{\partial \ell^*}{\partial \sigma_{e_{12}}} = -\frac{1}{2}\sum_{r=1}^{3}\left(tr(\hat{P}_r I_r^{(3)} - y_r^T \hat{P}_r I_r^{(3)} \hat{P}_r y_r\right),$$

Let $H = \left\{\dfrac{\partial \ell^{*2}}{\partial \Omega_s \partial \Omega_t}\right\}$ be the Hessian matrix (The matrix is too large to be included and is omitted here). The Fisher information matrix, $I(\Omega)$, in the REML procedure is obtained by taking the expectation of the negative Hessian matrix. The algorithm starts with $\Omega^{(t+1)} = \Omega^{(t)} + I^{-1}(\Omega^{(t)})u^{(t)}$. Given initial values $\Omega^{(0)}$, the iteration starts and stops until converges. Upon convergence, the REML estimator of β is just the generalized least squares estimator, that is,

$$\hat{\beta} = \left(X^T \hat{\Sigma}^{-1} X\right)^{-1} X^T \hat{\Sigma}^{-1} y$$

9. REFERENCES

1. C Li, A Zhou, T Sang: Rice domestication by reducing shattering. *Science* 311, 1936-1939 (2006)

2. J Bergelson, F Roux: Towards identifying genes underlying ecologically relevant traits in Arabidopsis thaliana. *Nat Rev Genet* 11, 867-879 (2010)

3. JT Williams, PV Eerdewegh, L Almasy, J Blangero: Joint multipoint linkage analysis of multivariate qualitative and quantitative traits. I. likelihood formulation and simulation results. *Am J Hum Genet* 65, 1134-1147 (1999)

4. L Almasy, J Blangero: Multipoint quantitative-trait linkage analysis in general pedigrees. *Am J Hum Genet* 62, 1198-1211 (1998)

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

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

7. K Pfeifer: Mechanisms of genomic imprinting. *Am J Hum Genet* 67, 777-787 (2000)

8. IM Morison, JP Ramsay, HG Spencer: A census of mammalian imprinting. *Trends Genet* 21, 457-465 (2005)

9. RL Hanson, S Kobes, RS Lindsay, WC Kmowler: Assessment of parent-of-origin effects in linkage analysis of quantitative traits. *Am J Hum Genet* 68, 951-962 (2001)

10. D-J de Koning, H Bovenhuis, JAM van Arendonk: On the detection of imprinted quantitative trait loci in experimental crosses of outbred species. *Genetics* 161, 931-938 (2002)

11. Y Cui: A statistical framework for genome-wide scanning and testing imprinted quantitative trait loci. *J Theo Biol* 244, 115-126 (2007)

12. Y Cui, JM Cheverud, R Wu: A statistical model for dissecting genomic imprinting through genetic mapping. *Genetica* 130, 227-239 (2007)

13. T Liu, RJ Todhunter, S Wu, W Hou, R Mateescu, Z Zhang, NI Burton-Wurster, GM Acland, G Lust, R Wu: A random model for mapping imprinted quantitative trait loci in a structured pedigree: An implication for mapping canine hip dysplasia. *Genomics* 90, 276-284 (2007)

14. Y Li, CM Coelho, T Liu, S Wu, J Wu, Y Zeng, Y Li, B Hunter, RA Dante, BA. Larkins, R Wu: A statistical strategy to estimate maternal-zygotic interactions and parent-of-origin effects of QTLs for seed development. *PLoS One* 3, e3131 (2008)

15. G Li, Y Cui: A statistical variance components framework for mapping imprinted quantitative trait loci in experimental crosses. *J Prob Stat* vol. 2009, Article ID 689489 (2009)

16. G Li, Y Cui: A general statistical framwork for dissecting parent-of-origin effects underlying endosperm traits in flowering plants. *Ann App Stat* 4, 1214-1233 (2010)

17. CM Coelho, S Wu, Y Li, B Hunter, RA Dante, Y Cui, R Wu, BA Larkins: Identification of quantitative trait loci that affect endoreduplication in maize endosperm. *Theor Appl Genet* 115, 1147-1162 (2007)

18. BP Dilkes, RA Dante, C Coelho, BA Larkins: Genetic analysis of endoreduplication in Zea mays endosperm: evidence of sporophytic and zygotic maternal control. *Genetics* 160, 1163-1177 (2002)

19. B Mangin, P Thoquet, N Grimsley: Pleiotropic QTL analysis. *Biometrics* 54, 88-99 (1998)

20. JP Grime, MA Mowforth: Variation in genome size: an ecological interpretation. *Nature* 299, 151-153 (1982)

21. M Abney, MS McPeek, C Ober: Estimation of variance components of quantitative traits in inbred populations. *Am J Hum Genet* 66, 629-650 (2000)

22. Y Cui, R Wu: A statistical model for characterizing epistatic control of triploid endosperm triggered by maternal and offspring QTL. *Genet Res* 86, 65-76 (2005)

23. RR Corbeil, SR Searle: A comparison of variance component estimators. *Biometrics* 32, 779-791 (1976)

24. SG Self, KY Liang: Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J Am Stat Asso* 82, 605-610 (1987)

25. G Li, Y Cui: Assessing statistical significance in genetic linkage analysis with the variance components model. Manuscript (2011)

26. RA Brink, D Cooper: The endosperm in seed development. *Bot Rev* 13, 423-541 (1947)

27. G Churchill, RW Doerge: Empirical threshold values for quantitative trait mapping. *Genetics* 138, 963-971 (1994)

28. JL Kermicle: Dependence of the R-mottled aleurone phenotype in maize on the modes of sexual transmission. Genetics 66, 69-85 (1970)

29. S Chaudhuri, J Messing: Allele-specific parental imprinting of dzrl, a post transcriptional regulator of zein accumulation. Proc Natl Acad Sci 91, 4867-4871 (1994)

30. T Kinoshita, R Yadegari, JJ Harada, RB Goldberg, RL Fishcher: Imprinting of the MEDEA polycomb gene in the Arabidopsis endosperm. Plant Cell 11, 1945-1952 (1999)

31. G Lund, J Messing, A Viotti: Endosperm-specific demethylation and activation of specific alleles of a-tubulin genes of Zea mays L. Mol Gen Genet 246, 716-722 (1995)

32. K Bomblies, JF Doebley: Pleiotropic effects of the duplicate maize FLORICAULA/LEAFY genes zfl1 and zfl2 on traits under selection during maize domestication. Genetics 172, 519-531 (2006)

33. RM Clark, TN Wagler, P Quijada, J Doebley: A distant upstream enhancer at the maize domestication gene tb1 has pleiotropic effects on plant and inflorescent architecture. *Nat Genet* 38, 594-597 (2006)

34. M Sturaro, H Hartings, E Schmelzer, R Velasco, F Salamini, M Motto: Cloning and characterization of GLOSSY1, a maize gene involved in cuticle membrane and wax production. *Plant Physiol* 138, 478-489 (2005)

Abbreviations: GWAS: genome-wide association studies, IBD: identical-by-descent, iQTL: imprinting quantitative trait loci, LR: likelihood ratio, QTL: quantitative trait loci, REML: restricted maximum likelihood, SNP: single nucleotide polymorphism

Key Words: Backcross, Close linkage, Likelihood ratio test, Maize, Maximum likelihood, Pleiotropic effect

Send correspondence to: Yuehua Cui, Department of Statistics and Probability, A-432 Wells Hall, Michigan State University, East Lansing, MI 48824, Tel: 517-432-7098, Fax: 517-432-1405, E-mail: cui@stt.msu.edu