*The Annals of Applied Statistics*
2010, Vol. 4, No. 3, 1214–1233
DOI: 10.1214/09-AOAS323


# A GENERAL STATISTICAL FRAMEWORK FOR DISSECTING PARENT-OF-ORIGIN EFFECTS UNDERLYING ENDOSPERM TRAITS IN FLOWERING PLANTS[1]

BY GENGXIN LI AND YUEHUA CUI

*Michigan State University*

Genomic imprinting has been thought to play an important role in seed development in flowering plants. Seed in a flowering plant normally contains diploid embryo and triploid endosperm. Empirical studies have shown that some economically important endosperm traits are genetically controlled by imprinted genes. However, the exact number and location of the imprinted genes are largely unknown due to the lack of efficient statistical mapping methods. Here we propose a general statistical variance components framework by utilizing the natural information of sex-specific allelic sharing among sibpairs in line crosses, to map imprinted quantitative trait loci (iQTL) underlying endosperm traits. We propose a new variance components partition method considering the unique characteristic of the triploid endosperm genome, and develop a restricted maximum likelihood estimation method in an interval scan for estimating and testing genome-wide iQTL effects. Cytoplasmic maternal effect which is thought to have primary influences on yield and grain quality is also considered when testing for genomic imprinting. Extension to multiple iQTL analysis is proposed. Asymptotic distribution of the likelihood ratio test for testing the variance components under irregular conditions are studied. Both simulation study and real data analysis indicate good performance and powerfulness of the developed approach.

**1. Introduction.** The life cycle of an angiosperm starts with the process of double fertilization, where the fertilization of the haploid egg with one sperm cell forms the embryo, and the fusion of the two polar nuclei with another sperm cell develops into endosperm [Chaudhury et al. (2001)]. Thus, endosperm is a tissue unique to angiosperm. The embryo and endosperm are genetically identical, except that the endosperm is triploid composed of one set of paternal and two identical sets of maternal chromosomes. In cereals, the endosperm of a grain is the major storage organ providing nutrition for early-stage seed development, and more than that, serves as the major source of food for human beings. The identification of important genes that underlie the variation of quantitative traits of various interests in endosperm is thus paramountly important.

Received August 2009; revised December 2009.

[1]Supported in part by NSF Grant DMS-0707031 and by Michigan State University intramural research Grant 06-IRGP-789.

*Key words and phrases.* Experimental cross, genomic imprinting, likelihood ratio test, quantitative trait loci, variance components model.

Genomic imprinting refers to the situation where the expression of the same genes is different depending on their parental origin [Pfifer (2000)]. It has been increasingly recognized that many endosperm traits are controlled by genomic imprinting. For example, endoreduplication is a commonly observed phenomenon which shows a maternally controlled parent-of-origin effect in maize endosperm [Dilkes et al. (2002)]. Cells undergoing endoreduplication are typically larger than other cells, which consequently results in larger fruits or seeds beneficial to human beings [Grime and Mowforth (1982)]. Other reports of genomic imprinting with paternal imprinting in maize endosperm include, for instance, the *r* gene in the regulation of anthocyanin [Kermicle (1970)], the seed storage protein regulatory gene *dsrl* [Chaudhuri and Messing (1994)], the *MEA* gene affecting seed development [Kinoshita et al. (1999)] and some *α-tubulin* genes [Lund, Messing and Viotti (1995)]. These studies underscore the value of developing statistical methods that empower geneticists to identify the distribution and effects of imprinted genes controlling endosperm traits.

Statistical methods for mapping imprinted genes or imprinted quantitative trait loci (iQTL) have been extensively studied. Focusing on different genetic designs and different segregation populations, methods were developed in mapping iQTL underlying quantitative traits in controlled experimental crosses [e.g., Cui, Cheverud and Wu (2007); Cui et al. (2006); Wolf et al. (2008)], in outbred population [e.g., de Koning, Bovenhuis and van Arendonk (2002)] and in human population [e.g., Hanson et al. (2001); Shete, Zhou and Amos (2003)]. Broadly speaking, these methods can be categorized into two frameworks: one based on the fixed effect model where the iQTL effect is considered as fixed [e.g., Cui et al. (2006, 2007); de Koning, Bovenhuis and van Arendonk (2002)], and the other considering iQTL effect as random and estimating the genetic variances contributed by an iQTL [e.g., Hanson et al. (2001); Shete, Zhou and Amos (2003); Li and Cui (2009a)]. The method proposed by Li and Cui (2009a) extended the variance components model to experimental crosses and showed relative merits in mapping iQTLs with inbred lines. However, all these approaches for iQTL mapping were developed based on diploid populations, whereby chromosomes are paired. Their applications are immediately limited when the ploidy level of the study population is more than two, for instance, the triploid endosperm.

In this study we propose to extend our previous work in iQTL mapping with the variance components approach in experimental crosses [Li and Cui (2009a)], and consider the unique genetic makeup of the triploid endosperm genome to map iQTLs underlying triploid endosperm traits. Cytoplasmic maternal effects are also considered and adjusted when testing for genomic imprinting. Motivated by a real experiment, we propose a reciprocal backcross design initiated with two inbred lines. The likelihood ratio test (LRT) is applied to test the significance of the variance components and its asymptotic distribution is evaluated under irregular conditions.

The article is organized as follows. Section 2 will illustrate the basic genetic design and the statistical mapping framework. We propose a new approach for calculating the parental specific allelic sharing among inbreeding triploid sibs. Statistical hypothesis testings are proposed to assess iQTL effects. The limiting distribution of the LRT under the proposed mapping framework is studied. The multiple iQTL model is also proposed to separate closely linked (i)QTLs. Sections 3 and 4 will be devoted to simulations and real application followed by a general discussion in Section 5.

## 2. Statistical method.

2.1. *The genetic design*. Using experimental crosses for QTL mapping has been the traditional means in targeting genetic regions harboring potential genes responsible for quantitative trait variations. Toward the goal of mapping iQTL underlying endosperm traits in line crosses, we propose a reciprocal backcross design. A similar design was proposed by Li and Cui (2009a) for diploid mapping populations. In brief, two inbred parents with genotypes $AA$ and $aa$ are crossed to produce an $F_1$ population ($Aa$). $F_1$ individuals are then backcrossed with one of the parents to generate backcross populations. We can use both parents as the maternal strain to cross with an $F_1$ individual to generate two backcross segregation populations. Or we can use $F_1$ individuals as the maternal strains to cross with both parents to produce another two sets of segregation populations. The so-called reciprocal backcross design generates four different segregation populations with each one being considered as one family. Large number of backcross families can be obtained by simply replicating each one of the above crosses.

To distinguish the allelic parental origin, we use subscript letters $f$ and $m$ to denote an allele inherited from the father and mother, respectively. A list of possible offspring genotypes considering the unique genetic makeups in the triploid endosperm genome is detailed in the second column in Table 1. Clearly, the endosperm genome carries one extra maternal copy due to the unique double fertilization step in flowering plants. When a dosage effect is considered, we do expect different expression values triggered by endosperm and embryo genes.

2.2. *The model*. In QTL mapping different line crosses can be combined together to increase the parameter inference space via a variance components method [Xie, Gessler and Xu (1998)]. VC method has been shown to be powerful in assessing genomic imprinting in human linkage analysis [Hanson et al. (2001)]. Recently, Li and Cui (2009a) extended the VC model to experimental crosses and proposed an iQTL mapping framework via combining different line crosses for iQTL detection. We extend our previous work to triploid endosperm tissue considering the unique genetic components in the endosperm genome.

Suppose total $K$ families are collected which are composed of the four distinct backcross families. Assume $n_k$ individuals are sampled in the $k$th family.

TABLE 1

*The allelic-specific IBD sharing coefficients for full-sib pairs in a reciprocal backcross design*

| Backcross | Offspring genotype | Parent-specific IBD sharing | | | | | | Total IBD | |
|---|---|---|---|---|---|---|---|---|---|
| | | $\pi_{mm}$ | | $\pi_{ff}$ | | $\pi_{m/f}$ | | $\pi$ | |
| | | $Q_mQ_mQ_f$ | $Q_mQ_mq_f$ | $Q_mQ_mQ_f$ | $Q_mQ_mq_f$ | $Q_mQ_mQ_f$ | $Q_mQ_mq_f$ | $Q_mQ_mQ_f$ | $Q_mQ_mq_f$ |
| $QQ \times Qq$ | $Q_mQ_mQ_f$ | 4/3 | 4/3 | 1/3 | 0 | 4/3 | 2/3 | 3 | 2 |
| | $Q_mQ_mq_f$ | 4/3 | 4/3 | 0 | 1/3 | 2/3 | 0 | 2 | 5/3 |
| | | $Q_mQ_mQ_f$ | $q_mq_mQ_f$ | $Q_mQ_mQ_f$ | $q_mq_mQ_f$ | $Q_mQ_mQ_f$ | $q_mq_mQ_f$ | $Q_mQ_mQ_f$ | $q_mq_mQ_f$ |
| $Qq \times QQ$ | $Q_mQ_mQ_f$ | 4/3 | 0 | 1/3 | 1/3 | 4/3 | 2/3 | 3 | 1 |
| | $q_mq_mQ_f$ | 0 | 4/3 | 1/3 | 1/3 | 2/3 | 0 | 1 | 5/3 |
| | | $q_mq_mQ_f$ | $q_mq_mq_f$ | $q_mq_mQ_f$ | $q_mq_mq_f$ | $q_mq_mQ_f$ | $q_mq_mq_f$ | $q_mq_mQ_f$ | $q_mq_mq_f$ |
| $qq \times Qq$ | $q_mq_mQ_f$ | 4/3 | 4/3 | 1/3 | 0 | 0 | 2/3 | 5/3 | 2 |
| | $q_mq_mq_f$ | 4/3 | 4/3 | 0 | 1/3 | 2/3 | 4/3 | 2 | 3 |
| | | $Q_mQ_mq_f$ | $q_mq_mq_f$ | $Q_mQ_mq_f$ | $q_mq_mq_f$ | $Q_mQ_mq_f$ | $q_mq_mq_f$ | $Q_mQ_mq_f$ | $q_mq_mq_f$ |
| $Qq \times qq$ | $Q_mQ_mq_f$ | 4/3 | 0 | 1/3 | 1/3 | 0 | 2/3 | 5/3 | 1 |
| | $q_mq_mq_f$ | 0 | 4/3 | 1/3 | 1/3 | 2/3 | 4/3 | 1 | 3 |

The phenotypic variation of a quantitative trait in family $k$ (denoted as $y_k$) can be explained by the genotype-specific cytoplasmic maternal effect (denoted as $\mu_k$), additive QTL effect (denoted as $a_k$), polygene effect (denoted as $g_k$) and random residual effect (denoted as $e_k$). To incorporate the parent-of-origin effect, the additive QTL effect ($a_k$) can be further partitioned into two separate effects, an effect due to the expression of the maternal allele (denoted as $a_{km}$) and an effect due to the expression of the paternal allele (denoted as $a_{kf}$). The model can thus be expressed as

$$y_{ki} = \mu_k + 2a_{kmi} + a_{kfi} + g_{ki} + e_{ki}, \qquad k = 1, \ldots, K; i = 1, \ldots, n_k, \tag{2.1}$$

where $a_{kmi}$, $a_{kfi}$, $g_{ki}$ and $e_{ki}$ are random effects with normal distribution, that is, $a_{kmi} \sim N(0, \pi_{i_m j_m|k}\sigma_m^2)$, $a_{kfi} \sim N(0, \pi_{i_m/j_f|k}\sigma_f^2)$, $g_{ki} \sim N(0, \phi_{ij|k}\sigma_g^2)$, $e_{ki} \sim N(0, \sigma_e^2)$; $g_{ki}$ and $e_{ki}$ are uncorrelated to $a_{kmi}$ and $a_{kfi}$; the coefficient 2 for $a_{kmi}$ adjusts for the effects of two identical maternal copies; $\mu_k$ models the maternal genotype-specific effect; $\pi_{i_m j_m|k}$, $\pi_{i_f j_f|k}$ and $\phi_{ij|k}$ are the IBD coefficients which are explained in the following section. With four distinct segregation populations, we have only three distinct maternal genotypes, $AA$, $Aa$ and $aa$. Thus, the parameter $\mu_k$ can be collapsed into three distinct values denoted as $\mu_1$, $\mu_2$ and $\mu_3$ corresponding to maternal genotypes $AA$, $Aa$ and $aa$, respectively. Letting $\beta = (\mu_1, \mu_2, \mu_3)$, then model (2.1) can be rewritten in a vector form as

$$\mathbf{y}_k = X_k\beta + 2\mathbf{a}_{km} + \mathbf{a}_{kf} + \mathbf{g}_k + \mathbf{e}_k, \qquad k = 1, \ldots, K, \tag{2.2}$$

where $X_k$ is an $n_k \times 3$ matrix with one column of ones and two columns of zeros.

2.3. *Parent-specific allele sharing and the covariance between two inbreeding sibs*. One of the major tasks in IBD-based iQTL mapping with the variance components model is to calculate the IBD sharing probabilities and the phenotypic covariances between sibs. Such a method has been developed in the human population [Hanson et al. (2001)], which, however, cannot be applied to a complete inbreeding population in experimental crosses, because the allelic sharing relationship among sibpairs does not follow the pattern as the one derived from a natural noninbreeding population. Instead, the IBD sharing probability can be calculated based on Malécot's coefficient of coancestry (1948) for an inbreeding population. Li and Cui (2009a) recently explored different allelic sharing patterns among sibpairs in a reciprocal backcross design with a diploid tissue. We extend the method to the triploid endosperm genome and derive covariances among sibpairs in a triploid tissue.

Consider two individuals $i$ and $j$ randomly selected from one backcross family with phenotype $y_i$ and $y_j$. Figure 1 shows all possible allelic sharing patterns between individuals $i$ and $j$. The solid line indicates IBD sharing for alleles derived from the same parent and the dotted line indicates IBD cross-sharing for alleles derived from different parents. The allelic cross-sharing is unique to inbreeding populations, whereby this cross-sharing probability reduces to zero for

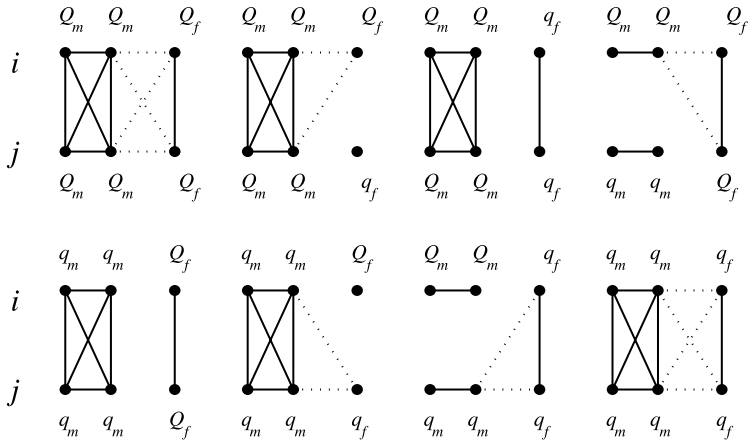


FIG. 1. *Possible alleles shared IBD for individuals $i$ and $j$ in inbreeding backcross families. The solid lines indicate IBD sharing for alleles inherited from the same parent. The dotted lines indicate IBD cross-sharing for alleles inherited from different parents.*

noninbreeding populations. Here we propose to calculate the IBD sharing between individuals $i$ and $j$ (denoted as $\pi_{ij}$) for a triploid genome as

$$\pi_{ij} = \begin{cases} 3\theta_{ij}, & \text{if } i \neq j, \\ \frac{1}{3}(5 + 3F_i), & \text{if } i = j, \end{cases} \tag{2.3}$$

where $\theta_{ij}$ is Malécot's coefficient of coancestry and $F_i$ is the inbreeding coefficient [Harris (1964); Cockerham (1983); Lynch and Walsh (1998)]. By definition, $\theta_{ij}$ is calculated as the probability of two randomly selected alleles from individuals $i$ and $j$ being identical by descent. The calculation of $\pi_{ij}$ is different from the usual IBD sharing calculation in noninbreeding populations. It is instead interpreted as triple the Malécot coefficient of coancestry [Xie, Gessler and Xu (1998)]. For easy notation, we still adopt the term "IBD sharing probability" for $\pi_{ij}$ in the rest of the presentation. The calculation of the inbreeding coefficient follows the procedure given in Lynch and Walsh (1998).

To illustrate the idea, consider two backcross individuals $i$ (with genotype $A_m A_m A_f$) and $j$ (with genotype $B_m B_m B_f$). The coefficient of coancestry $\theta_{ij}$ between these two individuals can be expressed as

$$\begin{aligned} \theta_{ij} &= \tfrac{1}{9}\{\Pr(A_{m1} \equiv B_{m1}) + \Pr(A_{m1} \equiv B_{m2}) + \Pr(A_{m2} \equiv B_{m1}) \\ &\qquad + \Pr(A_{m2} \equiv B_{m2}) + \Pr(A_{m1} \equiv B_f) + \Pr(A_{m2} \equiv B_f) \\ &\qquad + \Pr(A_f \equiv B_{m1}) + \Pr(A_f \equiv B_{m2}) + \Pr(A_f \equiv B_f)\} \\ &= \tfrac{1}{9}(4\theta_{i_m j_m} + 2\theta_{i_m j_f} + 2\theta_{i_f j_m} + \theta_{i_f j_f}), \end{aligned}$$

where the notation $\equiv$ refers to identical by decent; the subscript numbers 1 and 2 indicate two maternally inherited alleles; $\theta_{i \cdot j \cdot}$ is defined as the allelic kinship

coefficient [Lynch and Walsh (1998)]. Note that the two terms $\theta_{i_m j_f}$ and $\theta_{i_f j_m}$ are indistinguishable, but their sum denoted as $\theta_{i_m/j_f} (= \theta_{i_m j_f} + \theta_{i_f j_m})$ is unique. Thus, we have $\theta_{ij} = \frac{1}{9}(4\theta_{i_m j_m} + 2\theta_{i_m/j_f} + \theta_{i_f j_f})$. Following equation (2.3), we have

$$\pi_{ij} = 3\theta_{ij} = \tfrac{4}{3}\theta_{i_m j_m} + \tfrac{2}{3}\theta_{i_m/j_f} + \tfrac{1}{3}\theta_{i_f j_f} = \pi_{i_m j_m} + \pi_{i_m/j_f} + \pi_{i_f j_f} \qquad \text{for } i \neq j.$$

It can be seen that the IBD sharing between any two individuals can be decomposed as three separate components, one due to the IBD sharing for alleles derived from the maternal parent ($\pi_{i_m j_m} = \frac{4}{3}\theta_{i_m j_m}$), one due to the cross-sharing for alleles derived from different parents ($\pi_{i_m/j_f} = \frac{2}{3}\theta_{i_m/j_f}$) and one due to the IBD sharing for alleles derived from the paternal parent ($\pi_{i_f j_f} = \frac{1}{3}\theta_{i_f j_f}$). An exhaustive list of all possible IBD sharing probabilities for the four backcross families is given in Table 1.

Dropping the family index $k$, the covariance between any two individuals $i$ and $j$ can be expressed as

$$\begin{aligned}
&\operatorname{Cov}(y_i, y_j | \pi_{i_m j_m}, \pi_{i_m/j_f}, \pi_{i_f j_f}) \\
&\quad = \operatorname{Cov}(2a_{mi} + a_{fi} + g_i + e_i, 2a_{mj} + a_{fj} + g_j + e_j) \\
&\quad = 4\pi'_{i_m j_m}\sigma_m^2 + 2\pi'_{i_m/j_f}\sigma_{mf}^2 + \pi_{i_f j_f}\sigma_f^2 + \phi_{ij}\sigma_g^2 + I_{ij}\sigma_e^2,
\end{aligned}$$

where $\pi'_{i_m j_m} = \frac{1}{4}(\pi_{i_m j_m})$ and $\pi'_{i_m/j_f} = \frac{1}{2}(\pi_{i_m/j_f})$ are the IBD sharing and cross-sharing probabilities by considering one single maternal allele; $\sigma_{mf}^2$ measures the variation of trait distribution due to alleles cross-sharing; $\phi_{ij}$ is the expected alleles shared IBD; $I_{ij}$ is an indicator variable taking value 1 if $i = j$ and 0 if $i \neq j$. For a natural population without inbreeding, there is no allele cross-sharing for an individual with itself, hence, $\pi_{i_m/j_f} = 0$. For a diploid noninbreeding population, the trait covariance can be simplified as the one given in Shete, Zhou and Amos (2003). In matrix form, the phenotypic variance-covariance for individuals in the $k$th backcross family can then be expressed as

$$\boldsymbol{\Sigma}_k = \boldsymbol{\Pi}_{m|k}\sigma_m^2 + \boldsymbol{\Pi}_{m/f|k}\sigma_{mf}^2 + \boldsymbol{\Pi}_{f|k}\sigma_f^2 + \boldsymbol{\Phi}_{g|k}\sigma_g^2 + \mathbf{I}\sigma_e^2, \tag{2.4}$$

where the elements of $\boldsymbol{\Pi}_{m|k}$, $\boldsymbol{\Pi}_{f|k}$ and $\boldsymbol{\Pi}_{m/f|k}$ can be found in Table 1.

2.4. *QTL IBD sharing and genome-wide linkage scan.* The above described IBD sharing probability is calculated at a known marker position. Unless markers are dense enough, we have to search across the genome for potential (i)QTL positions and their effects. In general, the QTL position can be viewed as a fixed parameter by searching for a putative QTL at every 1 or 2 cM on a map interval bracketed by two markers throughout the entire linkage map. Thus, we need to estimate the QTL IBD sharing at every scan position. Since the conditional probability of an endosperm QTL given upon two flanking markers is the same as the one derived from a diploid genome [Cui and Wu (2005)], the same procedure termed

as the expected conditional IBD sharing described in Li and Cui (2009a) can be applied to calculate the QTL IBD sharing probability at every scan position.

Assuming multivariate normality of the trait distribution for data in each family and assuming independence between families, the joint log-likelihood function when $K$ backcross families are sampled can be formulated as

$$\ell = \sum_{k=1}^{K} \log[f(\mathbf{y}_k; \mu_k, \mathbf{\Sigma}_k)], \tag{2.5}$$

where $f$ is the multivariate normal density. Parameters to be estimated include $\beta = (\mu_1, \mu_2, u_3)$ and $\Omega = (\sigma_m^2, \sigma_f^2, \sigma_{mf}^2, \sigma_g^2, \sigma_e^2)$. Two commonly used methods in linkage analysis, the maximum likelihood (ML) method and the restricted maximum likelihood (REML) method, may be applied to estimate parameters. It is commonly recognized that the REML method gives less biased estimation compared to the ML method [Corbeil and Searle (1976)]. Here we adopt the REML method with the Fisher scoring algorithm to obtain the REML estimates [see Li and Cui (2009a) for details of the algorithm].

The conditional QTL IBD-sharing values vary at different testing positions. The amount of support for a QTL at a particular map position can be displayed graphically through the use of likelihood ratio profiles, which reflect the variation of the testing position of putative QTLs. The significant QTLs are detected by the peaks of the profile plot that pass a certain significant threshold (see Section 2.5 for more details).

2.5. *Hypothesis testing*. In iQTL mapping, we are interested in testing whether there is any significant genetic effect at a test position and would like to further quantify the imprinting effect if any. The hypothesis for testing the existence of a QTL can be expressed as

$$\begin{cases} H_0 : \sigma_m^2 = \sigma_f^2 = \sigma_{mf}^2 = 0, \\ H_1 : \text{at least one parameter is not zero.} \end{cases} \tag{2.6}$$

The LRT is applied for this purpose. Define $\widetilde{\Omega}$ and $\widehat{\Omega}$ to be the estimates of the unknown parameters under $H_0$ and $H_1$, respectively. The LRT statistic can be calculated as

$$\mathrm{LR} = -2[\log L(\widetilde{\Omega}|\mathbf{y}) - \log L(\widehat{\Omega}|\mathbf{y})]. \tag{2.7}$$

Let $\boldsymbol{\theta} = (\mu_1\ \mu_2\ \mu_3\ \theta_1\ \theta_2\ \theta_3\ \theta_4\ \theta_5)^T = (\mu_1\ \mu_2\ \mu_3\ \sigma_m^2\ \sigma_f^2\ \sigma_{mf}^2\ \sigma_g^2\ \sigma_e^2)^T \in \Omega = \mathbb{R}^3 \times [0,\infty) \times [0,\infty) \times [0,\infty) \times (0,\infty) \times (0,\infty)$ be the parameters to be estimated. Note that the polygene variance is bounded away from zero if we assume there are more than one QTL in the genome. Let the true parameters under the null hypothesis be $\boldsymbol{\theta}_0 = (\mu_{10}\ \mu_{20}\ \mu_{30}\ \sigma_{m_0}^2\ \sigma_{f_0}^2\ \sigma_{mf_0}^2\ \sigma_{g_0}^2\ \sigma_{e_0}^2)^T = (\mu_{10}\ \mu_{20}\ \mu_{30}\ 0\ 0\ 0\ \sigma_{g_0}^2\ \sigma_{e_0}^2)^T \in \Omega_0 = \mathbb{R}^3 \times \{0\} \times \{0\} \times \{0\} \times (0,\infty) \times (0,\infty)$.

The three tested genetic variance components under the null hypothesis lie on the boundaries of the parameter space $\Omega$. Thus, the standard conditions for obtaining the asymptotic $\chi^2$ distribution of the LRT are not satisfied [Self and Liang (1987)]. Following the results from Chernoff (1954), Shapiro (1985) and Self and Liang (1987), the following theorem states that the LR statistic follows a mixture chi-square distribution, whereby the mixture proportions depend on the estimated Fisher information matrix.

THEOREM 2.1. *Let $C_{\Omega_0}$ and $C_\Omega$ be closed convex cones with vertex at $\boldsymbol{\theta}_0$ to approximate $\Omega_0$ and $\Omega$, respectively. Let $\mathbf{Y}$ be a random variable with a multivariate normal distribution with mean $\boldsymbol{\theta}_0$, and variance–covariance matrix $I^{-1}(\boldsymbol{\theta}_0)$. Under the assumptions given in the Appendix, the LR statistic in (2.7) is asymptotically distributed as a mixture chi-square distribution with the form $\omega_3\chi_3^2 : \omega_2\chi_2^2 : \omega_1\chi_1^2 : \omega_0\chi_0^2$, where $\omega_3 = \frac{1}{4\pi}[2\pi - \cos^{-1}\rho_{12} - \cos^{-1}\rho_{13} - \cos^{-1}\rho_{23}]$, $\omega_2 = \frac{1}{4\pi}[3\pi - \cos^{-1}\rho_{12|3} - \cos^{-1}\rho_{13|2} - \cos^{-1}\rho_{23|1}]$, $\omega_1 = \frac{1}{4\pi}(\cos^{-1}\rho_{12} + \cos^{-1}\rho_{13} + \cos^{-1}\rho_{23})$, and $\omega_0 = \frac{1}{2} - \frac{1}{4\pi}[3\pi - \cos^{-1}\rho_{12|3} - \cos^{-1}\rho_{13|2} - \cos^{-1}\rho_{23|1}]$; $\rho_{ab}$ is the correlation between the variance terms a and b calculated from the Fisher information matrix, and $\rho_{ab|c} = \frac{(\rho_{ab} - \rho_{ac}\rho_{bc})}{(1-\rho_{ac}^2)^{1/2}(1-\rho_{bc}^2)^{1/2}}$.*

Note that the symbol $\pi$ in the above theorem is the irrational number (a mathematical constant) not the IBD sharing probability. The proof of the theorem is given in the Appendix.

REMARK. When the random parameter estimators are uncorrelated or the correlation is extremely small, that is, the Fisher information matrix is close to diagonal, the mixture proportions for the $\chi_k^2$ components are reduced to the binomial form with $\binom{3}{k}2^{-3}$, which is consistent with the result (Case 9) given in Self and Liang (1987).

Once a QTL is identified at a genomic position, we can further assess its imprinting property. To evaluate whether a QTL shows imprinting effect, the hypotheses can be formulated as

$$\begin{cases} H_0 : \sigma_f^2 = \sigma_m^2, \\ H_1 : \sigma_f^2 \neq \sigma_m^2. \end{cases} \tag{2.8}$$

Again, the likelihood ratio test can be applied which asymptotically follows a $\chi^2$ distribution with 1 degree of freedom since the tested parameter under the null is nonnegative and does not lie on the boundary of the parameter space. Rejecting $H_0$ indicates genomic imprinting, and the QTL can be called an iQTL. We denote this imprinting test as $\mathrm{LR}_{\mathrm{imp}}$. If the null is rejected, one would be interested in testing

whether the detected iQTL is completely maternally or paternally imprinted with the corresponding null hypothesis expressed as $H_0 : \sigma_m^2 = 0$ and $H_0 : \sigma_f^2 = 0$, respectively. The LRT statistic for the two tests asymptotically follows a mixture $\chi^2$ distribution with the form $\frac{1}{2}\chi_0^2 : \frac{1}{2}\chi_1^2$. Rejection of complete imprinting indicates partial imprinting.

Maternal effects can be tested by formulating hypothesis: $H_0 : \mu_1 = \mu_2 = \mu_3$. Note that these three parameters do not represent the true maternal effects, as they are confounded with the main genetic effects. But a test of pairwise differences can be applied to detect the significance of any maternal contribution.

2.6. *Multiple iQTL model.* In practice, there may be several QTLs to reflect the phenotypic variation in the whole genome. When testing QTL effects at one chromosome, effects from QTLs located at other chromosomes are absorbed by the polygenic effect ($g$). In some cases, two or more QTLs may be located at the same chromosome, which are termed as background QTL(s) in comparison to the tested one. When this happens, it is essential to adjust for the background QTL(s)' effects. Otherwise, it may lead to biased estimation for the putative QTL caused by the interference of QTL(s) close to the tested interval [Zeng (1994)].

In the previous work of Li and Cui (2009a), the authors proposed a multiple iQTL model following the idea of next-to-flanking markers proposed by Xu and Atchley (1995). We adopted a similar strategy in the current study. Briefly, assuming there are $S$ (i)QTLs in one chromosome, the multiple iQTL model considering parent-specific allele effect can be expressed as

$$y_{ki} = \mu_k + \sum_{s=1}^{S} 2a_{kmis} + \sum_{s=1}^{S} a_{kfis} + g_{ki} + e_{ki}, \qquad k = 1, \ldots, K; i = 1, \ldots, n_k,$$

where each (i)QTL effect is partitioned as two separate terms to reflect the contribution of the maternal and paternal alleles. In reality, the exact number and location of QTLs in a chromosome is generally unknown before doing a genome-wide search. This problem can be eased by applying the next-to-flanking markers idea proposed by Xu and Atchley (1995).

Denote a test interval with two flanking markers as $\mathcal{M}_l$–$\mathcal{M}_r$. The markers next to these two markers are denoted as $\mathcal{M}_L$ on the left of $\mathcal{M}_l$, and $\mathcal{M}_R$ on the right of $\mathcal{M}_r$ ($L = l - 1$ and $R = r + 1$). Conditional on the two markers, $\mathcal{M}_L$ and $\mathcal{M}_R$, we expect the effects of QTL(s) located outside of the tested interval can be absorbed by the IBD values calculated from the two next-to-flanking markers [Xu and Atchley (1995)]. Thus, the calculation of (i)QTL covariance conditional on these two markers will avoid the requirement for the position of QTLs outside of the tested interval. Dropping the family index, the phenotypic covariance between

two individuals $i$ and $j$ can be expressed as

$$\begin{aligned}
&\mathrm{Cov}(y_i, y_j | \pi_L, \hat{\pi}_{i_m j_m}, \hat{\pi}_{i_m/j_f}, \hat{\pi}_{i_f j_f}, \pi_R) \\
&= \sum_{l=1}^{L} K(\theta_{lL}, \pi_L)\sigma_l^2 + \hat{\pi}_{i_m j_m}\sigma_m^2 + \hat{\pi}_{i_m/j_f}\sigma_{mf}^2 + \hat{\pi}_{i_f j_f}\sigma_f^2 \\
&\quad + \sum_{r=1}^{R} K(\theta_{lR}, \pi_R)\sigma_r^2 + \phi_{ij}\sigma_g^2 + I_{ij}\sigma_e^2 \\
&= \pi_L\sigma_L^2 + \hat{\pi}_{i_m j_m}\sigma_m^2 + \hat{\pi}_{i_m/j_f}\sigma_{mf}^2 + \hat{\pi}_{i_f i_f}\sigma_f^2 + \pi_R\sigma_R^2 + \phi_{ij}\sigma_g^2 + I_{ij}\sigma_e^2,
\end{aligned}$$

where $\pi_L$ is the IBD sharing value at marker $L$, and $\sigma_L^2$ is a composite variance component which reflects the variation of (i)QTL effects on the left side of the tested interval [see Li and Cui (2009a) for details]. $\pi_R$ and $\sigma_R^2$ are defined similarly. The calculations of $\pi_L$ and $\pi_R$ reflect the triploid structure of the endosperm genome. Testing (i)QTL effects can then be focused on a tested interval while adjusting for the background QTLs' effects located in another place.

**3. Simulation.** Simulation studies are conducted to investigate the method performance. We assume a fixed total sample size of 400, then vary the family and offspring size with different combinations, that is, $4 \times 100$, $8 \times 50$, $20 \times 20$ and $100 \times 4$, in order to evaluate the effect of family and offspring size on testing power and parameter estimation. Simulation details are given in the Simulation and real data analysis. Here we briefly summarize the main results.

3.1. *Single iQTL simulation.* For the single iQTL simulation, the results show that both the $4 \times 100$ and the $100 \times 4$ designs yield lower QTL detection power and higher RMSE (root mean squared error) for QTL position estimation than the other two designs do. The $20 \times 20$ design slightly beats the $8 \times 50$ design with smaller imprinting type I error and higher QTL detection power. These results indicate that it is necessary to maintain a balance between the family size and the offspring size, in order to achieve optimal power and good effects estimation precision. For a given budget with a fixed total sample size, one should always try to avoid extreme designs with a large (or small) number of families, each with a small (or large) number of offsprings.

Focusing on the $20 \times 20$ design, simulations are performed to show the model behavior under different imprinting modes, that is, complete paternal imprinting, complete maternal imprinting, partial maternal imprinting and partial paternal imprinting. The results indicate that the power to detect imprinting depends on the underlying degree of imprinting. Relatively higher imprinting power is observed when an iQTL is maternally imprinting compared to the case when an iQTL is paternally imprinting.

3.2. *Multiple iQTL simulation.* In this simulation data are simulated by assuming two (i)QTLs located at two genomic positions and are subject to both the single iQTL and multiple iQTL analyses. The results indicate a clear benefit of analysis by fitting a multiple iQTL model rather than fitting a single iQTL model. While the single iQTL analysis detects one "ghost" QTL located between the two simulated QTLs, the multiple iQTL analysis can clearly separate the two QTLs with high precision. Note that the multiple iQTL analysis normally generates lower LR values than the single iQTL analysis does. Note that the distribution of the LR value under the multiple iQTL analysis is not clear, and permutation should be applied to assess significance of any (i)QTLs in multiple iQTL analysis [Xu and Atchley (1995)].

**4. A case study.** We apply our method to a real data set which has two endosperm traits of interests: mean ploidy level (denoted as Mploidy) and percentage of endoreduplicated nuclei (denoted as Endo). The two traits describe the level of endoreduplication in maize endosperm, which is thought to be genetically controlled by imprinted genes [Dilkes et al. (2002)]. Four backcross (BC) segregation populations, initiated with two inbred lines, Sg18 and Mo17, were sampled. The four BC populations were obtained following the design illustrated in Table 1. The data show a large degree of variation for endoreduplication among the four BC populations, and ten linkage groups were constructed from the observed marker data [Coelho et al. (2007)]. Readers are referred to Coelho et al. (2007) for more details about the data. The two traits are analyzed with our multiple iQTL model aimed to identify iQTLs across the ten linkage groups. The data are also analyzed with a Mendelian model. Results from both imprinting and Mendelian models are compared and summarized in the Supplementary Materials.

Figure 2 plots the LR values across the ten linkage groups for the two traits. The solid and dotted curves represent LR profiles for traits Endo and Mploidy, respectively. To adjust for the genome-wide error rate across the entire linkage group, permutation tests are applied in which the critical threshold value is empirically calculated on the basis of repeatedly shuffling the relationships between marker genotypes and phenotypes within each BC family [Churchill and Doerge (1994)]. The corresponding genome-wide significance thresholds (at 5% level) for the two traits are denoted by the horizontal solid (for Endo) and dotted (for Mploidy) lines. The 5% level chromosome-wide thresholds are denoted by the dashed (for Endo) and dash-dotted (for Mploidy) lines. QTLs that are significant at the chromosome-wide level are called suggestive QTLs. It can be seen that two QTLs (on G7 and G9) associated with Mploidy and one QTL (on G6) associated with Endo are detected at the 5% genome-wide significance level (denoted by "$*$" in Table 2). Two suggestive QTLs (on G2 and G10) associated with Endo and one suggestive QTL (on G6) associated with Mploidy are also identified. The detailed QTL location and effect estimates as well as the test results for imprinting are tabulated in Table 2. For the trait Mploidy, the identified three QTLs are all imprinted ($p_{\text{imp}} < 0.05$)

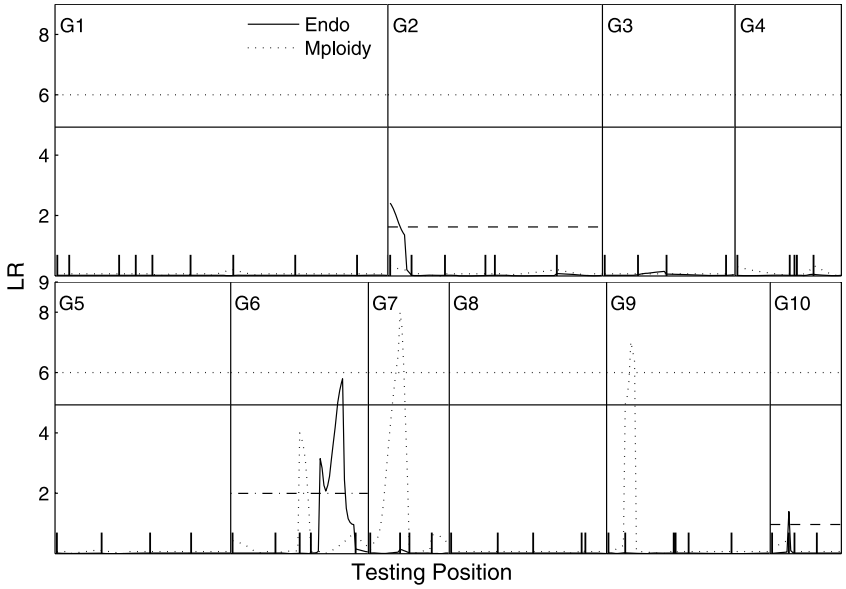


FIG. 2. *The profile of the log-likelihood ratios* (*LR*) *for testing the existence of QTLs underlying the two endosperm traits across the* 10 *maize linkage groups* $(G_1, \ldots, G_{10})$. *The genome-wide LR profiles for the percentage of endoreduplication* (*Endo*) *and mean ploidy* (*Mploidy*) *traits are indicated by solid and dotted curves*, *respectively. The threshold values for claiming the existence of QTLs are given as the horizonal solid and dotted line for the genome-wide threshold*, *and the dashed and dash-dotted line for the chromosome-wide threshold*, *for the two traits Endo and Mploidy*, *respectively. The genomic positions corresponding to the peak of the curves that pass the corresponding thresholds are the MLEs of the QTL location. The positions of markers on the linkage groups* [*Coelho et al*. (2007)] *are indicated at ticks*.

and all show completely maternal imprinting, that is, the maternal copy does not express. They are thus termed iQTLs. The cytoplasmic maternal effect does not show any evidence of significance for all the three iQTLs ($p_M > 0.05$). For the trait Endo, only the QTL detected on G6 shows imprinting effect ($p_{\text{imp}} < 0.05$) and it shows completely paternal imprinting ($p_f < 0.05$). The other two QTLs do not show evidence of imprinting ($p_{\text{imp}} > 0.05$). For this trait, significant maternal effects are detected ($p_M < 0.01$).

In our study, one maternally controlled iQTL is detected for trait Endo, which is consistent with the result given by Dilkes et al. (2002). Meanwhile, according to the genetic conflict theory proposed by Haig and Westoby (1991), maternally derived alleles tend to trigger a negative effect on the increase of endosperm growth, whereas paternally derived alleles tend to play an opposite effect to increase seed size. The identified iQTLs showing maternal imprinting for trait Mploidy can be well explained by the genetic conflict theory. Both empirical evidence and theoretical hypothesis support the current finding.

TABLE 2

*The estimated parameters for the three maternal effects and the variance components for two endosperm traits*: *mean ploidy* (*Mploidy*) *and percent of the endoreduplicated nuclei* (*Endo*)

| | | Maternal effects | | | Genetic effects | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Trait** | **Ch** | $\mu_1$ | $\mu_2$ | $\mu_3$ | $\sigma_m^2$ | $\sigma_f^2$ | $\sigma_{mf}^2$ | $\sigma_L^2$ | $\sigma_R^2$ | $\sigma_g^2$ | $\sigma_e^2$ | $p_M$ | $p_{\text{imp}}$ | $p_m$ | $p_f$ |
| Mploidy | 6* | 13.13 | 11.88 | 9.78 | 0.01 | 0.30 | 0.03 | ≈0 | 0.22 | 1.25 | 2.59 | 0.34 | 0.045 | 0.023 | 0.31 |
| | 7 | 11.78 | 11.19 | 9.16 | 0.15 | 0.60 | 0.94 | ≈0 | 0.12 | 1.07 | 2.69 | 0.31 | 0.048 | 0.024 | 0.49 |
| | 9 | 13.84 | 12.08 | 10.01 | ≈0 | 0.94 | 0.71 | ≈0 | 0.01 | 1.59 | 2.55 | 0.12 | 0.013 | 0.021 | 0.48 |
| Endo | 2* | 72.23 | 62.40 | 52.86 | 0.43 | 0.83 | 2.41 | 0.99 | ≈0 | 5.10 | 37.49 | <0.01 | 0.67 | – | – |
| | 6 | 68.37 | 63.18 | 54.92 | 2.92 | ≈0 | 7.14 | 1.42 | 0.92 | 1.28 | 38.91 | <0.01 | 0.02 | 0.28 | 0.01 |
| | 10* | 70.78 | 62.28 | 50.67 | 0.58 | 0.03 | 1.52 | ≈0 | 0.17 | 3.24 | 39.20 | <0.01 | 0.29 | – | – |

The three QTLs for trait Mploidy are located at marker umc1805, marker dupssr9 and umc1040 + 5.76cM on chromosome 6, 7 and 9, respectively. The three QTLs for trait Endo are located at marker umc2094, bnlg345 + 33.49cM and MMC501 + 18cM on chromosome 2, 6 and 10, respectively. QTLs showing significance at the genome-wide significance level are indicated by "$*$". $p_M$, $p_{\text{imp}}$, $p_m$ and $p_f$ are the $p$-values for testing maternal effect ($H_0 : \mu_1 = \mu_2 = \mu_3$), imprinting effect ($H_0 : \sigma_m^2 = \sigma_f^2$), complete maternal imprinting ($H_0 : \sigma_m^2 = 0$) and complete paternal ($H_0 : \sigma_f^2 = 0$), respectively.

**5. Discussion.** The role of genomic imprinting in endosperm development has been commonly recognized [Dilkes et al. (2002); Kinoshita et al. (1999); Chaudhuri and Messing (1994)]. But little is known about the exact location and effect size of imprinted genes in endosperm. As endosperm in cereal provides the most nutrition for human beings, it is important to identify imprinted genes that govern seed development, particularly endosperm development. In this article we develop a variance components linkage analysis method with an experimental cross design, aimed to identify iQTLs in endosperm. Our method is motivated by real applications and is evaluated through Monte Carlo simulations.

The proposed method is based on a particular genetic design (reciprocal BC design) with inbreeding populations. We treat iQTL effects as random, different from a fixed-effect iQTL model [e.g., Cui (2007)]. Variance components linkage analysis with a partial inbreeding human population was previously proposed [see Abney, McPeek and Ober (2000)]. However, extending the VC model to a completely inbreeding population is challenging. In our previous work, we proposed a VC-based iQTL mapping framework for an inbreeding diploid mapping population [Li and Cui (2009a)]. Extending the previous work, we propose a novel IBD partitioning approach to calculate allelic sharing in an inbreeding endosperm population. Extension to mapping multiple iQTLs is provided. Simulations indicate good performance of the multiple iQTL analysis compared to a single iQTL model. Meanwhile, to obtain a good balance of iQTL position and effect estimation as well as detection power, we have to avoid extreme sample designs. For a fixed total sample size, extremely large or small families should be always avoided.

In an application to two endosperm traits, we identified three iQTLs for trait Mploidy. All show paternal expression. We also identified one iQTL for trait Endo, which shows a maternal expression. According to the parental conflict theory proposed by Haig and Westoby (1991), maternally derived alleles trigger a negative effect on endosperm cell growth and inhibit endosperm development because the extra maternal copy could slower nuclear division in endosperm. On the contrary, paternally derived alleles tend to increase seed size. Thus, the three iQTLs identified for Mploidy can be explained by the genetic conflict theory. The occurrence of parental conflict theory explains parent-of-origin effects as an ubiquitous mechanism for the control of early seed development [Grossniklaus et al. (2001); Kinoshita et al. (1999)].

In VC-based linkage analyses, likelihood ratio test (LRT) has been commonly applied in assessing QTL significance. The LRT statistic asymptotically follows a mixture $\chi^2$ distribution with binomial mixture coefficients, as many investigators often claimed [following Case 9 in Self and Liang (1987)]. In a recent investigation, we found that the LRT in a regular VC-based linkage analysis without considering imprinting follows a mixture $\chi^2$ distribution with mixture proportions depending on the estimated Fisher information matrix [Li and Cui (2009b)]. The modified calculation of mixture proportion does give more reasonable type I error rate than the one with binomial coefficients. When imprinting is considered, we

show that the limiting distribution of the LRT also follows a mixture $\chi^2$ distribution, and we adopt the new criterion for power evaluation. Simulations show that the new criterion gives type I error closer to the nominal level than the one using binomial coefficients, and also produces power as good as the later one (data not shown). We recommend investigators adopt the new criterion in their analysis.

Increasing evidence has suggested that for correlated traits, multivariate approaches can increase the power and precision to identify genetic effects in genetic linkage analyses [e.g., Boomsma and Dolan (1998); Amos and Andrade (2001); Evans (2002)]. Also, the joint analysis of multivariate traits can provide a platform for testing a number of biologically interesting hypotheses, such as testing pleiotropic effects of QTL and testing pleiotropic vs close linkage. Moreover, if the putative QTL has pleiotropic effects on several traits, the joint analysis may perform better than mapping each trait separately [Jiang and Zeng (1995)]. Multivariate traits appear frequently in genetic mapping studies. For example, the two endosperm traits evaluated in this study are highly correlated [Coelho et al. (2007)]. We expect joint analysis may provide high mapping resolution and power for iQTL detection. This will be explored in our future investigation. A computer code written in R for implementing the current analysis is available upon request.

## APPENDIX

In standard human linkage analysis with a variance components model, many authors declare that the likelihood ratio statistic follows a mixture $\chi^2$ distribution with binomial coefficient for each mixture component [e.g., Amos and Andrade (2001); Hanson et al. (2001); Shete, Zhou and Amos (2003)]. Following Chernoff (1954), Shapiro (1985) and Self and Liang (1987), in the following we show that the mixture proportion actually depends on the estimated Fisher information matrix.

For a random sample $\mathbf{X}$ with density function $f(\mathbf{x};\boldsymbol{\theta})$, following Chernoff (1954) and Self and Liang (1987), assume that:

(i) For any true parameter $\boldsymbol{\theta}_0$, the neighborhood of $\boldsymbol{\theta}_0$ is closed and the intersection between this closure and $\Omega$ defined in the main text is also a closed set.

(ii) The first three derivatives of $\sum_i \log f(x_i;\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ on the intersection of the neighborhood of $\boldsymbol{\theta}_0$ and $\Omega$ almost surely exist. Moreover, $|\frac{\partial^3 \sum \log f}{\partial\theta_i \partial\theta_j \partial\theta_k}| < W(\mathbf{x})$ for all $\theta$ on the intersection, and $E[W(\mathbf{x})] < \infty$.

(iii) The information matrix $\mathcal{I}(\boldsymbol{\theta})$ is positive definite on neighborhoods of $\boldsymbol{\theta}_0$.

(vi) The set $\Omega$ is convex.

Assuming the above assumptions, the consistency, weak convergence and asymptotic normality of the estimators can be established [see Chernoff (1954); Self and Liang (1987); Shapiro (1985)]. Here we cite the main results from Chernoff (1954), Shapiro (1985) and Self and Liang (1987) to show the asymptotic distribution of the LRT in our case.

Defining two closed polyhedral convex cones $C_{\Omega_0}$ and $C_{\Omega_1}$ to approximate $\Omega_0$ and $\Omega_1$ at $\boldsymbol{\theta}_0$, the parameter space under the null hypothesis is approximated as $C_{\Omega_0} = \{\boldsymbol{\theta} : \boldsymbol{\theta} \in \mathbb{R}^3 \times \{0\} \times \{0\} \times \{0\} \times (0,\infty) \times (0,\infty)\}$, against $C_{\Omega_1} = \{\boldsymbol{\theta} : \boldsymbol{\theta} \in \mathbb{R}^3 \times [0,\infty) \times [0,\infty) \times [0,\infty) \times (0,\infty) \times (0,\infty)\}$ under the alternative. Let $\mathbf{Y}'$ be a random variable generated from the multivariate normal distribution, that is, $\mathbf{Y}' \sim N(\boldsymbol{\theta}_0,\ I^{-1}(\boldsymbol{\theta}_0))$. Following Chernoff [(1954), Theorem 1], the asymptotic distribution of the LRT in (2.7) is equivalent to the following quadratic approximation:

$$\text{(A1)}\quad LR^* = \inf_{\boldsymbol{\theta}\in C_{\Omega_0}} (\mathbf{Y}' - \boldsymbol{\theta})' I(\boldsymbol{\theta}_0)(\mathbf{Y}' - \boldsymbol{\theta}) - \inf_{\boldsymbol{\theta}\in C_{\Omega_1}} (\mathbf{Y}' - \boldsymbol{\theta})' I(\boldsymbol{\theta}_0)(\mathbf{Y}' - \boldsymbol{\theta}).$$

Subtracting $\boldsymbol{\theta}_0$ from $\mathbf{Y}'$ and $\boldsymbol{\theta}$, the expression in (A1) is given by

$$\text{(A2)}\quad LR^* = \inf_{\boldsymbol{\theta}\in C_{\Omega_0}-\boldsymbol{\theta}_0} (\mathbf{Y} - \boldsymbol{\theta})' I(\boldsymbol{\theta}_0)(\mathbf{Y} - \boldsymbol{\theta}) - \inf_{\boldsymbol{\theta}\in C_{\Omega_1}-\boldsymbol{\theta}_0} (\mathbf{Y} - \boldsymbol{\theta})' I(\boldsymbol{\theta}_0)(\mathbf{Y} - \boldsymbol{\theta}),$$

where $\mathbf{Y} = \mathbf{Y}' - \boldsymbol{\theta}_0 \sim N(\mathbf{0}, I^{-1}(\boldsymbol{\theta}_0))$ under the linear transformation.

Let $C^{\ddagger} = (C_{\Omega_1} - \boldsymbol{\theta}_0) \cap (C_{\Omega_0} - \boldsymbol{\theta}_0)^c = \{\boldsymbol{\theta} : \theta_1 > 0, \theta_2 > 0, \theta_3 > 0\}$, which is a closed polyhedral convex cone with 3 dimensions. By the Pythagoras theorem, the statistic in (A2) can be expressed as

$$\text{(A3)}\qquad LR^* = \inf_{\boldsymbol{\theta}\in C^{\ddagger}} (\mathbf{Y} - \boldsymbol{\theta})' I(\boldsymbol{\theta}_0)(\mathbf{Y} - \boldsymbol{\theta}).$$

Let $\mathcal{F}(C^{\ddagger})$ be the set of all faces of $C^{\ddagger}$. $C^{\ddagger 0} = \{\gamma \in \mathbb{R}^3 : \gamma'\boldsymbol{\theta} \le 0, \forall \boldsymbol{\theta} \in C^{\ddagger}\}$ is defined to be a polar cone such that $(C^{\ddagger 0})^0 = C^{\ddagger}$. Following Shapiro (1985), we can select a face $\nu \in \mathcal{F}(C^{\ddagger})$ corresponding to the polar face $\nu^0 \in \mathcal{F}(C^{\ddagger 0})$ such that the linear spaces generated by $\nu$ and $\nu^0$ are orthogonal to each other. For one face $\nu$ (or $\nu^0$), a projection $T_\nu$ (or $T_{\nu^0}$) [a symmetric idempotent matrix giving projection onto the space generated by $\nu$ (or $\nu^0$)] can be found such that $T_\nu = I - T_{\nu_0}$ since they are orthogonal. Then $T_\nu \mathbf{Y}$ (or $T_{\nu^0}\mathbf{Y}$) is a projection of $\mathbf{Y}$ onto $C^{\ddagger}$ (or $C^{\ddagger 0}$).

For a given $\mathbf{Y}$, let $g(\mathbf{Y})$ be the minimizer to achieve the infimum in (A3). Define $\psi_{\nu|\mathbf{Y}} = \{\mathbf{Y} \in \mathbb{R}^3 : g(\mathbf{Y}) \in \nu\}$ so that $g(\mathbf{Y}) \in \nu$ if and only if $T_\nu \mathbf{Y} \in C^{\ddagger}$ and $T_{\nu 0}\mathbf{Y} \in C^{\ddagger 0}$. By Shapiro (1985), $g(\mathbf{Y}) = T_\nu \mathbf{Y} \in C^{\ddagger}$, $\forall \mathbf{Y} \in \psi_{\nu|\mathbf{Y}}$.

Note that the set $\psi_{\nu|\mathbf{Y}}$ is composed of $2^3$ disjoint sets in $\mathbb{R}^3$. All these disjoint sets can be classified into four categories as follows:

(1) $\psi^1_{\nu|\mathbf{Y}} = \{\mathbf{Y}; Y_1 > 0, Y_2 > 0, Y_3 > 0, g(\mathbf{Y}) \in \nu\}$,

(2) $\psi^2_{\nu|\mathbf{Y}} = \{\mathbf{Y}; Y_1 > 0, Y_2 > 0, Y_3 \le 0, g(\mathbf{Y}) \in \nu\}$; $\psi^3_{\nu|\mathbf{Y}} = \{\mathbf{Y}; Y_1 > 0, Y_2 \le 0, Y_3 > 0, g(\mathbf{Y}) \in \nu\}$; $\psi^4_{\nu|\mathbf{Y}} = \{\mathbf{Y}; Y_1 \le 0, Y_2 > 0, Y_3 > 0, g(\mathbf{Y}) \in \nu\}$,

(3) $\psi^5_{\nu|\mathbf{Y}} = \{\mathbf{Y}; Y_1 \le 0, Y_2 \le 0, Y_3 > 0, g(\mathbf{Y}) \in \nu\}$; $\psi^6_{\nu|\mathbf{Y}} = \{\mathbf{Y}; Y_1 > 0, Y_2 \le 0, Y_3 \le 0, g(\mathbf{Y}) \in \nu\}$; $\psi^7_{\nu|\mathbf{Y}} = \{\mathbf{Y}; Y_1 \le 0, Y_2 > 0, Y_3 \le 0, g(\mathbf{Y}) \in \nu\}$,

(4) $\psi^8_{\nu|\mathbf{Y}} = \{\mathbf{Y}; Y_1 \le 0, Y_2 \le 0, Y_3 \le 0, g(\mathbf{Y}) \in \nu\}$.

By linear transformation, we cab define $C^* = \{\boldsymbol{\theta}^* : \boldsymbol{\theta}^* = \Lambda^{1/2} P' \boldsymbol{\theta}, \forall \boldsymbol{\theta} \in C^{\ddagger}\}$ which is a polyhedral closed convex cone. Then (A3) can be further expressed as

$$LR^* = \inf_{\boldsymbol{\theta}^* \in C^*} \|\mathbf{z} - \boldsymbol{\theta}^*\|^2, \tag{A4}$$

where $\mathbf{z} = \Lambda^{1/2} P' \mathbf{Y}$ [$P \Lambda P^T = I(\boldsymbol{\theta}_0)$] has a multivariate normal distribution with mean $\mathbf{0}$ and identity covariance matrix.

Let $C^{*0}$ be a polar cone of $C^*$ and $(C^{*0})^0 = C^*$. Two faces $\nu^*$ and $\nu^{*0}$ can be defined with respect to $\mathcal{F}(C^*)$ and $\mathcal{F}(C^{*0})$. The relevant orthogonal projections $T_{\nu^*}$ and $T_{\nu^{*0}}$ corresponding to $\nu^*$ and $\nu^{*0}$ can be defined. Suppose $h(\mathbf{z})$ is the minimizer to achieve the infimum in (A4). Following Shapiro (1985), a set $\psi_{\nu^*|\mathbf{z}}$ can be defined similarly as $\psi_{\nu|\mathbf{Y}}$, such that $h(\mathbf{z}) = T_{\nu^*}\mathbf{z} \in C^*$, $\forall \mathbf{z} \in \psi_{\nu^*|\mathbf{z}}$. It satisfies the conditions of Lemma 3.1 [Shapiro (1985)]. Then we have

$$LR^* = \|\mathbf{z} - h(\mathbf{z})\|^2 = \|\mathbf{z} - T_{\nu^*}\mathbf{z}\|^2 = \mathbf{z}'(I - T_{\nu^*})\mathbf{z} = \mathbf{z}' T_{\nu^{*0}} \mathbf{z} \qquad \forall \mathbf{z} \in \psi_{\nu^*|\mathbf{z}}. \tag{A5}$$

Thus, the distribution of $LR^*$ in (A3) can be evaluated by

$$\begin{aligned}
&\Pr(LR^* > c^2) \\
&\quad = \Pr\left( (\mathbf{Y} - g(\mathbf{Y}))' I(\boldsymbol{\theta}_0)(\mathbf{Y} - g(\mathbf{Y})) > c^2, \mathbf{Y} \in \bigcup_{i=1}^{2^3} \psi^i_{\nu|\mathbf{Y}} \right) \\
&\quad = \sum_{i=1}^{2^3} \Pr(\mathbf{Y} \in \psi^i_{\nu|\mathbf{Y}}) \Pr\big((\mathbf{Y} - g(\mathbf{Y}))' I(\boldsymbol{\theta}_0)(\mathbf{Y} - g(\mathbf{Y})) > c^2 | \mathbf{Y} \in \psi^i_{\nu|\mathbf{Y}}\big) \\
&\quad = \sum_{i=1}^{2^3} \Pr(\mathbf{Y} \in \psi^i_{\nu|\mathbf{Y}}) \Pr(\mathbf{z}' T_{\nu^{*0}} \mathbf{z} > c^2 | \mathbf{z} \in \psi^i_{\nu^*|\mathbf{z}}),
\end{aligned}$$

where, conditional on $\mathbf{z} \in \psi^i_{\nu^*|\mathbf{z}}$, $\mathbf{z}' T_{\nu^{*0}} \mathbf{z}$ is a chi-square distribution [Lemma 3.1, Shapiro (1985)]. By Bayes' theorem, the distribution of $LR^*$ follows a mixture chi-square distribution with mixing proportions $\Pr(\mathbf{Y} \in \psi^i_{\nu|\mathbf{Y}})$ $(i = 1, \ldots, 2^3)$ and $\sum_{i=1}^{2^3} \Pr(\mathbf{Y} \in \psi^i_{\nu|\mathbf{Y}}) = 1$.

The calculation of the mixture proportions follows Plackett (1954). Specifically, when $\mathbf{Y} \in \psi^1_{\nu|\mathbf{Y}}$, $LR^* \sim \chi^2_3$, and the corresponding mixture proportion $w_3 = \Pr(\mathbf{Y} \in \psi^1_{\nu|\mathbf{Y}}) = \frac{1}{4\pi}[2\pi - \cos^{-1}\rho_{12} - \cos^{-1}\rho_{13} - \cos^{-1}\rho_{23}]$. For category (2), $LR^* \sim \chi^2_2$ for $\mathbf{Y} \in \psi^i_{\nu|\mathbf{Y}}$, $i = 2, 3, 4$, with the corresponding mixture probability calculated by $w_2 = \sum_{j=2}^{4} \Pr(\mathbf{Y} \in \psi^i_{\nu|\mathbf{Y}}) = \frac{1}{4\pi}[3\pi - \cos^{-1}\rho_{12|3} - \cos^{-1}\rho_{13|2} - \cos^{-1}\rho_{23|1}]$. Correspondingly, $LR^* \sim \chi^2_1$ for $\mathbf{Y} \in \psi^i_{\nu|\mathbf{Y}}$, $i = 5, 6, 7$, with the relevant mixture probability evaluated as $w_1 = \sum_{j=5}^{7} \Pr(\mathbf{Y} \in \psi^i_{\nu|\mathbf{Y}}) = \frac{1}{2} - w_3$ in category (3). For the last category, $LR^* \sim \chi^2_0$ for $\mathbf{Y} \in \psi^8_{\nu|\mathbf{Y}}$ with the mixture probability $w_0 = \Pr(\mathbf{Y} \in \psi^8_{\nu|\mathbf{Y}}) = \frac{1}{2} - w_2$. Note $\rho_{ab}$ is the correlation between the terms $a$ and $b$ calculated from the Fisher information matrix, and $\rho_{ab|c} = \frac{(\rho_{ab} - \rho_{ac}\rho_{bc})}{(1-\rho^2_{ac})^{1/2}(1-\rho^2_{bc})^{1/2}}$. For more details of the derivation, the readers are referred to Li and Cui (2009b).

**Acknowledgments.** We thank B. Larkins for providing the endosperm mapping data. We also thank the Editor and two anonymous reviewers for helpful comments.

## SUPPLEMENTARY MATERIAL

**Simulation and real data analysis:** (DOI: 10.1214/09-AOAS323SUPP; .zip). Details for simulation are included in the supplemental file. We also analyze the data with a Mendelian model. A comparison of results with both imprinting and Mendelian models is summarized in the supplemental file.

## REFERENCES

ABNEY, M., MCPEEK, S. M. and OBER, C. (2000). Estimation of variance components of quantitative traits in inbred populations. *Am. J. Hum. Genet.* **66** 629–650.

AMOS, C. and ANDRADE, M. (2001). Genetic linkage methods for quantitative traits. *Stat. Methods Med. Res.* **10** 3–25.

BOOMSMA, D. I. and DOLAN, C. V. (1998). A comparison of power to detect a QTL in sib-pair data using multivariate phenotypes, mean phenotypes, and factor scores. *Behav. Genet.* **28** 329–340.

CHAUDHURY, A. M., KOLTUNOW, A., PAYNE, T., LUO, M., TUCKER, M. R., DENNIS, E. S. and PEACOCK, W. J. (2001). Control of early seed development. *Ann. Rew. Cell Dev. Biol.* **17** 677–699.

CHAUDHURI, S. and MESSING, J. (1994). Allele-specific parental imprinting of *dzrl*, a post transcriptional regulator of zein accumulation. *Proc. Natl. Acad. Sci.* **91** 4867–4871.

CHERNOFF, H. (1954). On the distribution of the likelihood ratio. *Ann. Math. Statist.* **25** 573–578. MR0065087

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

COCKERHAM, C. C. (1983). Covariances of relatives from self-fertilization. *Crop. Sci.* **23** 1177–1180.

COELHO, C. M., WU, S., LI, Y., HUNTER, B., DANTE, R. A., CUI, Y., WU, R. and LARKINS, B. A. (2007). Identification of quantitative trait loci that affect endoreduplication in maize endosperm. *Theor. Appl. Genet.* **115** 1147–1162.

CORBEIL, R. R. and SEARLE, S. R. (1976). A comparison of variance component estimators. *Biometrics* **32** 779–791. MR0443239

CUI, Y. H. (2007). A statistical framework for genome-wide scanning and testing imprinted quantitative trait loci. *J. Theoret. Biol.* **244** 115–126. MR2280488

CUI, Y., CHEVERUD, J. M. and WU, R. (2007). A statistical model for dissecting genomic imprinting through genetic mapping. *Genetica* **130** 227–239.

CUI, Y. H., LU, Q., CHEVERUD, J. M., LITTEL, R. L. and WU, R. L. (2006). Model for mapping imprinted quantitative trait loci in an inbred $F_2$ design. *Genomics* **87** 543–551.

CUI, Y. H. and WU, R. L. (2005). A statistical model for characterizing epistatic control of triploid endosperm triggered by maternal and offspring QTL. *Genet. Res.* **86** 65–76.

DE KONING, D.-J., BOVENHUIS, H. and VAN ARENDONK, J. A. M. (2002). On the detection of imprinted quantitative trait loci in experimental crosses of outbred species. *Genetics* **161** 931–938.

DILKES, B. P., DANTE, R. A., COELHO, C. and LARKINS, B. A. (2002). Genetic analysis 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.

GRIME, J. P. and MOWFORTH, M. A. (1982). Variation in genome size: An ecological interpretation. *Nature* **299** 151–153.

GROSSNIKLAUS, U., SPILLANE, C., PAGE, D. R. and KOEHLER, C. (2001). Genomic imprinting and seed development: Endosperm formation with and without sex. *Curr. Opin. Plant Biol.* **4** 21–27.

HAIG, D. and WESTOBY, M. (1991). Genomic imprinting in endosperm: Its effect on seed development in crosses between species, and between different ploidies of the same species, and its implications for the evolution of apomixis. *Philos. Trans. R. Soc. Lond.* **333** 1–13.

HANSON, R. L., KOBES, S., LINDSAY, R. S. and KMOWLER, W. C. (2001). Assessment of parent-of-origin effects in linkage analysis of quantitative traits. *Am. J. Hum. Genet.* **68** 951–962.

HARRIS, D. L. (1964). Genotypic covariances between inbred relatives. *Genetics* **50** 1319–1348.

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

KERMICLE, J. L. (1970). Dependence of the *R*-mottled aleurone phenotype in maize on the modes of sexual transmission. *Genetics* **66** 69–85.

KINOSHITA, K., YADEGARI, M., HARADA, J. J., GOLDBERG, R. B. and FISHCHER, R. L. (1999). Imprinting of the *MEDEA polycomb* gene in the *Arabidopsis* endosperm. *Plant Cell* **11** 1945–1952.

LI, G. X. and CUI, Y. H. (2009a). A statistical variance components framework for mapping imprinted quantitative trait loci in experimental crosses. *J. Probab. Statist.* Article ID 689489.

LI, G. X. and CUI, Y. H. (2009b). On the limiting distribution of the likelihood ratio test in linkage analysis with the variance components model. Unpublished manuscript.

LUND, G., MESSING, J. and VIOTTI, A. (1995). Endosperm-specific demethylation and activation of specific alleles of *alpha*-tubulin genes of *Zea mays* L. *Mol. Gen. Genet.* **246** 716–722.

LYNCH, M. and WALSH, B. (1998). *Genetics and Analysis of Quantitative Traits.* Sinauer, Sunderland, MA.

MALÉCOT, G. (1948). *Les mathématiques del'hérédité*. Masson et Cie, Paris.

PFEIFER, K. (2000). Mechanisms of genomic imprinting. *Am. J. Hum. Genet.* **67** 777–787.

PLACKETT, R. L. (1954). A reduction formula for normal multivariate integrals. *Biometrika* **41** 351–360. MR0065047

SELF, S. G. and LIANG, K. Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions *J. Amer. Statist. Assoc.* **82** 605–610. MR0898365

SHAPIRO, A. (1985). Asymptotic distribution of test statistics in the analysis of moment structures under inequality constraints. *Biometrika* **72** 133–144. MR0790208

SHETE, S., ZHOU, X. and AMOS, C. I. (2003). Genomic imprinting and linkage test for quantitative trait loci in extended pedigrees. *Am. J. Hum. Genet.* **73** 933–938.

WOLF, J., CHEVERUD, J., ROSEMAN, C. and HAGER, R. (2008). Genome-wide analysis reveals a complex pattern of genomic imprinting in mice. *PLoS Genetics* **4**.

XIE, C., GESSLER, D. D. G. and XU, S. (1998). Combining different line crosses for mapping quantitative trait loci using the identical by descent-based variance component method. *Genetics* **149** 1139–1146.

XU, S. and ATCHLEY, W. R. (1995). A random model approach to interval mapping of quantitative trait loci. *Genetics* **141** 1189–1197.

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

DEPARTMENT OF STATISTICS AND PROBABILITY
MICHIGAN STATE UNIVERSITY
EAST LANSING, MICHIGAN 48824
USA
E-MAIL: ligengxi@stt.msu.edu
cui@stt.msu.edu