Hindawi Publishing Corporation
Journal of Probability and Statistics
Volume 2009, Article ID 689489, 27 pages
doi:10.1155/2009/689489

Research Article

A Statistical Variance Components Framework for Mapping Imprinted Quantitative Trait Locus in Experimental Crosses

Gengxin Li and Yuehua Cui

Department of Statistics & Probability, Michigan State University, East Lansing, MI 48824, USA

Correspondence should be addressed to Yuehua Cui, cui@stt.msu.edu

Received 3 October 2008; Accepted 22 January 2009

Recommended by Rongling Wu

Current methods for mapping imprinted quantitative trait locus (iQTL) with inbred line crosses assume fixed QTL effects. When an iQTL segregates in experimental line crosses, combining different line crosses with similar genetic background can improve the accuracy of iQTLs inference. In this article, we develop a general interval-based statistical variance components framework to map iQTLs underlying complex traits by combining different backcross line crosses. We propose a new iQTL variance partition method based on the nature of marker alleles shared identical-by-decent (IBD) in inbred lines. Maternal effect is adjusted when testing imprinting. Efficient estimation methods with the maximum likelihood and the restricted maximum likelihood are derived and compared. Statistical properties of the proposed mapping strategy are evaluated through extensive simulations under different sampling designs. An extension to multiple QTL analysis is given. The proposed method will greatly facilitate genetic dissection of imprinted complex traits in inbred line crosses.

1. Introduction

The genetic architecture of complex phenotypes in agriculture, evolution, and biomedicine is generally complex involving a network of multiple genetic and environmental factors that interact with one another in complicated ways [1]. The development of molecular markers makes it possible to identify genetic locus (i.e., quantitative trait locus (QTL)) underlie various traits of interest. Genetic designs with controlled crosses are generally pursued to generate mapping populations aimed to identify QTL underlying the variation of phenotypes. Statistical method for QTL mapping with experimental crosses dates back to the seminal work of Lander and Botstein [2]. Various extensions have been developed since then (e.g., [3, 4]).

For a diploid organism, the expression products of most functional regions from each one of a chromosome pair are equal. A broken of this equivalence, that is, nonequivalent genetic contribution of each parental genome to offspring phenotype, can result in *genomic imprinting*, a phenomenon also called parent-of-origin effect [5]. Since its discovery, imprinting-like phenomena have been commonly observed in mammals and seed plants (reviewed by Burt and Trivers [6]). However, statistical methods for identifying imprinted genes have not been extensively studied and well developed.

The imprinted inheritance violates the Mendelian theory and brings challenges in statistical modeling. Currently, there are two frameworks in mapping imprinted genes. One is based on the random effect model with pedigree-based natural population such as humans. Hanson et al. [7] first proposed a variance components framework by partitioning the additive variance component as two parts, a component due to maternal gene and a component due to paternal gene. The variance component method is developed based on the identical-by-decent (IBD) idea in which the expression of the gene for a pair of individuals is expected to be similar if they share alleles IBD. Liu et al. [8] recently applied the model to map iQTL underlying canine hip dysplasia in a structured canine population. However, the current IBD-based variance components method for mapping imprinted genes assumes noninbreeding population. Their applications are immediately limited with fully or partially inbreeding population such as the controlled inbreeding design in plants and animals. With inbred mapping population in humans, Abney et al. [9] proposed a method to estimate variance components of quantitative traits. However, the extension of the method to map imprinted gene is not straightforward. No variance components method has been proposed to map imprinted genes with inbred population in the literature.

Another general framework for mapping imprinted genes is based on the fixed-effect model in which the effects of genetic factors are considered as fixed. A number of studies were proposed under this framework for mapping imprinted QTL (iQTL) with controlled crosses of outbred parents [10–12]. One potential limitation of these methods is that allelic heterozygosity at a locus between two outbred parents could cause confounding effects for genomic imprinting. The genetic differences detected by such a fixed-effect model could be caused by allelic heterozygosity of the parents rather than the imprinted effect of iQTL [13]. A natural alternative for the mapping population is the inbred lines. Fixed-effect models based on backcross (BC) and F_2 population were recently proposed under the maximum likelihood framework [14–17]. When inbred lines are used, Xie et al. [18] pointed out that it is more meaningful to inference QTL effect by its variance rather than by the allele substitution effect. The QTL variance is generally calculated conditional on the cross, and it, as a variable, is different from one cross to another [18]. In a single-line cross, the estimated QTL variance cannot be simply extended to a statistical inference space beyond that [18]. Multiple parental lines are needed for QTL variance inference. A solution to this is to combine data from multiple line crosses [18]. An IBD-based variance component method was proposed by Xie et al. [18] with multiple line crosses. Extension of the IBD-based variance component method with multiple line crosses to iQTL mapping has not been studied.

Motivated by the limitations of current methods aforementioned and by the pressing need for efficient iQTL mapping procedure, in this article, we propose a statistical variance components framework for iQTL mapping by combining data from multiple inbred line crosses. The proposed model is robust in iQTL variance inference by extending the iQTL inference space from single-line cross to multiple line crosses. A parent-specific IBD sharing partition method is proposed by considering the inbreeding structure in line crosses. As discussed by Cui in [14], the phenotype of an offspring is not only controlled by its own genetic profiles, but also controlled by maternal genotype. The effect of maternal genotype on the phenotype of her offspring, termed maternal effect, is one potential source of confounding effect in the inference of genomic imprinting. The existence of such parental effect may lead to incorrect interpretations of imprinting when they are not properly accounted for in the analysis. Parameters that model the maternal effect are also included and adjusted when testing imprinting.

With the developed model, we propose an interval-based method for genome-wide scan and testing of iQTL. Both maximum likelihood (ML) and restricted maximum likelihood (REML) methods are proposed and compared for parameter estimation and power analysis. An extension to multiple QTL is also proposed in which the multiple QTL model provides improved resolution for QTL inference. Extensive simulations are conducted to compare the performance of the proposed model under different sampling designs with different combinations of family and offspring size. Coparisons of the ML and REML methods, single QTL and multiple QTL methods are discussed. The proposed method provides a general framework in iQTL mapping with multiple line crosses and has significant implications in real application.

2. Statistical Methods

2.1. Genetic Design

The dissection of imprinting effects in line crosses depends on appropriate mating designs, where the allele parental origin can be traced and distinguished. Most commonly used inbred line crosses are the backcross, F_2, and recombinant inbred line (RIL). Reciprocal backcross design has been proposed in iQTL mapping [14, 16]. Considering parental origin of an allele, we use the subscripts m and f to refer to an allele inherited from the maternal and paternal parents, respectively. The merit of a backcross design is that two reciprocal heterozygotes in offsprings, $A_m a_f$ and $a_m A_f$, can be distinguished and their mean effects can be estimated and tested to assess imprinting [14, 16]. While all individuals in an F_2 segregation population share the same parental information, theoretically it is impossible to distinguish the phenotypic distribution of $A_m a_f$ and $a_m A_f$ without extra information. Considering sex-specific recombination rates, Cui et al. [15] recently developed an imprinting model by incorporating this information into an interval mapping framework. No study has been reported to use RILs for iQTL mapping.

The methods proposed in Cui [14] and Cui et al. [16] are fixed-effects QTL models where the effects of an iQTL are considered as fixed. While only four backcross families are considered, when extending to multiple backcross families, the inference of iQTL variance calculation is less efficient. The variance components method, initially proposed in human linkage analysis [19], offers a powerful alternative in assessing genomic imprinting [7]. In this paper, we will extend the variance components method to inbred line populations by combining different backcross lines to map iQTL.

A typical backcross design often starts with the cross between one of the parental lines and their F_1 progeny to create a segregation population. Then large number of offsprings are collected for QTL mapping. When imprinting effect is considered, reciprocal backcrosses are needed. A basic design framework is illustrated in Table 1 in Cui [14]. The two reciprocal backcrosses are treated as the base mapping units. Multiple backcross families are sampled based on these four crosses. For simplicity, we sample equal number of families for each backcross category. For example, a sample of 8 families would require two of each of the four backcrosses. Noted that the variance components method assesses the degree of allele sharing among siblings. When it is applied to inbred line crosses, each backcross population is considered as one family and different families are considered as independent. For fixed total sample size, one issue is to assess whether we should sample large number of families each with small offspring size or small number of families each with large offspring size. For example, to sample 400 individuals, shall we sample 4 backcross families each with 100 offsprings or 100 families each with 4 progenies or other sampling strategies? The choice of optimal designs is intensively evaluated through simulations.

2.2. The Mixed-Effect Variance Components Model

Suppose there is a putative QTL with two segregating alleles Q and q, located in an interval responsible for the variation of a quantitative trait. The phenotype, y_{ik}, for individual i measured in backcross family $k(=1,\ldots,K)$ can be written as a linear function of QTL, polygene, and environmental effects:

$$y_{ik} = \mu + a_{ik} + G_{ik} + e_{ik}, \quad k = 1,\ldots,K;\ i = 1,\ldots,n_k, \tag{2.1}$$

where n_k is the number of offsprings in the kth backcross family; μ denotes the overall mean; a_{ik} is the random additive effect of the major monogenic QTL assuming normal distribution with mean zero; G_{ik} is the polygenic effect that reflects the effects of unlinked genes and is assumed to be normally distributed with mean zero; $e_{ik} \sim N(0,\sigma_e^2)$ is the random environmental error uncorrelated to other effects. The phenotypic variance-covariance for the kth family can be expressed as

$$\Sigma_k = \Pi_k\sigma_a^2 + \mathbf{\Phi}_g\sigma_g^2 + \mathbf{I}\sigma_e^2, \tag{2.2}$$

where σ_a^2 and σ_g^2 are the additive and polygene variances; Π_k is a matrix containing the proportion of marker alleles shared IBD for individuals in the kth backcross family; $\mathbf{\Phi}_g$ is a matrix of the expected proportion of alleles shared IBD; $\mathbf{I}$ is the identity matrix. The calculation of the IBD sharing matrix with inbred lines can be found in Xie et al. [18] which is based on the Malécot coefficient of coancestry [20].

Noted that a backcross offspring with genotype $Q_m q_f$ may be obtained by the $QQ\times Qq$ or the $Qq \times QQ$ cross. When there is a significant maternal effect, the mean expression for genotype $Q_m q_f$ may be different depending on whether its maternal parents carrying QQ or Qq genotype. As described in Cui [14], maternal effect is one source of potential confounding factor for genomic imprinting. It should be appropriately modeled and adjusted when testing imprinting. Here, we model the cytoplasmic maternal effects as fixed effects, and the overall mean μ is replaced by μ_k which models the maternal effect of the kth distinct backcross family.

To accommodate parent-of-origin effects, the QTL additive effect (a) can be partitioned as two terms:

(1) a component that reflects the influence of the QTL carried on the maternally derived chromosome (a_m);

(2) a component that reflects the influence of the QTL carried on the paternally derived chromosome (a_f).

The model that accommodates the parent-specific effects can be expressed as

$$y_{ik} = \mu_k + a_{ikm} + a_{ikf} + G_{ik} + e_{ik}, \quad k = 1, \ldots, K; \; i = 1, \ldots, n_k. \tag{2.3}$$

Note that the proposed design contains three distinct maternal parent genotypes. Thus the k maternal effects indexed by μ_k can be compressed to three distinct maternal effects, instead of k terms. For data vector $\mathbf{y}$ in family k, the above model can be reexpressed as

$$\mathbf{y}_k = X_k\beta + \mathbf{a}_{km} + \mathbf{a}_{kf} + \mathbf{G}_k + \mathbf{e}_k, \quad k = 1, \ldots, K, \tag{2.4}$$

where X_k is an indicator matrix corresponding to the kth backcross family and β contains parameters associated with the three maternal effects; $\mathbf{a}_{km} \sim N(\mathbf{0}, \Pi_{m|k}\sigma_m^2)$, $\mathbf{a}_{kf} \sim N(\mathbf{0}, \Pi_{f|k}\sigma_f^2)$, $\mathbf{G}_k \sim N(\mathbf{0}, \mathbf{\Phi}_g\sigma_g^2)$, $\mathbf{e}_k \sim N(\mathbf{0}, \mathbf{I}\sigma_e^2)$, where $\Pi_{m|k}$ and $\Pi_{f|k}$ are matrices containing the proportion of marker alleles shared IBD that are derived from the mother and father, respectively; $\mathbf{\Phi}_g$ is a matrix of the expected proportion of alleles shared IBD; $\mathbf{I}$ is the identity matrix; σ_m^2 and σ_f^2 are the variance of alleles inherited from the maternal and paternal parents, respectively.

With noninbreeding mapping population, Hanson et al. [7] expressed the phenotypic variance-covariance for the kth family as

$$\Sigma_k = \Pi_{m|k}\sigma_m^2 + \Pi_{f|k}\sigma_f^2 + \mathbf{\Phi}_g\sigma_g^2 + \mathbf{I}\sigma_e^2. \tag{2.5}$$

However, for an inbred mapping population, this IBD-based variance partition method cannot be directly applied. New method considering the inbreeding structure is needed.

2.3. Parent-Specific Allele Sharing and Covariances between Two Inbreeding Full-Sibs

Before we get the phenotypic variance-covariance of a pair of individuals i and j, let us first consider the parent-specific allele sharing status. Within each BC family, there are two alleles segregating at each locus. Because of inbreeding, the IBD values between two backcross individuals are different from those calculated from outbred full-sibs. Consider two sibs i and j in the kth backcross family. Without considering allelic parental origin, Xie et al. [18] proposed to calculate the IBD value at a QTL as

$$\pi_{ij} = 2\theta_{ij} = \begin{cases} 2 & \text{for } QQ - QQ, \\ 1 & \text{for } QQ - Qq \text{ or } Qq - Qq \end{cases} \tag{2.6}$$

with θ_{ij} being the Malécot coefficient of coancestry [20]. Thus, for an inbred population, π_{ij} is not the actual IBD value between individuals i and j, rather interpreted as twice the coefficient of coancestry [18, 21]. For individuals with itself,

$$\pi_{ii} = 1 + F_i = \begin{cases} 2 & \text{for } QQ - QQ, \\ 1 & \text{for } Qq - Qq, \end{cases} \tag{2.7}$$

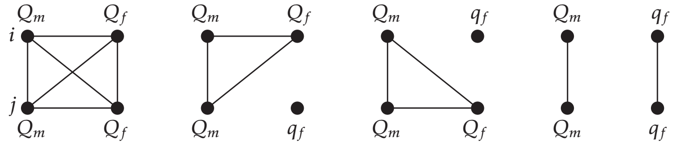

Figure 1: Possible alleles-shared IBD for individuals i and j in inbreeding backcross families. Lines indicate alleles-shared IBD.

where F_i is the inbreeding coefficient for individual i at the QTL. The elements in $\mathbf{\Phi}_g$ matrix are just the expected values of π_{ij} and π_{ii} which are $\phi_{ij} = 5/4$ and $\phi_{ii} = 3/2$ [18].

When allelic parental origin is considered, the IBD sharing matrix can also be calculated based on the coefficient of coancestry. By definition, the coefficient of coancestry is defined as the probability that two randomly drawn alleles from individuals i and j are identical by descent. Figure 1 displays possible alleles shared IBD for sibs drawn in backcross families. Consider two backcross individuals i (with two alleles A_{i_m} and A_{i_f}) and j (with two alleles A_{j_m} and A_{j_f}). Define θ_{ij} as the coefficient of coancestry between individuals i and j. By definition, θ_{ij} can be calculated as

$$\begin{aligned}\theta_{ij} &= \frac{1}{4}\left\{\Pr\left(A_{i_m} = A_{j_m}\right) + \Pr\left(A_{i_m} = A_{j_f}\right) + \Pr\left(A_{i_f} = A_{j_m}\right) + \Pr\left(A_{i_f} = A_{j_f}\right)\right\} \\ &= \frac{1}{4}\left(\theta_{i_m j_m} + \theta_{i_m j_f} + \theta_{i_f j_m} + \theta_{i_f j_f}\right),\end{aligned} \tag{2.8}$$

where $\theta_{i\cdot j\cdot}$ can be interpreted as the allelic kinship coefficient, that is, the probability that a randomly chosen allele from individual i is IBD to a randomly chosen allele from individual j. Note that the two terms $\theta_{i_m j_f}$ and $\theta_{i_f j_m}$ are not distinguishable. However, their sum is unique and therefore the two terms can be combined as one single term, denoted as $\theta_{i_m/j_f}(= \theta_{i_m j_f} + \theta_{i_f j_m})$. After the manipulation, the coefficient of coancestry for individuals i and j can be expressed as $\theta_{ij} = (1/4)(\theta_{i_m j_m} + \theta_{i_m/j_f} + \theta_{i_f j_f})$ which is composed of three components.

Following Xie et al. [18], the alleles shared IBD between individuals i and j can be expressed as

$$\begin{aligned}\pi_{ij} &= 2\theta_{ij} \\ &= \frac{1}{2}\left(\theta_{i_m j_m} + \theta_{i_m/j_f} + \theta_{i_f j_f}\right) \\ &= \pi_{i_m j_m} + \pi_{i_m/j_f} + \pi_{i_f j_f},\end{aligned} \tag{2.9}$$

where $\pi_{i_m j_m} = (1/2)\theta_{i_m j_m}$ and $\pi_{i_f j_f} = (1/2)\theta_{i_f j_f}$ are the alleles shared IBD derived from the mother and father, respectively; $\pi_{i_m/j_f} = (1/2)\theta_{i_m/j_f}$ is the alleles shared IBD due to alleles cross sharing, a special case for inbreeding sibs. Without inbreeding, π_{i_m/j_f} takes value of zero.

For completely inbreeding population, the inbreeding coefficient F_i is 1 if alleles inherited from both parents are the same since these alleles can be traced back to the same grandparent. For example, for an individual with genotype $Q_m Q_f$, $\Pr(Q_m = Q_f) = 1$ since both alleles Q_m and Q_f are inherited from the same grandparent. Therefore, for individuals with itself, $\pi_{ii} = 1 + F_i$ would be the same as π_{ij} $(i \neq j)$ when i and j carry the same genotypes. The expected proportion of alleles shared IBD ϕ_{ij} can also be calculated.

Thus, the proportion of alleles-shared IBD can be partitioned as three components for inbreeding sibs, rather than two components considering parent-of-origin effects proposed by Hanson et al. [7]. To further illustrate the idea, we use one backcross family to demonstrate the derivation. A full list of possible IBD sharing values for the two reciprocal backcrosses is given in Table 1. Considering a backcross family initiated with the $Qq \times QQ$ cross. Randomly selecting two individuals i and j with genotype Q_mQ_f and Q_mQ_f, the Malécot coefficient of coancestry can be calculated as

$$\begin{aligned}
\pi_{ij} &= 2\theta_{ij} \\
&= \frac{1}{2}\left\{\Pr\left(Q_{im} = Q_{jm}\right) + \Pr\left(Q_{im} = Q_{jf}\right) + \Pr\left(Q_{if} = Q_{jm}\right) + \Pr\left(Q_{if} = Q_{jf}\right)\right\} \\
&= \frac{1}{2}[1 + 1 + 1 + 1] \\
&= 2.
\end{aligned} \tag{2.10}$$

Thus, $\pi_{i_m j_m} = \pi_{i_f j_f} = 0.5$ and $\pi_{i_m/j_f} = 1$. For sib pairs i (with genotype Q_mQ_f) and j (with genotype Q_mq_f), $\pi_{i_m j_m} = 0.5$, $\pi_{i_f j_f} = 0$ and $\pi_{i_m/j_f} = 0.5$, and $\pi_{ij} = 1$ which is the same as given in (2.6) without considering parent-of-origin partition.

Considering the allelic sharing status in a complete inbreeding population, the relationship between the maternal and paternal alleles is no longer independent if the two alleles are in identical form. There exists a covariance term (denoted as σ_{mf}^2) due to alleles cross sharing for two inbreeding full-sibs when calculating the phenotypic variance. Corresponding to the partition of the IBD-sharing considering allelic parental origin, the major QTL additive variance component can be partitioned into three components, that is, σ_f^2, σ_m^2, and σ_{mf}^2, in which σ_{mf}^2 can be interpreted as the covariance due to alleles cross sharing in inbreeding families. Thus, the trait covariance between two individuals i and j can be expressed as

$$\operatorname{Cov}\left(y_i, y_j\right) = \pi_{i_m j_m}\sigma_m^2 + \pi_{i_f j_f}\sigma_f^2 + \pi_{i_m/j_f}\sigma_{mf}^2 + \phi_{ij}\sigma_g^2 + I_{ij}\sigma_e^2, \tag{2.11}$$

where I_{ij} is an indicator variable taking value 1 if $i = j$ and 0 if $i \neq j$. The variance-covariance matrix for a phenotypic vector in the kth backcross family can then be expressed as

$$\Sigma_k = \Pi_{m|k}\sigma_m^2 + \Pi_{m/f|k}\sigma_{mf}^2 + \Pi_{f|k}\sigma_f^2 + \mathbf{\Phi}_g\sigma_g^2 + \mathrm{I}\sigma_e^2, \tag{2.12}$$

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

For noninbreeding sib pairs with random mating, $\pi_{i_m/j_f} = 0$ and hence $\operatorname{Cov}(a_m, a_f) = 0$. Model (2.12) reduces to $\sum_k = \Pi_{m|k}\sigma_m^2 + \Pi_{f|k}\sigma_f^2 + \mathbf{\Phi}_g\sigma_g^2 + \mathbf{I}\sigma_e^2$, the same as the variance components partition model considering parent-of-origin effects given by Hanson et al. [7].

Table 1: The IBD sharing coefficients for full-sib pairs in a reciprocal backcross design considering allelic parental origin.

	Offspring	Parent-specific IBD sharing						Total IBD	
Backcross	genotype	π_{mm}		π_{ff}		$\pi_{m/f}$		π	
		Q_mQ_f	Q_mq_f	Q_mQ_f	Q_mq_f	Q_mQ_f	Q_mq_f	Q_mQ_f	Q_mq_f
$QQ \times Qq$	Q_mQ_f	0.5	0.5	0.5	0	1	0.5	2	1
	Q_mq_f	0.5	0.5	0	0.5	0.5	0	1	1
		Q_mQ_f	q_mQ_f	Q_mQ_f	q_mQ_f	Q_mQ_f	q_mQ_f	Q_mQ_f	q_mQ_f
$Qq \times QQ$	Q_mQ_f	0.5	0	0.5	0.5	1	0.5	2	1
	q_mQ_f	0	0.5	0.5	0.5	0.5	0	1	1
		q_mQ_f	q_mq_f	q_mQ_f	q_mq_f	q_mQ_f	q_mq_f	q_mQ_f	q_mq_f
$qq \times Qq$	q_mQ_f	0.5	0.5	0.5	0	0	0.5	1	1
	q_mq_f	0.5	0.5	0	0.5	0.5	1	1	2
		Q_mq_f	q_mq_f	Q_mq_f	q_mq_f	Q_mq_f	q_mq_f	Q_mq_f	q_mq_f
$Qq \times qq$	Q_mq_f	0.5	0	0.5	0.5	0	0.5	1	1
	q_mq_f	0	0.5	0.5	0.5	0.5	1	1	2

2.4. Likelihood Function and Parameter Estimation

Assuming multivariate normality, the density function of observing a particular vector of data $\mathbf{y}$ for family k is given by

$$f(\mathbf{y}_k; \mu_k, \Sigma_k) = \frac{1}{(2\pi)^{n_k/2}|\Sigma_k|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{y}_k - \mu_k)^{\mathrm{T}}\Sigma_k{}^{-1}(\mathbf{y}_k - \mu_k)\right], \tag{2.13}$$

where $\mathbf{y}_k = (y_{1k}, \ldots, y_{n_k k})^T$ is an $n_k \times 1$ vector of phenotypes for the kth backcross family, and n_k is the kth backcross family size. The overall log likelihood function for K independent backcross families is give by

$$\ell = \sum_{k=1}^{K} \log[f(\mathbf{y}_k; \mu_k, \Sigma_k)]. \tag{2.14}$$

Note that the maternal effect μ_k is the same for families with the same maternal genotype. Thus, only three maternal effects need to be estimated. Two commonly used methods can be applied to estimate parameters in a mixed effects model, the ML method and the REML method. Both methods have been applied in genetic linkage analysis in a variance components model framework [19, 22]. In general, ML estimators tend to be downwardly biased given that it does not account for the loss in degrees of freedom resulted from estimation of the fixed effects [23]. The REML is based on a linear transformation of the data such that the fixed effects are eliminated from the model, hence it provides less-biased estimators. Even though standard softwares such as SAS have standard procedures to estimate parameters for a mixed effects model, the estimation for the proposed model cannot be directly fitted into a standard software. The estimation procedures for the two methods are detailed here.

2.4.1. The ML Estimation

The phenotype vector in the kth backcross family follows a multivariate normal distribution, that is, $\mathbf{y}_k \sim MVN(X_k\beta, \Sigma_k)$. Parameters that need to be estimated are $\Omega = (\beta, \sigma_m^2, \sigma_f^2, \sigma_{mf}^2, \sigma_g^2, \sigma_e^2)$ with $\beta = (\mu_1, \mu_2, \mu_3)^T$.

Define $\sigma^2 = \sigma_m^2 + \sigma_f^2 + \sigma_{mf}^2 + \sigma_g^2 + \sigma_e^2$, $h_m^2 = (\sigma_m^2/\sigma^2)$, $h_f^2 = (\sigma_f^2/\sigma^2)$, $h_{mf}^2 = (\sigma_{mf}^2/\sigma^2)$, $h_g^2 = (\sigma_g^2/\sigma^2)$, and $h_e^2 = 1 - h_m^2 - h_f^2 - h_{mf}^2 - h_g^2$. σ^2 is the total phenotypic variance and hence h_m^2 and h_f^2 can be considered as the heritability of maternaland paternal alleles, $h_m^2 + h_f^2 + h_{mf}^2$ is the total genetic heritability due to the major QTL, h_g^2 is the polygene heritability, and $h^2 = h_m^2 + h_f^2 + h_{mf}^2 + h_g^2$ is the overall heritability. The phenotypic variance-covariance between any two individuals i and j in the kth backcross family can then be reexpressed as

$$\operatorname{Var}\begin{pmatrix} y_{ik} \\ y_{jk} \end{pmatrix} = \sigma^2 H_{ij|k}, \tag{2.15}$$

where

$$H_{ij|k} = \begin{pmatrix} \delta_i & \delta_{ij} \\ \delta_{ij} & \delta_j \end{pmatrix} \tag{2.16}$$

with $\delta_i = \pi_{i_m i_m} h_m^2 + \pi_{i_f i_f} h_f^2 + \pi_{i_m/i_f} h_{mf}^2 + \phi_{ii} h_g^2 + h_e^2$; δ_j is defined similarly; $\delta_{ij} = \pi_{i_m j_m} h_m^2 + \pi_{i_f j_f} h_f^2 + \pi_{i_m/j_f} h_{mf}^2 + \phi_{ij} h_g^2$.

If there are n_k sibs in each backcross family, $H_k = \{H_{ij|k}\}_{n_k \times n_k}$ is simply an $n_k \times n_k$ matrix. Instead of estimating $\Omega = (\beta, \sigma_m^2, \sigma_f^2, \sigma_{mf}^2, \sigma_g^2, \sigma_e^2)$, we can estimate $\Omega = (\beta, \sigma^2, h_m^2, h_f^2, h_{mf}^2, h_g^2)$ and solve above equations to get the original variance estimates. Now, the log-likelihood can be expressed as

$$\ell(\Omega) = \sum_{k=1}^{K} \log[f(\mathbf{y}_k \mid \Omega)] \propto -\sum_{k=1}^{K} \left\{ \frac{n_k}{2} \log \sigma^2 - \frac{1}{2} \log|H_k| - \frac{1}{2\sigma^2} (\mathbf{y}_k - X_k\beta)' H_k^{-1} (\mathbf{y}_k - X_k\beta) \right\}. \tag{2.17}$$

Maximizing likelihood (2.14) is equivalent to maximize (2.17). Here, we take an iterated estimation procedure to estimate the parameters contained in Ω. For given values of $h_m^2, h_f^2, h_{mf}^2, h_g^2$, we can get the maximum likelihood estimates (MLEs) of parameters (β, σ^2) by setting the partial derivative of the log-likelihood function (2.17) to zero, that is,

$$\begin{gathered} \widehat{\beta} = \sum_{k=1}^{K} \left(X_k^T H_k^{-1} X_k\right)^{-1} \left(X_k^T H_k^{-1} \mathbf{y}_k\right), \\ \widehat{\sigma}^2 = \frac{1}{\sum_{k=1}^{K} n_k} \sum_{k=1}^{K} \left(\mathbf{y}_k - X_k\widehat{\beta}\right)^T H_k^{-1} \left(\mathbf{y}_k - X_k\widehat{\beta}\right). \end{gathered} \tag{2.18}$$

It can be seen that $\widehat{\beta}$ and $\widehat{\sigma}^2$ are functions of h_m^2, h_f^2, h_{mf}^2, and h_g^2. Plug the updated parameter values for β and σ^2 into likelihood equation (2.17), the log-likelihood function can be simplified as

$$\ell(\Omega) = \sum_{k=1}^{K} \log\left[f\left(\mathbf{y}_k \mid \Omega\right)\right] \propto -\sum_{k=1}^{K} \frac{n_k}{2} \log \widehat{\sigma}^2 - \frac{1}{2}\sum_{k=1}^{K} \log|H_k|. \tag{2.19}$$

The simplex algorithm [24] can be applied to maximize the function (2.19) with respect to parameters h_m^2, h_f^2, h_{mf}^2 and h_g^2 subject to the the constraints that $0 \le h_m^2, h_f^2, h_{mf}^2, h_g^2 \le 1$ and $0 \le h^2 \le 1$.

To guarantee a positive definite covariance matrix when searching for these heritability values over the constraint parameter space, a reparameterization technique is adopted [25]. Taking $\delta_{ij} = h^2(\pi_{i_m j_m}\gamma_m^2 + \pi_{i_f j_f}\gamma_f^2 + \pi_{i_m/j_f}\gamma_{mf}^2 + \phi_{ij}\gamma_g^2)$, where $\gamma_m^2 = h_m^2/h^2$, $\gamma_f^2 = h_f^2/h^2$, $\gamma_{mf}^2 = h_{mf}^2/h^2$, $\gamma_g^2 = h_g^2/h^2$, and $h^2 = h_m^2 + h_f^2 + h_{mf}^2 + h_g^2$. We now have four new unknowns with the constraints: $0 \le h^2 \le 1$, $\gamma_m^2 + \gamma_f^2 + \gamma_{mf}^2 + \gamma_g^2 = 1$, and $\gamma_m^2, \gamma_f^2, \gamma_{mf}^2, \gamma_g^2 \ge 0$.

The new constraints can be easily satisfied by a reparameterization technique. Let u, v_m, v_f, v_{mf}, and v_g be any real numbers. Estimating h^2, γ_m^2, γ_f^2, γ_{mf}^2, and γ_g^2 can be done by maximizing the likelihood function (2.19) via searching through the real domain space with respect to u, v_m, v_f, v_{mf}, and v_g with the reparameterization

$$\begin{aligned} h^2 &= \frac{e^u}{1+e^u}, \\ \gamma_m^2 &= \frac{e^{v_m}}{e^{v_m} + e^{v_f} + e^{v_{mf}} + e^{v_g}}, \\ \gamma_f^2 &= \frac{e^{v_f}}{e^{v_m} + e^{v_f} + e^{v_{mf}} + e^{v_g}}, \\ \gamma_{mf}^2 &= \frac{e^{v_{mf}}}{e^{v_m} + e^{v_f} + e^{v_{mf}} + e^{v_g}}, \\ \gamma_g^2 &= \frac{e^{v_g}}{e^{v_m} + e^{v_f} + e^{v_{mf}} + e^{v_g}}. \end{aligned} \tag{2.20}$$

MLEs of h^2, γ_m^2, γ_f^2, γ_{mf}^2, and γ_g^2 can be obtained through the estimated values for u, v_m, v_f, v_{mf}, and v_g according to the invariance property of MLEs. These estimated MLEs are used to update h^2, h_m^2, h_f^2, h_{mf}^2, and h_g^2, and hence σ^2 and β. The iteration steps continue until converge.

2.4.2. The REML Estimation

The REML method was first proposed by Patterson and Thompson [26]. This method has been broadly applied to estimate variance components in a mixed-effect model framework.

Taking $\Omega = (\beta, \Theta)$, where $\Theta = (\sigma_m^2, \sigma_f^2, \sigma_{mf}^2, \sigma_g^2, \sigma_e^2)$. The REML method starts with maximizing the following likelihood function:

$$\ell^*(\Theta) = \sum_{k=1}^{K} \log\left[f(\mathbf{y}_k \mid \Theta)\right] = -\frac{1}{2}\sum_{k=1}^{K}\left\{\log\left|\Sigma_k\right| + \log\left(\left|X_k'\Sigma_k^{-1}X_k\right|\right) + \mathbf{y}_k' P_k \mathbf{y}_k\right\}, \tag{2.21}$$

where $P_k = \Sigma_k^{-1} - \Sigma_k^{-1}X_k\left(X_k'\Sigma_k^{-1}X_k\right)^{-1}X_k'\Sigma_k^{-1}$. We can combine all family data together as one $N \times 1$ vector denoted as $\mathbf{y}$, where $N = \sum_{k=1}^{K} n_k$. All the X_k and the variance-covariance matrix $\sum_k$ corresponding to each family can be combined. The log-likelihood function for the combined data is expressed as

$$\ell^*(\Theta) = \log\left[f(\mathbf{y} \mid \Theta)\right] = -\frac{1}{2}\left\{\log\left|\Sigma\right| + \log\left(\left|X'\Sigma^{-1}X\right|\right) + \mathbf{y}'P\mathbf{y}\right\}, \tag{2.22}$$

where $\sum$ is a block diagonal matrix with the kth diagonal block $\sum_k$ corresponding to the kth family and off-diagonal blocks being zeros; P is also a block diagonal matrix with block elements given by P_k. The dimension of $\sum$ is $N \times N$. With this combination, we develop the following REML estimation procedure.

We apply the Fisher scoring algorithm to estimate the unknowns, which has the form

$$\Theta^{(t+1)} = \Theta^{(t)} + \mathcal{I}\left(\Theta^{(t)}\right)^{-1}\frac{\partial \ell^*(\Theta)}{\partial \Theta} \mid \Theta^{(t)}, \tag{2.23}$$

where $\mathcal{I}(\Theta^{(t)})$ is the Fisher information matrix evaluated at $\Theta^{(t)}$ which can be expressed as

$$\mathcal{I}\begin{pmatrix} \sigma_m^2 \\ \sigma_f^2 \\ \sigma_{mf}^2 \\ \sigma_g^2 \\ \sigma_e^2 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} \operatorname{tr}\left(P\Pi_m P\Pi_m\right) & \operatorname{tr}\left(P\Pi_m P\Pi_f\right) & \operatorname{tr}\left(P\Pi_m P\Pi_{m/f}\right) & \operatorname{tr}\left(P\Pi_m P\Phi_g\right) & \operatorname{tr}\left(P\Pi_m P\right) \\ \operatorname{tr}\left(P\Pi_f P\Pi_m\right) & \operatorname{tr}\left(P\Pi_f P\Pi_f\right) & \operatorname{tr}\left(P\Pi_f P\Pi_{m/f}\right) & \operatorname{tr}\left(P\Pi_f P\Phi_g\right) & \operatorname{tr}\left(P\Pi_f P\right) \\ \operatorname{tr}\left(P\Pi_{m/f} P\Pi_m\right) & \operatorname{tr}\left(P\Pi_{m/f} P\Pi_f\right) & \operatorname{tr}\left(P\Pi_{m/f} P\Pi_{m/f}\right) & \operatorname{tr}\left(P\Pi_{m/f} P\Phi_g\right) & \operatorname{tr}\left(P\Pi_{m/f} P\right) \\ \operatorname{tr}\left(P\Phi_g P\Pi_m\right) & \operatorname{tr}\left(P\Phi_g P\Pi_f\right) & \operatorname{tr}\left(P\Phi_g P\Pi_{m/f}\right) & \operatorname{tr}\left(P\Phi_g P\Phi_g\right) & \operatorname{tr}\left(P\Phi_g P\right) \\ \operatorname{tr}\left(P\Pi_m P\right) & \operatorname{tr}\left(P\Pi_f P\right) & \operatorname{tr}\left(P\Pi_{m/f} P\right) & \operatorname{tr}\left(P\Phi_g P\right) & \operatorname{tr}\left(PP\right) \end{pmatrix}. \tag{2.24}$$

The first derivative of the log-likelihood function ℓ^* with respective to each variance components is given by

$$\begin{aligned}
\frac{\partial \ell^*}{\partial \sigma_m^2} &= -\frac{1}{2}\left(\operatorname{tr}(P\Pi_m) - \mathbf{y}^T P\Pi_m P\mathbf{y}\right),\\
\frac{\partial \ell^*}{\partial \sigma_f^2} &= -\frac{1}{2}\left(\operatorname{tr}(P\Pi_f) - \mathbf{y}^T P\Pi_f P\mathbf{y}\right),\\
\frac{\partial \ell^*}{\partial \sigma_{mf}^2} &= -\frac{1}{2}\left(\operatorname{tr}(P\Pi_{m/f}) - \mathbf{y}^T P\Pi_{m/f} P\mathbf{y}\right),\\
\frac{\partial \ell^*}{\partial \sigma_g^2} &= -\frac{1}{2}\left(\operatorname{tr}(P\Phi_g) - \mathbf{y}^T P\Phi_g P\mathbf{y}\right),\\
\frac{\partial \ell^*}{\partial \sigma_e^2} &= -\frac{1}{2}\left(\operatorname{tr}(PI_N) - \mathbf{y}^T PP\mathbf{y}\right).
\end{aligned} \tag{2.25}$$

The REML estimator of β is the generalized least squares estimator, that is,

$$\widehat{\beta} = \left(X^T\widehat{\Sigma}^{-1}X\right)^{-1}X^T\widehat{\Sigma}^{-1}Y. \tag{2.26}$$

Under the current mapping framework, the likelihood function is very complex and no global maxima is guaranteed. Thus, initial values are very important. In both ML and REML estimation procedures, the estimated values under the null are set as the initial parameters for the alternative model. The additional variance component(s) to be estimated under the alternative model is(are) set to a small positive number. In this way, we can guarantee that the alternative model always produces larger likelihood values than the null model does, and hence positive likelihood ratio value.

2.5. QTL IBD Sharing and Genome-Wide Linkage Scan

The above IBD computation procedure assumes that a putative QTL is located right on a marker. When a QTL is located within an interval, a more efficient approach would be to do an interval scan and to test the imprinting property of QTLs at positions across the entire linkage group. Under the proposed framework, essentially, we need to estimate the proportion of putative QTL alleles shared IBD at every genome position. Here, we propose a method to calculate QTL alleles-shared IBD inside an interval conditional on the flanking markers. The so-called expected conditional IBD values can be derived at each test position as a function of recombination fraction between the two flanking markers, and the one between one flanking marker and the QTL.

We use one backcross initiated with the cross $QQ \times Qq$ as an example to illustrate the idea. For a putative QTL with two alleles Q and q, four QTL genotype pairs $QQ - QQ$, $QQ - Qq$, $Qq - QQ$, and $Qq - Qq$ can be formed. If the QTL genotype is observed, the corresponding QTL alleles-shared IBD can be calculated (see Table 1). In general, the QTL genotype is unobservable, but its conditional distribution can be calculated from the two flanking markers. For individuals i and j with flanking marker genotypes g_i and g_j, let $\pi_{v|G_iG_j}$ be the IBD values calculated at the QTL position between individual i carrying QTL genotype G_i (= 1 or 2 corresponding to QQ or Qq, resp.) and individual j carrying genotype G_j (similarly 1 or 2), where $v = i_m j_m, i_f j_f$ or i_m / j_f. For example, $\pi_{i_m j_m | G_i G_j}$ is the proportion of IBD sharing between individual i carrying QTL genotype G_i and individual j carrying genotype G_j for alleles derived from the mother.

Let $\varphi_{G_i|g_i}$ and $\varphi_{G_j|g_j}$ be the conditional distribution of QTL genotype G_i and G_j for individuals i and j given on the flanking markers g_i and g_j, respectively. This conditional probabilities can be easily calculated and can be found at standard QTL mapping literature (see [27]). The probability to observe $\pi_{v|G_i G_j}$ is just $\varphi_{G_i|g_i}\varphi_{G_j|g_j}$. Thus, the expected IBD values between individuals i and j at the tested QTL position conditioning on the flanking markers g_i and g_j can be calculated as $\widehat{\pi}_v = \mathrm{E}(\pi_{v|G_i G_j}) = \sum_{G_i=1}^{2}\sum_{G_j=1}^{2}\pi_{v|G_i G_j}\varphi_{G_i|g_i}\varphi_{G_j|g_j}$. For the above example, the IBD values derived from the maternal and paternal parents can be calculated as $\widehat{\pi}_{i_m j_m} = \mathrm{E}(\pi_{i_m j_m|G_i G_j}) = 0.5\ \varphi_{1|g_i}\varphi_{1|g_j} + 0.5\ \varphi_{1|g_i}\varphi_{2|g_j} +0.5\ \varphi_{2|g_i}\varphi_{1|g_j} + 0.5\ \varphi_{2|g_i}\varphi_{2|g_j}$ and $\widehat{\pi}_{i_f j_f} = \mathrm{E}(\pi_{i_f j_f|G_i G_j}) = 0.5\ \varphi_{1|g_i}\varphi_{1|g_j} + 0.5\ \varphi_{2|g_i}\varphi_{2|g_j}$. Similarly, we can calculate the conditional expectation of IBD sharing for other backcross families.

Since $\varphi_{G_i|g_i}$ and $\varphi_{G_j|g_j}$ are functions of recombinations, the conditional QTL IBD values vary at different testing positions. Once the estimated IBD matrix is calculated at every 1 or 2 cm on an interval bracketed by two markers throughout the entire genome, a grid search can be done at all 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 plot the likelihood ratio test statistic as a function of testing positions of putative QTL (see details in Section 2.6). The peaks of the profile plot that pass certain significant threshold correspond to the positions of significant QTL.

2.6. Hypothesis Testing

With the estimated parameters using either the ML or REML method, we are interested in testing the existence of QTL across the genome and assess their imprinting mechanism. The first hypothesis is to test the existence of major QTL, termed overall QTL test, which can be formulated as

$$\begin{gathered} H_0 : \sigma_m^2 = \sigma_f^2 = \sigma_{mf}^2 = 0, \\ H_1 : \text{at least one parameter is not zero.} \end{gathered} \tag{2.27}$$

Likelihood ratio (LR) test is applied which is computed between the full (there is a QTL) and the reduced model (there is no QTL) corresponding to H_1 and H_0, respectively. Let $\widetilde{\Omega}$ and $\widehat{\Omega}$ be the estimates of the unknown parameters under H_0 and H_1, respectively. The log-likelihood ratio can be calculated as

$$\mathrm{LR}_1 = -2\left[\log L\left(\widetilde{\Omega} \mid \mathbf{y}\right) - \log L\left(\widehat{\Omega} \mid \mathbf{y}\right)\right]. \tag{2.28}$$

When testing the hypothesis, the polygene and the residual variances are nuisance parameters which are constrained to be nonnegative. The three tested genetic variance components under the null are lied on the boundaries of their alternative parameter spaces. Following Self and Liang [28], when the null is true, LR_1 may asymptotically follows a mixture of χ^2 distribution on $0,\ldots,3$ degrees of freedom (df) with the mixture proportion for the χ_k^2 components being $\binom{3}{k}2^{-3}$. The theoretical distribution can be used to assess significance in linkage scan. However, since there are many point tests across the genome, the pointwise significance value may not guarantee an appropriate genome-wide error rate. Another approach to assess significance is to use nonparametric permutation tests in which the critical threshold value can be empirically calculated on the basis of repeatedly shuffling the relationships between marker genotypes and phenotypes [29]. In simulation studies, we also simulate the null distribution and compare it with the theoretical distribution.

For those detected QTL, the next step is to assess their imprinting property. An identified QTL can be imprinted, completely imprinted, partially imprinted, or not imprinted at all. These can be tested through the following sequential tests. The first imprinting test is to assess whether a QTL shows imprinting effect, which can be done by formulating the following hypotheses:

$$H_0 : \sigma_m^2 = \sigma_f^2 = \sigma^2, \qquad H_1 : \sigma_m^2 \neq \sigma_f^2. \tag{2.29}$$

Rejection of H_0 provides evidence of genomic imprinting and the QTL is called iQTL. Again, likelihood ratio test can be applied in which the log-likelihood ratio test statistics asymptotically follows an χ^2 with one df [7]. We denote the log-likelihood ratio test statistic as $\mathrm{LR}_{\mathrm{imp}}$. If the null is rejected, one would be interested to test if the detected iQTL is completely maternally or paternally imprinted. The corresponding hypotheses can be formulated as

$$H_0 : \sigma_m^2 = 0, \qquad H_1 : \sigma_m^2 \neq 0 \tag{2.30}$$

for testing completely maternal imprinting and

$$H_0 : \sigma_f^2 = 0, \qquad H_1 : \sigma_f^2 \neq 0 \tag{2.31}$$

for testing completely paternal imprinting. The likelihood ratio test statistics for the above two tests asymptotically follow a 50 : 50 mixture of χ_0^2 and χ_1^2 distributions [28]. Rejection of complete imprinting indicates partial imprinting.

2.7. Multiple QTL Model

In reality, more than one QTL may contribute to the phenotypic variation located in one chromosome region or across the whole genome. The polygenic effect in model (2.4) absorbs the effects of multiple QTL located on other chromosomes. However, when there are multiple QTL located on the same linkage group as the tested QTL, if their effects are not properly adjusted, the estimation could be biased due to interference caused by these QTL outside of the testing interval [3, 30–32]. A multiple QTL model that can test the putative QTL effect while adjusting the effects of interference QTL deserves more attention.

Zeng [32] previously showed that IBD variables share the same property as the indicator variables in which the shared proportion of alleles IBD for a QTL conditional on the IBD of one flanking marker is independent of that of a QTL on the other side of that flanking marker. Thus, conditional on one flanking marker, the interference of QTL located on the other side of the marker can be eliminated. By conditional on the IBD of the flanking markers, the IBD sharing of a QTL is uncorrelated with that outside this interval. Xu and Atchley [25] showed that one marker is enough to block the interference caused by other QTL located on the same linkage group. The authors derived the next-to-flanking markers structure to block additional QTL effects from both sides of testing region in one chromosome. We derive a multiple QTL model adopting a similar idea as Xu and Atchley [25]. Assume there are total S QTL located on a linkage group. Considering parent-specific allelic effects, the multiple QTL model can be expressed in general as

$$y_{ik} = \mu_k + \sum_{s=1}^{S} a_{ikms} + \sum_{s=1}^{S} a_{ikfs} + G_{ik} + e_{ik}, \quad k = 1, \ldots, K;\ i = 1, \ldots, n_k. \tag{2.32}$$

In an interval-based linkage scan, only one putative QTL is considered at each testing position conditioning on the effects of all other QTL. Assuming there are total L and R QTL located on the left and right sides of the putative QTL on a linkage group, model (2.32) can be modified as

$$y_{ik} = \mu_k + \sum_{l=1}^{L} a_{ikl} + \left(a_{ikm} + a_{ikf}\right) + \sum_{r=1}^{R} a_{ikr} + G_{ik} + e_{ik}, \quad k = 1, \ldots, K;\ i = 1, \ldots, n_k, \tag{2.33}$$

where a_{ikl} and a_{ikr} are the lth and rth QTL random effects on the left and right sides of the putative QTL, respectively. When testing the putative QTL effect, we are only interested in blocking the total effects of QTL outside of the tested interval. Therefore, in the modified model, the effects of QTL outside of the tested interval are not partitioned. This, however, does not affect the inference of the tested QTL.

As shown by Zeng [32] and Jansen [31, 33], one marker is enough to block the correlation between a locus on its left and a locus on its right. Therefore, only two additional markers flanking the current interval are needed to block interference caused by outside QTL [25]. Let $\mathcal{M}_l$ and $\mathcal{M}_r$ denote two flanking markers for the tested interval, and $\mathcal{L}$ and $\mathcal{R}$ denote the two markers next to $\mathcal{M}_l$ and $\mathcal{M}_{l+1}$ with the marker order $\mathcal{L}$ - $\mathcal{M}_l$ - $\mathcal{M}_{l+1}$ - $\mathcal{R}$. With the modified model given in (2.33), the covariance of phenotypes between individuals i and j in the kth backcross family can be expressed as

$$\begin{aligned}
\operatorname{Cov}\left(y_{ik}, y_{jk}\right) &= \sum_{l=1}^{L} \operatorname{Cov}\left(a_{ikl}, a_{jkl}\right) + \operatorname{Cov}\left(a_{ikm}, a_{jkm}\right) + \operatorname{Cov}\left(a_{ikf}, a_{jkf}\right) + \operatorname{Cov}\left(a_{ikm}, a_{jkf}\right) \\
&\quad + \sum_{r=1}^{R} \operatorname{Cov}\left(a_{ikr}, a_{jkr}\right) + \phi_{ij}\sigma_g^2 + I_{ij}\sigma_e^2 \\
&= \sum_{l=1}^{L} \pi_{l|k}\sigma_l^2 + \pi_{i_m j_m}\sigma_m^2 + \pi_{i_m/j_f}\sigma_{mf}^2 + \pi_{i_f j_f}\sigma_f^2 + \sum_{r=1}^{R} \pi_{r|k}\sigma_r^2 + \phi_{ij}\sigma_g^2 + I_{ij}\sigma_e^2,
\end{aligned} \tag{2.34}$$

where $\pi_{l|k}$ and $\pi_{r|k}$ are the IBD values for QTL located on the left and right sides of the putative QTL in the kth backcross family, and can be calculated following (2.6) and (2.7) if their genotype information is known. Unfortunately, the number and exact locations of QTL outside the testing interval are unknown. Hence, $\pi_{l|k}$ and $\pi_{r|k}$ are not observable. Xu and Atchley [25] showed that when $\pi_{l|k}$ and $\pi_{r|k}$ are unknown, they can be estimated by some composite terms $K(\theta_{l\mathcal{L}}, \pi_{\mathcal{L}|k})$ and $K(\theta_{l\mathcal{R}}, \pi_{\mathcal{R}|k})$, where $K(\theta_{l\mathcal{L}}, \pi_{\mathcal{L}|k})$ is a function of the recombination fraction between the lth QTL and the left marker $\mathcal{L}$ as well as a function of $\pi_{\mathcal{L}|k}$, the IBD value for a pair of individuals at the left marker $\mathcal{L}$. $K(\theta_{l\mathcal{R}}, \pi_{\mathcal{R}|k})$ can be similarly defined. Following Xu and Atchley [25], $K(\theta_{l\mathcal{L}}, \pi_{\mathcal{L}|k})$ can be expressed as a function of $\pi_{\mathcal{L}|k}$ multiplied by a function of recombination frequency between the lth QTL and the marker $\mathcal{L}$, $f(\theta_{l\mathcal{L}})$, that is, $K(\theta_{l\mathcal{L}}, \pi_{\mathcal{L}|k}) = \pi_{\mathcal{L}|k} f(\theta_{l\mathcal{L}})$. Similarly, $K(\theta_{l\mathcal{R}}, \pi_{\mathcal{R}|k}) = \pi_{\mathcal{R}|k} f(\theta_{r\mathcal{R}})$. When doing an interval scan, the covariance function given in (2.34) between individuals i and j can be reexpressed as

$$\begin{aligned}
&\operatorname{Cov}\left(y_{ik}, y_{jk} \mid \pi_{\mathcal{L}|k}, \widehat{\pi}_{i_m j_m}, \widehat{\pi}_{i_m/j_f}, \widehat{\pi}_{i_f j_f}, \pi_{\mathcal{R}|k}\right)\\
&\quad = \sum_{l=1}^{L} K\left(\theta_{l\mathcal{L}}, \pi_{\mathcal{L}|k}\right)\sigma_l^2 + \widehat{\pi}_{i_m j_m}\sigma_m^2 + \widehat{\pi}_{i_m/j_f|k}\sigma_{mf}^2 + \widehat{\pi}_{i_f j_f}\sigma_f^2\\
&\qquad + \sum_{r=1}^{R} K\left(\theta_{l\mathcal{R}}, \pi_{\mathcal{R}|k}\right)\sigma_r^2 + \phi_{ij}\sigma_g^2 + I_{ij}\sigma_e^2\\
&\quad = \pi_{\mathcal{L}|k}\sum_{l=1}^{L} f\left(\theta_{l\mathcal{L}}\right)\sigma_l^2 + \widehat{\pi}_{i_m j_m}\sigma_m^2 + \widehat{\pi}_{i_m/j_f}\sigma_{mf}^2 + \widehat{\pi}_{i_f j_f}\sigma_f^2\\
&\qquad + \pi_{\mathcal{R}|k}\sum_{r=1}^{R} f\left(\theta_{r\mathcal{R}}\right)\sigma_r^2 + \phi_{ij}\sigma_g^2 + I_{ij}\sigma_e^2\\
&\quad = \pi_{\mathcal{L}|k}\sigma_L^2 + \widehat{\pi}_{i_m j_m}\sigma_m^2 + \widehat{\pi}_{i_m/j_f|k}\sigma_{mf}^2 + \widehat{\pi}_{i_f i_f}\sigma_f^2 + \pi_{\mathcal{R}|k}\sigma_R^2 + \phi_{ij}\sigma_g^2 + I_{ij}\sigma_e^2.
\end{aligned} \tag{2.35}$$

Instead of estimating individual variance components σ_l^2 and σ_r^2, now we estimate the composite terms $\sum_{l=1}^{L} f(\theta_{l\mathcal{L}})\sigma_l^2 = \sigma_L^2$ and $\sum_{r=1}^{R} f(\theta_{r\mathcal{R}})\sigma_r^2 = \sigma_R^2$. By conditioning the IBD sharing information for the left and right markers $\mathcal{L}$ and $\mathcal{R}$, the effects of those interference QTL are blocked. σ_L^2 and σ_R^2 absorb the random effects of all QTL that are outside of the testing interval but are on the same linkage group as the putative QTL. Estimation of the variance components terms follows the same procedure as the single QTL analysis with slight modification to consider multiple variance components.

3. Results

3.1. Simulation Design

To investigate the performance of the proposed models and estimation methods, we conduct intensive computer simulations. We start with the single-QTL simulation followed by the multiple QTL analysis. Six evenly spaced markers ($\mathcal{M}_1 - \mathcal{M}_6$) are simulated. The total length for the simulated linkage group is 100 cm. We assume that all the backcross families share the same linkage map constructed using Haldane map function. For simplicity, we assume the sample size for all backcross families is the same (i.e., $n_k = n$). The position of the simulated QTL is assumed to be located 48 cm away from the first marker ($\mathcal{M}_1$). The effect of the putative QTL is simulated by assuming different imprinting mechanisms, that is, no imprinting, completely imprinting, and partial imprinting. Once QTL genotypes are simulated, phenotypes can be simulated by randomly drawing multivariate normal distribution with the covariance structure given in (2.12) with different parameter combinations.

To evaluate the effect of family and offspring size combination on testing power and parameter estimation, we simulate data assuming different sample size combinations. We fix the total sample size as 400 and vary the family and offspring size with different combinations, that is, 4×100, 8×50, 20×20, and 100×4. The first number for each combination indicates the family size. For example, in the combination 4 × 100, 4 families each containing 100 offsprings are simulated. For each sib-pair, the IBD value at a putative position at every 2 cm along the linkage group is calculated as described in Section 2.7. For each simulation scenario, 100 simulation replications are recorded, and the ML and the REML methods are used to estimate the unknown parameters.

3.2. Simulation Results

3.2.1. Single QTL Analysis

The single QTL model assumes one QTL is located at the third interval in the simulated linkage group, 48 cm away from the first marker. Results using both ML and REML estimation methods are summarized in Table 2. n_F denotes the number of families and n_k denotes the number of offsprings for each family. Without loss of generality, we assume equal offspring size for all families in each simulation scenario. The simulated parameter values are listed under each parameter. The root mean square errors (RMSEs) are recorded for each parameter estimate to assess the estimation precision. Overall, the fixed effects (three means) and most variance components can be better estimated with large number of families. For example, the RMSE of parameter μ_1 is reduced from 1.869 (2.45) to 0.321 (0.305) when the number of families increases from 4 to 100 with the ML (REML) estimation method. The only exception is the two variance components terms (σ_m^2 and σ_f^2) which are better estimated with the 20 × 20 combination design. Through the combination of different line crosses, the parameter inference space is expanded, and as a result, better estimations are achieved as expected. However, the QTL position is better estimated with the 8 × 50 and 20 × 20 designs than the other two among the four simulation scenarios. The 100 × 4 design gives the worst QTL position estimation with the largest RMSEs for both estimation methods. Therefore, a balance of family and offspring size is needed. A moderate family size with moderate offspring size would be necessary in order to achieve reasonable parameter estimation for both QTL effects and position.

Table 2 also lists the results of power analysis under different scenarios with two different estimation methods. Power^1 denotes the empirical power calculated from the simulated null distribution corresponding to hypothesis test (2.27). We simulate the null distribution by simulating data assuming no QTL effect (i.e., $\sigma_m^2 = \sigma_f^2 = \sigma_{mf}^2 = 0$). The LR test statistics is calculated for each simulation run, and the 95% cutoff is reported as the threshold value. Power^2 refers to the theoretical power which is calculated assuming the mixture chi-square distribution, that is, $(1/8)\chi_0^2 + (3/8)\chi_1^2 + (3/8)\chi_2^2 + (1/8)\chi_3^2$ [28]. Results show that the threshold calculated from the theoretical distribution is smaller than the one calculated from the simulation. Thus, the testing power based on the theoretical cutoff is greater than the empirical power. The testing powers under different sampling designs are very comparable except for the 100 × 4 design in which the power is dramatically reduced compared to other designs. No remarkable difference in power for both estimation methods is observed. Figure 2 shows the log-likelihood ratio test statistic calculated under the four sampling designs across the simulated linkage group by using both ML and REML estimation

Table 2: The power, MLEs, and REMLs of the QTL position, and effect parameters estimated based on 100 simulation replicates for a QTL with no imprinting effect under different sampling designs. The square roots of the mean squared errors of the parameter estimates are given in parentheses.

	Estimation	Position	$\mu 1$	$\mu 2$	$\mu 3$	σ_m^2	σ_f^2	σ_{mf}^2	σ_g^2	σ_e^2			Type I
$n_F \times n_k$	method	48 cm	10	8	6	1.5	1.5	0.5	0.5	2	Power^1	Power^2	error
4 × 100	ML	48.78	9.85	8.08	5.80	0.91	0.73	0.78	0.17	2.12	0.91	0.99	0.42
		(17.87)	(1.869)	(1.246)	(1.768)	(1.296)	(1.122)	(1.242)	(0.572)	(0.229)			
	REML	47.2	9.83	8.13	6.23	1.51	1.42	1.52	1.06	2.02	0.92	0.96	0.17
		(11.95)	(2.45)	(1.39)	(2.42)	(1.65)	(1.33)	(2.88)	(2.35)	(0.57)			
8 × 50	ML	50.04	9.68	7.97	5.87	1.09	1.28	0.98	0.37	2.06	0.81	1.00	0.26
		(14.73)	(1.373)	(0.736)	(1.157)	(0.943)	(0.990)	(1.302)	(0.701)	(0.233)			
	REML	46.9	9.84	8.12	5.91	1.61	1.64	1.02	0.70	2.03	0.98	1.00	0.08
		(7.93)	(1.16)	(0.83)	(1.10)	(0.99)	(1.26)	(1.44)	(1.10)	(0.37)			
20 × 20	ML	48.96	9.99	7.95	6.05	1.34	1.25	0.60	0.32	2.01	0.98	0.98	0.12
		(10.70)	(0.706)	(0.495)	(0.633)	(0.768)	(0.771)	(0.810)	(0.478)	(0.182)			
	REML	49.74	10.04	7.94	5.89	1.55	1.48	0.69	0.58	1.98	0.95	0.98	0.06
		(13.96)	(0.597)	(0.458)	(0.675)	(0.964)	(0.786)	(0.838)	(0.608)	(0.257)			
100 × 4	ML	51.34	10.02	7.95	6.02	1.50	1.45	0.58	0.41	2.01	0.55	0.86	0.11
		(18.00)	(0.321)	(0.277)	(0.345)	(0.875)	(0.878)	(0.553)	(0.444)	(0.242)			
	REML	48.4	9.95	8.01	6.01	1.51	1.57	0.62	0.52	1.99	0.67	0.81	0.07
		(20.84)	(0.305)	(0.220)	(0.319)	(0.902)	(0.892)	(0.654)	(0.485)	(0.222)			

The locations of the QTL are described by the map distances (in cm) from the first marker of the linkage group (100 cm long). The true QTL is located at 48 cm; Powr^1 is calculated using the empirical distribution through simulation. Powr^2 is calculated using the theoretical distribution assuming mixture chi-square distribution. Type I error refers to the imprinting type I error.

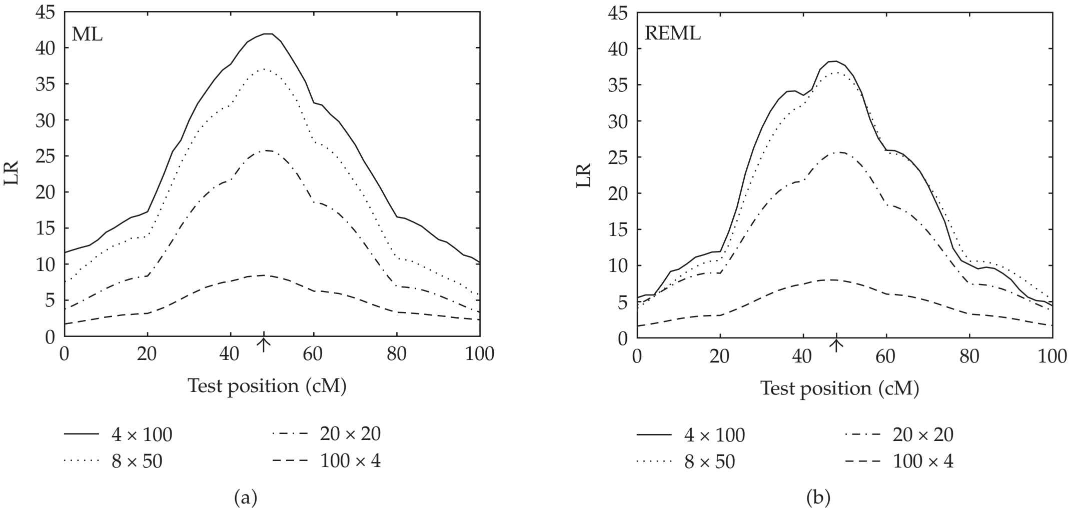

Figure 2: The LR profile plot. The left and right figures correspond to the LR profiles generated using the ML and REML methods, respectively. The arrow indicates the true QTL position.

methods. The plotted LR curve is from averaged LR values out of 100 replications. It is clear that large offspring size always gives large test statistics. As the family size increases from 4 to 100 and so decreased offspring size, we observe a huge LR value decrease. Clearly, the 100 × 4 design is less powerful than the others. The last column listed in Table 2 shows type I error for testing genomic imprinting, that is, $H_0 : \sigma_m^2 = \sigma_f^2$. The simulated data assume no imprinting ($\sigma_m^2 = \sigma_f^2 = 1.5$). The imprinting test is only conducted at the position where the overall QTL test shows significance. The imprinting test statistic LR_{imp} is compared with a chi-square distribution with 1 df. Overall, the REML estimation method results in smaller type I error rate than the ML method does. As the number of families increase, type I error decreases. The 4 ×100 design yields the largest type I error.

In comparison of the ML and REML methods, the REML method gives smaller estimation biases but larger RMSEs than the ML method does. This reflects the large variability of the REML estimation. In terms of computation speed, the ML method is faster than the REML method. For example, in a single-simulation run with the one-QTL model, the ML method takes about 9 minutes to scan the linkage group compared to 26 minutes with the REML method. The difference is more remarkable with the multiple QTL model (e.g., 10 minutes for ML versus 43 minutes for REML). Even though the QTL position estimation is better estimated by using the REML method when family size is small, as family size increases, the REML method performs worse than the ML method (Table 2). In checking the LR profile plot in Figure 2 and the power analysis in Table 2, we do not observe significant gain in power by using the REML method. The two methods do no dominate each other and are very comparable in power analysis. With large sample size and limited computing resources, one might want to try the ML method first. However, the REML method is suggested when testing imprinting since it has small type I error.

In a short summary of the results listed in Table 2, the 8 × 50 and 20 × 20 designs give better QTL position estimation and testing power. In terms of the type I error for imprinting test, the 20 × 20 and 100 × 4 designs provide reasonable type I error. Thus, a practical guidance is to choose the 20 × 20 design, and one should always avoid designs with extremely large or extremely small family size.

To evaluate the proposed model under different imprinting mechanisms, we simulated data assuming different degrees of imprinting. Since the results in Table 2 indicate that a 20 × 20 design provides relatively reasonable parameter estimation, good power, and small type I error rate for imprinting test, the evaluation of imprinting analysis is thus focused on this design. The results for 100 simulation replications are summarized in Table 3. Three imprinting models are assumed complete maternal imprinting ($\sigma_m^2 = 0$ and $\sigma_f^2 = 3$), complete paternal imprinting ($\sigma_m^2 = 3$ and $\sigma_f^2 = 0$), and partial maternal imprinting ($\sigma_m^2 = 1$ and $\sigma_f^2 = 2$). Both ML and REML estimators are reported. Overall, the two estimation methods produce very comparable results with less-biased estimations by the REML method as we expected. All the parameters can be properly estimated with reasonable precision. Large imprinting power is observed when the variance difference between the two parent-specific variance components is large. When the difference between the two parent-specific variance components is reduced, the power to detect imprinting is largely reduced. For example, when data are simulated assuming complete paternal imprinting, the power is 0.91(0.86) by using the ML(REML) estimation method. With partially imprinted data, the imprinting power reduces to 0.24(0.09) by using the ML(REML) method, even though it can be increased by increasing the offspring sample size (data not shown).

In reality, whether a QTL is imprinted or not is an unknown prior. When a QTL has Mendelian effect and is not imprinted, is there any power loss by analyzing with the proposed imprinting model? Or when a QTL is actually imprinted, is there any power loss by analyzing with regular variance components approach? To answer these two questions, we simulated data under different scenarios and analyzed with both Mendelian and imprinting models. The first and second columns in Table 4 refer to the simulation and analysis models, respectively. M refers to the Mendelian model without variance components partition and I refers to the imprinting model with allelic-specific partition of the variance components. For comparison purpose, heritabilities are recorded instead of original variance components estimates. The polygene and residual variances are fixed as 0.5 ($h_g^2 = 0.083$) and 2, respectively for all the simulation scenarios. We first simulated data with one additive genetic effect without partitioning variance into allelic-specific components. This is equivalent to simulate data assuming the Mendelian model. A single additive variance component of 3.5 is assumed which corresponds to a heritability of $h_a^2 = 0.583$. The second scenario is to simulate data with three allelic-specific variance components. Simulation models I_1 and I_2 correspond to a complete maternal imprinting model (i.e., $h_m^2 = 0$ and $h_f^2 = 0.5$) and a partial maternal imprinting model (i.e., $h_m^2 = 0.083$ and $h_f^2 = 0.417$), respectively. The variance component σ_{mf}^2 is assumed to be 0.5 ($h_{mf}^2 = 0.083$) for I_1 and I_2. In all the simulations, we use the 20×20 design to make the comparison. Similar results are expected under the other sampling designs. Since the true variance components values for the imprinting model are unknown when data are simulated assuming Mendelian effect and vice versa, only standard deviations for these parameter estimates are recorded (listed as italic font in the parentheses).

The simulation results are summarized in Table 4 analyzed with the ML method. When the simulated model is Mendelian, QTL position is better estimated with the Mendelian model than with the imprinting model. No remarkable difference in power is observed for both models. The estimated parent-specific variances due to maternal and paternal alleles are almost identical and no imprinting is detected. When data are simulated assuming imprinting (model I_1 and I_2), large power is observed when analyzed with the imprinting

Table 3: The power, MLEs, and 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. The square roots of the mean squared errors of the parameters are given in parentheses.

Estimation	Position	$\mu1$	$\mu2$	$\mu3$	σ_m^2	σ_f^2	σ_{mf}^2	σ_g^2	σ_e^2			
method	48 cm	10	8	6			0.5	0.5	2	Powr[1]	Powr[2]	ipower
					3	0						
ML	48.98	10.06	8.02	5.94	2.94	0.10	0.65	0.25	2.04	0.98	0.99	0.91
	(7.36)	(0.716)	(0.388)	(0.722)	(1.197)	(0.217)	(0.694)	(0.426)	(0.207)			
REML	48.54	9.96	7.95	5.90	2.95	0.09	0.64	0.57	1.97	0.97	0.99	0.86
	(9.62)	(0.705)	(0.428)	(0.747)	(1.501)	(0.180)	(0.598)	(0.609)	(0.226)			
					0	3						
ML	46.62	10.05	8.02	6.08	0.09	2.76	0.69	0.30	2.05	0.97	1.00	0.89
	(8.11)	(0.680)	(0.442)	(0.584)	(0.201)	(1.052)	(0.722)	(0.430)	(0.193)			
REML	48.7	10.04	7.98	5.91	0.09	2.95	0.61	0.57	1.98	0.96	0.98	0.88
	(7.40)	(0.560)	(0.506)	(0.615)	(0.195)	(1.279)	(0.663)	(0.544)	(0.225)			
					1	2						
ML	47.8	10.09	8.00	6.11	0.84	2.04	0.66	0.32	2.02	0.97	1.00	0.24
	(7.83)	(0.634)	(0.468)	(0.666)	(0.615)	(0.994)	(0.811)	(0.475)	(0.203)			
REML	49.66	10.02	7.96	5.90	1.08	1.73	0.68	0.61	2.00	0.97	0.99	0.09
	(11.06)	(0.606)	(0.548)	(0.668)	(0.679)	(0.958)	(0.707)	(0.616)	(0.216)			

Por[1] and Por[2] correspond to the overall QTL effect test (2.27) calculated using the empirical and theoretical cutoffs, respectively; ipower refers to the imprinting test power corresponding to test (2.29). See Table 2 for explanations of other parameters.

Table 4: The power, MLEs of the QTL position, and effect parameters estimated based on 100 simulation replicates for data simulated with Mendelian and imprinting models based on the 20 × 20 sampling design.

Simulation	Analysis	Position	$\mu1$	$\mu2$	$\mu3$		h_a^2		h_g^2	σ_e^2		
model	model	48 cm	10	8	6	h_m^2	h_f^2	h_{mf}^2	0.083	2	Power[1]	Power[2]
M	M	48.08	10.123	7.890	6.023		0.544		0.093	1.982	1.00	1.00
		(2.981)	(0.887)	(0.730)	(1.046)		(0.121)		(0.137)	(0.250)		
	I	48.260	9.922	8.081	6.074	0.263	0.272	0.239	0.049	1.983	1.00	1.00
		(3.015)	(1.010)	(0.720)	(0.949)	(0.097)	(0.110)	(0.124)	(0.080)	(0.249)		
I_1	M	48.940	10.079	7.913	6.023		0.323		0.111	2.022	0.86	0.96
		(11.660)	(0.588)	(0.511)	(0.712)		(0.287)		(0.134)	(0.205)		
	I	47.86	10.077	7.900	6.018	0.008	0.473	0.089	0.091	2.015	0.95	0.98
		(7.524)	(0.579)	(0.575)	(0.665)	(0.022)	(0.151)	(0.099)	(0.053)	(0.86)		
I_2	M	48.640	10.070	7.917	6.028		0.331		0.103	2.023	0.91	0.98
		(8.389)	(0.613)	(0.517)	(0.706)		(0.278)		(0.129)	(0.195)		
	I	49.200	10.059	7.926	6.008	0.079	0.396	0.089	0.089	2.023	0.93	0.98
		(9.818)	(0.617)	(0.518)	(0.673)	(0.075)	(0.156)	(0.097)	(0.050)	(0.184)		

M and I refer to Mendelian and imprinting models, respectively. Simulated parameters for model M: (h_a^2 = 0.583); I_1 : (h_m^2, h_f^2, h_{mf}^2) = (0, 0.5, 0.083); I_2 : (h_m^2, h_f^2, h_{mf}^2) = (0.083, 0.417, 0.083). Power[1] and Power[2] correspond to the power calculated using the empirical cutoff and the theoretical threshold, respectively. The numbers given in the parenthesis with normal and italic fonts correspond to the RMSEs and standard errors of the parameter estimates, respectively. See Table 2 for other explanations.

Table 5: The MLEs and REMLs of the QTL position and effect parameters estimated based on 100 simulation replicates for data simulated with two QTL under the 20 × 20 design. The square roots of the mean squared errors are given in parentheses.

	Q_1				Q_2			
Estimation	Position	σ_m^2	σ_f^2	σ_{mf}^2	Position	σ_m^2	σ_f^2	σ_{mf}^2
method	28 cm			0.25	68 cm			0.25
		0.75	0.75			0.75	0.75	
ML	31.06	0.960	0.819	0.381	64.560	0.806	0.891	0.414
	(12.825)	(0.760)	(0.693)	(0.533)	(12.126)	(0.552)	(0.679)	(0.646)
REML	32.32	0.949	0.924	0.440	64.60	0.943	1.002	0.461
	(13.951)	(0.739)	(0.686)	(0.678)	(12.169)	(0.713)	(0.711)	(0.601)
		1.5	0			0.75	0.75	
ML	29.52	1.568	0.210	0.601	63.86	1.210	0.532	0.533
	(12.153)	(0.872)	(0.382)	(0.836)	(11.913)	(0.932)	(0.572)	(0.847)
REML	31.22	1.668	0.282	0.433	64.16	1.355	0.665	0.529
	(12.557)	(0.881)	(0.457)	(0.598)	(13.030)	(1.128)	(0.554)	(0.650)

Q_1 and Q_2 refer to two QTL located at 28 cm and 68 cm.

model. For example, the power is 86% when analyze the I_1 imprinting data by the Mendelian model. The power is increased to 95% when data are analyzed by the imprinting model. When imprinting data are analyzed with the Mendelian model, the major QTL variance is underestimated, and the polygene variance is slightly overestimated. No remarkable differences are observed for the estimation of the three fixed mean effects and the residual variance under all simulation cases. In any case, the imprinting model performs better or no worse than the Mendelian model in terms of power. In checking type I error rate based on the theoretical threshold, we find the imprinting model has slightly higher type I error rate compared with the Mendelian model. In real data analysis, it is more important to control the false negatives than the false positives. Thus, it is safe to apply the imprinting model for data with any inheritance pattern in this regard.

3.2.2. Multiple QTL Analysis

To see the relative merit of multiple QTL analysis against single QTL analysis when multiple QTL are located on the same linkage group, two QTL are simulated with QTL 1 (denoted as Q_1) located at the second interval, 28 cm away from the first marker ($\mathcal{M}_1$), and QTL 2 (denoted as Q_2) located at the fourth interval, 68 cm away from the first marker. Two simulation scenarios are considered. The first scenario considers two nonimprinted QTL with equal genetic effects. The second scenario assumes Q_1 is imprinted and Q_2 is not imprinted. Simulated parameters for the two QTL are listed in Table 5. Data are simulated assuming the 20×20 design. Parameters are estimated by the ML and REML approaches with 100 replicates.

Figure 3 shows the LR profile plots for the single and multiple QTL analyses. The single QTL model indicates three major peaks. The highest peak for the single QTL analysis is located at the wrong QTL interval where no QTL is assumed. The so-called “ghost image” of QTL can be removed and the positions of the two QTL can be precisely mapped on the chromosome by the multiple QTL model. Two clear peaks indicating the correct QTL positions (arrow signs) are observed by the multiple QTL analysis. However, we observe a

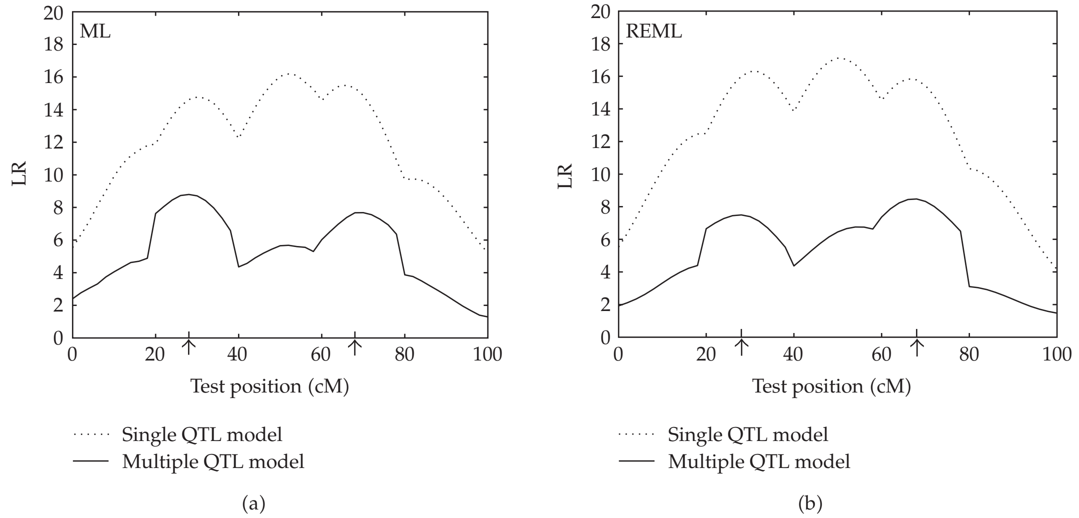

Figure 3: The LR profile plot for single QTL and multiple QTL analyses. The true QTL positions are simulated at 28 cm and 68 cm (see the arrow sign). The dotted curve and the solid curve represent the LR profiles by single QTL and multiple QTL analyses, respectively. The left and right figures correspond to the LR profiles generated using the ML and REML methods, respectively.

remarkable reduction in LR values by multiple QTL analysis compared to those by the single QTL analysis. Since the threshold for multiple QTL analysis is unknown, we cannot make the conclusion that multiple QTL analysis is less powerful than the single QTL analysis. It is possible that we may gain accuracy in QTL position estimation at the cost of power loss. Similar phenomenon and issues were also observed and discussed in the literature [3, 25].

The results of the multiple QTL analysis are summarized in Table 5. The fixed mean effects, the polygene, and residual variance components can be reasonably estimated with small RMSEs, similar results shown in Table 2 for the 20 × 20 design and hence are not reported here. Only the genetic factors for the two simulated QTLs are reported. It can be seen that both ML and REML methods provide reasonable parameter estimates and are very comparable. Under the first simulation scenario in which both QTL are not imprinted, the genetic effects are all slightly overestimated by both methods. This might be due to the interference of the two QTL in the same linkage group. The multiple QTL model may not completely block the effects of QTL outside of the tested interval. For the second simulation scenario, an interesting pattern is observed. When one QTL is imprinted (Q_1), the maternal and paternal variance components for the second one (Q_2) tend to be estimated with bias in the direction as the first imprinted QTL, that is, σ_m^2 tends to be overestimated and σ_f^2 tends to be underestimated. As we gain accuracy in QTL position estimation, we lose precision for the parameter estimation. These effects are expected as described in Zeng [3] and Xu and Atchley [25]. More investigations are needed in multiple QTL analysis in order to maintain a good balance of QTL position and parameter inference.

4. Discussion

Statistical methods assuming fixed effect models for iQTL mapping in controlled outbred and inbred lines have been proposed (e.g., [11, 14–16]). Considering the limitation of fixed-effect models, a random model that estimates the QTL variance by extending single line cross to multiple line crosses should be more powerful in QTL variance inference [18]. The IBD-based variance components method assuming random genetic effect for iQTL mapping has been developed in human linkage analysis [7]. However, no study has been proposed to map iQTL using variance components method with inbred or partially inbred line cross. In this article, we have first time presented an IBD-based variance components framework to search for the existence and distribution of iQTL throughout the entire genome in multiple experimental line crosses. The idea of the method is demonstrated through a backcross design. It can also be extended to multiple F_2 line crosses using the sex-specific recombination information as proposed by Cui et al. [15].

The key point of the proposed iQTL variance components analysis is to partition the additive genetic variance into parent-specific components. We have proposed a new parent-specific allelic sharing method which characterizes the relatedness of parent-specific alleles between pairs of individuals in a backcross pedigree. The calculation of parent-specific allelic sharing is based on the information of the coefficient of coancestry. More complicated calculation of the coefficient of coancestry can be found at Harris [21]. The quantification of the coefficient of the coancestry proposed by Harris [21] can also be utilized to calculate the parent-specific IBD sharing in an inbred human population, and thus for iQTL mapping in inbred human populations.

There have been extensive studies in the literature about various methods in the estimation of variance components in a mixed-effect model framework. The ML and REML are two commonly applied methods in variance components estimation with less-biased estimation by the REML method. Simulations show that the ML method yields high precision in parameter estimation but with relatively large bias than the REML method. Power analysis indicates that the ML method is a little more powerful than the REML method but with large type I error when testing imprinting. In terms of computing speed, the ML method is faster than the REML method. Thus, no single method dominates the other. In terms of overall QTL test, we suggest to use the ML method for the genome-wide linkage scan and use the REML method for the imprinting test.

The effect of sampling design is investigated by extensive simulations. Results indicate that one can always achieve large power with large offspring size when the total sample size is fixed. The LR value differences under different sampling designs are shown in Figure 2. However, the combination of small families each with large offsprings gives poor parameter estimation and large type I error for imprinting test (Table 2). As the number of families increase, we observe less-biased parameter estimates for both fixed and random effects, but with poor QTL position estimation and small power. This information implies that it is necessary to enlarge the number of families to improve precision of parameter estimation. Meanwhile, a balance of family and offspring size is needed to maintain good QTL detection power and position estimation. Our simulations indicate that for a fixed total sample size ($n = 400$), both 8×50 and 20×20 designs yield comparable results and both designs outperform the other two designs (Table 2). Moreover, the 20×20 design produces relatively small type I error in imprinting test. With the 20×20 design, results in Table 4 indicate that the imprinting model is better or as good as the regular Mendelian analysis without considering imprinting. In real data analysis, it should be safe to apply the proposed imprinting model for data with any imprinting pattern.

In this study, we have extended the single marker-based analysis to an interval-based mapping for genome-wide scan and testing of iQTL effects. Considering the interference of QTL located on the same linkage group, we have extended the single QTL model to multiple

QTL analysis following the derivation of Xu and Atchley [25]. Simulation results indicate the relative merits of the multiple QTL analysis with improved QTL position inference, but with possible power loss (Figure 3). This, however, has been a common issue in multiple QTL modeling (see [3, 25]). More investigations are needed in deriving efficient and robust multiple QTL mapping models to improve precision without suffering too much from power loss.

The theoretical distribution for the likelihood ratio test has been a challenging problem in QTL mapping. Dupuis and Siegmund [34] first proposed theoretical properties for LR test statistics in a genome-wide linkage scan for QTL in an interval mapping framework with a fixed-effect model. Currently, most linkage analysis using the variance components method assumes that the LR test statistic follows a mixture of chi-square distribution [35]. The mixture distribution is derived following Self and Liang [28]. With multiple testings and multiple nuisance parameters in a genome-wide scan, the assumptions to get the mixture chi-square distribution may not satisfy. Moreover, the multivariate normal assumption for the phenotypic data required to get the mixture distribution may not even valid. No theoretical work has been done to investigate this in an IBD-based variance components linkage mapping. Our simulations indicate that the theoretical threshold calculated from the mixture chi-square distribution is smaller than the simulated cutoff. Thus, the power calculated with the theoretical threshold is slightly inflated. A modified mixture chi-quare distribution may be more appropriate. More theoretical investigations are needed in this regard.

Acknowledgment

This work was supported by NSF Grant DMS-0707031 and by Michigan State University intramural research Grant 06-IRGP-789.

References

[1] M. Lynch and B. Walsh, *Genetics and Analysis of Quantitative Traits*, Sinauer, Sunderland, Mass, USA, 1998.

[2] E. S. Lander and D. Botstein, "Mapping mendelian factors underlying quantitative traits using RFLP linkage maps," *Genetics*, vol. 121, no. 1, pp. 185–199, 1989.

[3] Z.-B. Zeng, "Precision mapping of quantitative trait loci," *Genetics*, vol. 136, no. 4, pp. 1457–1468, 1994.

[4] C.-H. Kao, Z.-B. Zeng, and R. D. Teasdale, "Multiple interval mapping for quantitative trait loci," *Genetics*, vol. 152, no. 3, pp. 1203–1216, 1999.

[5] K. Pfeifer, "Mechanisms of genomic imprinting," *The American Journal of Human Genetics*, vol. 67, no. 4, pp. 777–787, 2000.

[6] A. Burt and R. L. Trivers, *Genes in Conflict*, Harvard University Press, Cambridge, Mass, USA, 2006.

[7] R. L. Hanson, S. Kobes, R. S. Lindsay, and W. C. Knowler, "Assessment of parent-of-origin effects in linkage analysis of quantitative traits," *The American Journal of Human Genetics*, vol. 68, no. 4, pp. 951–962, 2001.

[8] T. Liu, R. J. Todhunter, S. Wu, et al., "A random model for mapping imprinted quantitative trait loci in a structured pedigree: an implication for mapping canine hip dysplasia," *Genomics*, vol. 90, no. 2, pp. 276–284, 2007.

[9] M. Abney, M. S. McPeek, and C. Ober, "Estimation of variance components of quantitative traits in inbred populations," *The American Journal of Human Genetics*, vol. 66, no. 2, pp. 629–650, 2000.

[10] S. A. Knott, L. Marklund, C. S. Haley, et al., "Multiple marker mapping of quantitative trait loci in a cross between outbred wild boar and large white pigs," *Genetics*, vol. 149, no. 2, pp. 1069–1080, 1998.

[11] D.-J. de Koning, A. P. Rattink, B. Harlizius, J. A. M. van Arendonk, E. W. Brascamp, and M. A. M. Groenen, "Genome-wide scan for body composition in pigs reveals important role of imprinting,"

Proceedings of the National Academy of Sciences of the United States of America, vol. 97, no. 14, pp. 7947–7950, 2000.

[12] D.-J. de Koning, H. Bovenhuis, and J. A. M. van Arendonk, “On the detection of imprinted quantitative trait loci in experimental crosses of outbred species,” *Genetics*, vol. 161, no. 2, pp. 931–938, 2002.

[13] M. Lin, X.-Y. Lou, M. Chang, and R. Wu, “A general statistical framework for mapping quantitative trait loci in nonmodel systems: issue for characterizing linkage phases,” *Genetics*, vol. 165, no. 2, pp. 901–913, 2003.

[14] Y. Cui, “A statistical framework for genome-wide scanning and testing of imprinted quantitative trait loci,” *Journal of Theoretical Biology*, vol. 244, no. 1, pp. 115–126, 2007.

[15] Y. Cui, Q. Lu, J. M. Cheverud, R. C. Littell, and R. Wu, “Model for mapping imprinted quantitative trait loci in an inbred F_2 design,” *Genomics*, vol. 87, no. 4, pp. 543–551, 2006.

[16] Y. Cui, J. M. Cheverud, and R. Wu, “A statistical model for dissecting genomic imprinting through genetic mapping,” *Genetica*, vol. 130, no. 3, pp. 227–239, 2007.

[17] Y. Li, C. M. Coelho, T. Liu, et al., “A statistical model for estimating maternal-zygotic interactions and parent-of-origin effects of QTLs for seed development,” *PLoS ONE*, vol. 3, no. 9, p. e3131, 2008.

[18] C. Xie, D. D. G. Gessler, and S. Xu, “Combining different line crosses for mapping quantitative trait loci using the identical by descent-based variance component method,” *Genetics*, vol. 149, no. 2, pp. 1139–1146, 1998.

[19] C. I. Amos, “Robust variance-components approach for assessing genetic linkage in pedigrees,” *The American Journal of Human Genetics*, vol. 54, no. 3, pp. 535–543, 1994.

[20] G. Malécot, *Les mathématiques del'hérédité*, Masson et Cie, Paris, France, 1948.

[21] D. L. Harris, “Genotypic covariances between inbred relatives,” *Genetics*, vol. 50, no. 6, pp. 1319–1348, 1964.

[22] L. Almasy and J. Blangero, “Multipoint quantitative-trait linkage analysis in general pedigrees,” *The American Journal of Human Genetics*, vol. 62, no. 5, pp. 1198–1211, 1998.

[23] R. R. Corbeil and S. R. Searle, “A comparison of variance component estimators,” *Biometrics*, vol. 32, no. 4, pp. 779–791, 1976.

[24] J. A. Nelder and R. Mead, “A simplex method for function minimization,” *Computer Journal*, vol. 7, no. 4, pp. 308–313, 1965.

[25] S. Xu and W. R. Atchley, “A random model approach to interval mapping of quantitative trait loci,” *Genetics*, vol. 141, no. 3, pp. 1189–1197, 1995.

[26] H. D. Patterson and R. Thompson, “Recovery of inter-block information when block sizes are unequal,” *Biometrika*, vol. 58, pp. 545–554, 1971.

[27] R. Wu, C.-X. Ma, and G. Casella, *Statistical Genetics of Quantitative Traits: Linkage, maps, and QTL*, Statistics for Biology and Health, Springer, New York, NY, USA, 2007.

[28] S. G. Self and K.-Y. Liang, “Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions,” *Journal of the American Statistical Association*, vol. 82, no. 398, pp. 605–610, 1987.

[29] G. A. Churchill and R. W. Doerge, “Empirical threshold values for quantitative trait mapping,” *Genetics*, vol. 138, no. 3, pp. 963–971, 1994.

[30] O. Martínez and R. N. Curnow, “Estimating the locations and the sizes of the effects of quantitative trait loci using flanking markers,” *Theoretical and Applied Genetics*, vol. 85, no. 4, pp. 480–488, 1992.

[31] R. C. Jansen, “Controlling the type I and type II errors in mapping quantitative trait loci,” *Genetics*, vol. 138, no. 3, pp. 871–881, 1994.

[32] Z.-B. Zeng, “Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 90, no. 23, pp. 10972–10976, 1993.

[33] R. C. Jansen, “Interval mapping of multiple quantitative trait loci,” *Genetics*, vol. 135, no. 1, pp. 205–211, 1993.

[34] J. Dupuis and D. Siegmund, “Statistical methods for mapping quantitative trait loci from a dense set of markers,” *Genetics*, vol. 151, no. 1, pp. 373–386, 1999.

[35] D. B. Allison, M. C. Neale, R. Zannolli, N. J. Schork, C. I. Amos, and J. Blangero, “Testing the robustness of the likelihood-ratio test in a variance-component quantitative-trait loci-mapping procedure,” *The American Journal of Human Genetics*, vol. 65, no. 2, pp. 531–544, 1999.