DOI: 10.1534/genetics.106.056416

Mapping Quantitative Trait Loci Using the Experimental Designs of Recombinant Inbred Populations

Chen-Hung Kao[1]

Institute of Statistical Science, Academia Sinica, Taipei 11529, Taiwan, Republic of China

Manuscript received January 27, 2006
Accepted for publication July 6, 2006

ABSTRACT

In the data collection of the QTL experiments using recombinant inbred (RI) populations, when individuals are genotyped for markers in a population, the trait values (phenotypes) can be obtained from the genotyped individuals (from the same population) or from some progeny of the genotyped individuals (from the different populations). Let F_u be the genotyped population and F_v ($v \geq u$) be the phenotyped population. The experimental designs that both marker genotypes and phenotypes are recorded on the same populations can be denoted as (F_u/F_v, $u = v$) designs and that genotypes and phenotypes are obtained from the different populations can be denoted as (F_u/F_v, $v > u$) designs. Although most of the QTL mapping experiments have been conducted on the backcross and $F_2(F_2/F_2)$ designs, the other (F_u/F_v, $v \geq u$) designs are also very popular. The great benefits of using the other (F_u/F_v, $v \geq u$) designs in QTL mapping include reducing cost and environmental variance by phenotyping several progeny for the genotyped individuals and taking advantages of the changes in population structures of other RI populations. Current QTL mapping methods including those for the (F_u/F_v, $u = v$) designs, mostly for the backcross or F_2/F_2 design, and for the F_2/F_3 design based on a one-QTL model are inadequate for the investigation of the mapping properties in the (F_u/F_v, $u \leq v$) designs, and they can be problematic due to ignoring their differences in population structures. In this article, a statistical method considering the differences in population structures between different RI populations is proposed on the basis of a multiple-QTL model to map for QTL in different (F_u/F_v, $v \geq u$) designs. In addition, the QTL mapping properties of the proposed and approximate methods in different designs are discussed. Simulations were performed to evaluate the performance of the proposed and approximate methods. The proposed method is proven to be able to correct the problems of the approximate and current methods for improving the resolution of genetic architecture of quantitative traits and can serve as an effective tool to explore the QTL mapping study in the system of RI populations.

MOST biologically important traits show continuous variations and have poor heritability. Traditional study of quantitative genetics based on the phenotype evaluation to investigate quantitative trait loci (QTL) controlling these traits is difficult and limited. Recently, the advent of fine-scale molecular markers has provided researchers with an efficient tool for the detection of the underlying QTL. Most QTL detection experiments for producing marker genotypes and phenotypic traits in species have been conducted with populations derived from crosses between inbred lines, *e.g.*, backcross, advanced backcross, F_2, recombinant inbred (RI) populations, intermated recombinant inbred (IRI) populations, advanced intercross (AI) populations, advanced backcross populations, double haploid (DH) populations, and NC Design III, etc. (Comstock and Robinson 1952; Stuber *et al.* 1992; Beavis *et al.* 1994; Veldboom *et al.* 1994; Darvasi and Soller 1995; Austin and Lee 1996; Liu *et al.* 1996; Chapman *et al.* 2003; Winkler *et al.* 2003; Complex Trait Consortium 2004; Broman 2005). These different populations may show different properties in QTL mapping as they have different population structures, such as homozygosity, genotypic frequencies, and linkage disequilibrium (Weir 1996, Chap. 5). In principle, the use of the information about genotypes and phenotypes of individuals in these populations has become a key approach to detect the underlying QTL for the understanding of the genetic basis and the improvement of important traits in genetic study.

In the data collection of these QTL experiments, when individuals are genotyped for markers in a population, the trait values (phenotypes) can be recorded on the genotyped individuals (on the same population) or on some progeny of the genotyped individuals (on the progeny population). Fisch *et al.* (1996) illustrated the situations of data collection by F_u/F_v, where F_u is the genotyped population and F_v ($v \geq u$) is the phenotyped population, in the system of RI populations

[1] *Author e-mail:* chkao@stat.sinica.edu.tw

(see RECOMBINANT INBRED POPULATIONS for the population structure of RI populations). For example, F_2/F_2 denotes the typical F_2 design, where genotypes and phenotypes are obtained from the same individuals in the F_2 population, and F_2/F_4 denotes the design to genotype F_2 individuals and phenotype their progeny in the F_4 population. Although the (F_u/F_v, $u = v$) designs are typical (DOERGE *et al.* 1997; LYNCH and WALSH 1998, Chap. 15), the (F_u/F_v, $u < v$) designs are also very popular and important for QTL detection in the genetic analysis of complex traits. For example, the F_3/F_3, F_2/F_3, F_2/F_4, F_4/F_4, F_5/F_5, and F_6/F_7 designs have been used to detect QTL in maize by STUBER *et al.* (1992), BEAVIS *et al.* (1994), VELDBOOM *et al.* (1994), AUSTIN and LEE (1996), MIHALJEVIC *et al.* (2004, 2005), ZHANG and XU (2004), and SALA *et al.* (2006) and the F_4/F_6 design was used to study QTL in soybean by CHAPMAN *et al.* (2003). There are some benefits of using the (F_u/F_v, $u < v$) with multiple phenotyping individuals and (F_u/F_u, $u < 2$) designs. For example, the cost can be economical for not genotyping the progeny for markers, the environmental variance can be reduced by phenotyping multiple progeny for trait measurement, and homozygotes can be accumulated so that QTL mapping may be improved (COWEN 1988; LANDER and BOTSTEIN 1989; KNAPP and BRIDGES 1990; EDWARDS *et al.* 1992; AUSTIN and LEE 1996).

Traditional QTL mapping methods developed to date mostly assume that both marker genotypes and phenotypic traits are obtained from the same population [the (F_u/F_v, $u = v$) designs], and they especially focus on the F_2/F_2 and backcross designs (LANDER and BOTSTEIN 1989; JENSEN 1993; ZENG 1994; SATAGOPAN *et al.* 1996; KAO *et al.* 1999; NAKAMICHI *et al.* 2001; SEN and CHURCHILL 2001; KAO and ZENG 2002; YI *et al.* 2003; CARLBORG and HALEY 2004; ZOU *et al.* 2004). Some researchers have applied these traditional (approximate) methods to QTL mapping study by regarding the traits (trait means) of progeny as the traits of genotyped individuals, *i.e.*, by treating (F_u/F_v, $u < v$) designs as (F_u/F_v, $u = v$) designs, in the analysis (STUBER *et al.* 1992; BEAVIS *et al.* 1994; VELDBOOM *et al.* 1994; AUSTIN and LEE 1996; CHAPMAN *et al.* 2003; ZHANG and XU 2004). Such application implicitly ignores the fact that the traits are controlled by the progeny (F_v) genomes, not by their ancestral (F_u) genomes, and that the segregation of heterozygotes will vary their population structures. Consequently, the power of QTL detection may be affected and the estimates of QTL effects can be biased by the approximate methods as shown in ZHANG and XU (2004) and in this article. Statistical methods are generally lacking or inadequate for the (F_u/F_v, $u < v$) designs. FISCH *et al.* (1996) suggested to propose an adequate model for the (F_u/F_v, $u \leq v$) designs, and ZHANG and XU (2004) considered the nature of segregation to propose a one-QTL model for the F_2/F_3 design in QTL mapping. As shown in this article, the one-QTL model by ZHANG and XU (2004) and the approximate method may have confounding problems in the estimation of QTL parameters and lose power of QTL detection. Ideally, we would like to extend the one-QTL model to a multiple-QTL model and the F_2/F_3 design to the more general (F_u/F_v, $u \leq v$) designs for more practical and broad use in a way that multiple QTL and their possible epistasis can be considered in the model to correct the problems and the benefit of other RI populations as mentioned can be utilized to further improve and study QTL mapping. In this article, a statistical method considering the differences in population structures between different RI populations is developed on the basis of a more complete multiple-QTL model for the more general (F_u/F_v, $u \leq v$) designs. In addition, the QTL mapping properties of the proposed and approximate methods in the different (F_u/F_v, $u \leq v$) designs are also derived and discussed. A simulation study was performed for evaluating the relative efficiencies of different (F_u/F_v, $u \leq v$) designs and comparing the performance of the proposed and current methods in these designs. The proposed method is capable of improving the resolution of the genetic architecture of quantitative traits and can serve as a tool to study QTL mapping in the (F_u/F_v, $u \leq v$) designs.

RECOMBINANT INBRED POPULATIONS

Assume that two parental inbred lines, P_1 and P_2, differ substantially in the quantitative trait of interest and are fixed for alternative alleles at QTL and markers. A cross between the parental lines produces an F_1 population with all the same heterozygous individuals. If the F_1 individuals are selfed or intermated, it produces an F_2 population. In the F_2 population, the genotypic frequencies of P_1 homozygote, heterozygote, and P_2 homozygote are $\frac{1}{4}$, $\frac{1}{2}$, and $\frac{1}{4}$, respectively (the heterozygosity is 0.5), if one locus is considered. The frequency of recombinants (r) between any two loci in the F_2 population is equivalent to the recombination fraction (c). If the F_2 individuals are further selfed for $t - 2$ generations, it produces a so-called RI F_t population. For $t \rightarrow \infty$, the derived population is called recombinant inbred lines (RILs). In an F_t population, the frequencies of P_1 homozygote, heterozygote, and P_2 homozygote in a locus are expected to be $(1/2) - (1/2^t)$, $1/2^{t-1}$, and $(1/2) - (1/2^t)$, respectively (the heterozygosity is $1/2^{t-1}$), and the frequency of recombinants between two loci, denoted as r_t, is increasing as t is increasing and can be obtained according to HALDANE and WADDINGTON (1931). Haldane and Waddington showed that $r_\infty = c/(1 + 2c)$.

Genetic model: In a RI population, any individual can have three possible QTL genotypes, *QQ*, *Qq*, and *qq*, if only one QTL, say *Q*, is considered. Let the genotypic value, G_i, of an individual i have the following relation with the genetic parameters as

$$G_i = \begin{cases} \mu + a - d/2 & \text{for } QQ \\ \mu + d/2 & \text{for } Qq \\ \mu - a - d/2 & \text{for } qq, \end{cases} \tag{1}$$

where μ is the intercept, and a and d are the additive and dominance effects according to Cockerham's model (Kao and Zeng 2002). If multiple, say m, QTL are considered, the extension of the one-QTL genetic model in Equation 1 to a multiple-QTL model with epistasis is straightforward (Kao and Zeng 2002). If an individual i produces k progeny, the mean genotypic value of the k progeny, $\overline{G}_i$, is

$$\overline{G}_i = \mu + \frac{K_2 - K_0}{k} a + \frac{K_1 - K_2 - K_0}{2k} d, \tag{2}$$

where K_2, K_1, and K_0 denote the numbers of progeny with QQ, Qq, and qq genotypes among the k progeny, respectively. If the genotype of the individual i is QQ (qq), all the k progeny have the same QQ (qq) genotype, *i.e.*, $K_2 = k$ ($K_0 = k$), and the mean genotypic value is $\mu + a - d/2$ ($\mu - a - d/2$). If the genotype of the individual i is Qq, the possible genotype of the progeny can be QQ, Qq, or qq, and the mean genotypic value depends on K_2, K_1, and K_0. The possible allocations of (K_2, K_1, K_0) have $(k + 1)(k + 2)/2$ combinations and follow a trinomial distribution with k trials and cell probabilities $(1/2) - (1/2^{t-1})$, $(1/2^{t-2})$, and $(1/2) - (1/2^{t-1})$. The number of possible mean genotypic values corresponds to the number of possible allocations of (K_2, K_1, K_0). Let $\overline{g}_1, \overline{g}_2, \ldots, \overline{g}_{(k+1)(k+2)/2}$ denote the $(k + 1)(k + 2)/2$ genotypic means. For simplicity, the mean genotypic value in Equation 2 is expressed as

$$\overline{G}_i = \mu + a x_i + d z_i, \tag{3}$$

where

$$x_i = \frac{K_2 - K_0}{k} \quad \text{and} \quad z_i = \frac{K_1 - K_2 - K_0}{2k}$$

are to characterize the status of the additive and dominance effects in the genotypic means. Under the expression of Equation 3, the extension of the model for mean genotypic value from one QTL to multiple QTL is straightforward. For m QTL without epistasis, the genetic model can be written as

$$\overline{G}_i = \mu + \sum_{j=1}^{m} a_j x_{ij} + \sum_{j=1}^{m} d_j z_{ij}, \tag{4}$$

where x_{ij}'s and z_{ij}'s are the coded variables for Q_j's, $j = 1, 2, \ldots, m$, and are defined similarly as x_i and z_j in Equation 3. The extension of this model to consider epistasis is straightforward by introducing the cross-product terms as the terms of epistasis.

Variance components: When m QTL with complete marginal and epistatic effects are considered together, the genetic variances of a quantitative trait can be decomposed into $2m^2$ variances and $2m^4 - m^2$ covariances in a RI population. Taking $m = 2$ as an example, there are 8 genetic variances and 28 genetic covariances. If the two QTL are unlinked, the genetic variance in an F_t population can be found as

$$\begin{aligned} V_G^{[t]} = {} & \left[1 - \frac{1}{2^{(t-1)}}\right] a_1^2 + \frac{1}{2^{(t-1)}}\left[1 - \frac{1}{2^{(t-1)}}\right] d_1^2 + \left[1 - \frac{1}{2^{(t-1)}}\right] a_2^2 \\ & + \frac{1}{2^{(t-1)}}\left[1 - \frac{1}{2^{(t-1)}}\right] d_2^2 \\ & + \left[1 - \frac{1}{2^{(t-1)}}\right]^2 i_{aa}^2 + \frac{1}{4}\left[1 - \frac{1}{2^{(t-1)}}\right] i_{ad}^2 + \frac{1}{4}\left[1 - \frac{1}{2^{(t-1)}}\right] i_{da}^2 \\ & + \frac{1}{16}\left\{1 - \left[1 - \frac{1}{2^{(t-2)}}\right]^4\right\} i_{dd}^2 - \left[1 - \frac{1}{2^{(t-1)}}\right]\left[1 - \frac{1}{2^{(t-2)}}\right] a_1 i_{ad} \\ & - \left[1 - \frac{1}{2^{(t-1)}}\right]\left[1 - \frac{1}{2^{(t-2)}}\right] a_2 i_{da} \\ & - \left\{\frac{1}{4} - \frac{1}{2^t} - 2\left[\frac{1}{2} - \frac{1}{2^{(t-1)}}\right]^3\right\} d_1 i_{dd} \\ & - \left\{\frac{1}{4} - \frac{1}{2^t} - 2\left[\frac{1}{2} - \frac{1}{2^{(t-1)}}\right]^3\right\} d_2 i_{dd}, \end{aligned} \tag{5}$$

where i_{aa}, i_{ad}, i_{da}, and i_{dd} are the additive-by-additive, additive-by-dominance, dominance-by-additive, and dominance-by-dominance epistatic effects, respectively, under the setting of the digenic model of Equation 1. Some of the covariances are zeros. For $t = 2$, there is no genetic covariance and the genetic variance reduces to eight independent components $a_1^2/2 + d_1^2/4 + a_2^2/2 + d_2^2/4 + i_{aa}^2/4 + i_{ad}^2/8 + i_{da}^2/8 + i_{dd}^2/16$. As t is increasing, the additive variances are increasing due to the accumulation of homozygotes, and the dominance variances are decreasing for the loss of heterozygotes. For example, the additive and dominance variance components are $\frac{3}{4}a_1^2$ $(\frac{3}{4}a_2^2)$ and $\frac{3}{16}d_1^2$ $(\frac{3}{16}d_2^2)$ for the F_3 population, and these components are $\frac{7}{8}a_1^2$ $(\frac{7}{8}a_2^2)$ and $\frac{7}{64}d_1^2$ $(\frac{7}{64}d_2^2)$ for the F_4 population. These two components approach a_1^2 and zero for $t \rightarrow \infty$. This shows that the RI F_t, $t > 2$, populations can benefit the estimation of additive effects by cumulating the homozygotes, but may hurt the estimation of dominance effects due to the loss of heterozygotes. Also, the epistatic variances involving additive effects (i_{aa}, i_{ad}, and i_{da}) are increasing, and the dominance-by-dominance variance is decreasing. For example, the epistatic variances involving the additive effects are $\frac{9}{16}i_{aa}^2$ $(\frac{49}{64}i_{aa}^2)$, $\frac{3}{16}i_{ad}^2$ $(\frac{7}{32}i_{ad}^2)$, and $\frac{3}{16}i_{da}^2$ $(\frac{7}{32}i_{da}^2)$, and the dominance-by-dominance variance is $\frac{15}{256}i_{dd}^2$ $(\frac{175}{4096}i_{dd}^2)$ in the F_3 (F_4) population. The four variance components approach i_{aa}^2, $i_{ad}^2/4$, $i_{da}^2/4$, and zero, respectively, as $t \rightarrow \infty$. Also, the covariances between genetic effects become present in the F_t, $t > 2$, populations, and they will cause confounding problems in estimation for the one-QTL approach or if epistasis is present and ignored in QTL mapping.

THE STATISTICAL METHODS

Data structure: Consider a sample of size n from a $(\mathrm{F}_u/\mathrm{F}_v,\ u < v)$ design or a $(\mathrm{F}_u/\mathrm{F}_v,\ u = v)$ design. The n individuals from the F_u population are genotyped for markers $(X_i,\ i = 1, 2, \ldots, n)$. If the sample is from the $(\mathrm{F}_u/\mathrm{F}_v,\ u = v)$ design, the n genotyped individuals are phenotyped to obtain the n trait values (y_i's, $i = 1, 2, \ldots, n$). If the sample is from the $(\mathrm{F}_u/\mathrm{F}_v,\ u < v)$ design, each of the n genotyped individuals produces k progeny in the F_v generation for phenotyping, and their traits (y_{ij}'s, $j = 1, 2, \ldots, k$) or trait means ($\bar{y}_i$'s) are recorded. For QTL mapping using the data from $(\mathrm{F}_u/\mathrm{F}_v,\ u < v)$ designs, both the traditional (approximate) and the proposed QTL mapping methods have been used and are discussed here. When applying the traditional methods to $(\mathrm{F}_u/\mathrm{F}_v,\ u < v)$ designs, one assumes that the mean trait is controlled by the QTL in the F_u individuals, referred to as $Q^{[u]}$'s hereafter, rather than by the QTL in the F_v progeny, referred to as $Q^{[v]}$'s hereafter ($Q^{[u]}$ and $Q^{[v]}$ have the same dimension). As a result, the problems, such as bias in estimation and loss in power of QTL detection, will occur in QTL mapping for the traditional methods. The proposed method intends to connect the trait of the F_v progeny with $Q^{[v]}$'s using the marker information in the F_u individuals; hence it can correct the problems to improve QTL mapping in the $(\mathrm{F}_u/\mathrm{F}_v,\ u < v)$ designs as shown below.

The proposed method: Without loss of generality in inferring QTL mapping in the $(\mathrm{F}_u/\mathrm{F}_v,\ u \le v)$ designs, consider that the sample is obtained from a $(\mathrm{F}_u/\mathrm{F}_v,\ u < v)$ design and the trait means ($\bar{y}_i$'s) measured on the F_v progeny are used in the analysis. The proposed method attempts to relate the mean traits with the mean genotypic values at $Q^{[v]}$ using the marker information of the F_u individuals so that the genetic structure of the F_v population can be taken into account in modeling. If a quantitative trait is controlled by m nonepistatic QTL, $\bar{y}_i$ can be related to the m QTL by the model

$$\bar{y}_i = \mu + \sum_{j=1}^{m} a_j x_{ij}^* + \sum_{j=1}^{m} d_j z_{ij}^* + \bar{\epsilon}_i, \qquad (6)$$

where x_{ij}^*'s and z_{ij}^*'s, $j = 1, 2, \ldots, m$, are the coded variables associated with the additive and dominance effects at $Q_j^{[v]}$'s, $j = 1, 2, \ldots, m$, in the genotypic means, and they have the same definitions as x_i^* and z_i^* in Equation 4. The residual error ϵ_i is assumed to follow a normal distribution with mean zero and variance σ^2. As multiple (m) intervals are used to infer the multiple QTL, this model is a multiple-interval mapping-based (MIM-based) method (Kao *et al.* 1999) for the $(\mathrm{F}_u/\mathrm{F}_v,\ u < v)$ designs. A single-QTL model for the $\mathrm{F}_2/\mathrm{F}_3$ design was first proposed by Zhang and Xu (2004).

As QTL could be located in the marker intervals, the genotypic means (x_i^*'s and z_i^*'s) for the m QTL are unobservable and need to be inferred from the flanking marker genotype of the F_u individual. For k progeny, there are $[(k+1)(k+2)/2]^m$ genotypic means (possible values for x_i^*'s and z_i^*'s) for m QTL. Given a sample with size n, the likelihood function of the model in Equation 6 for $\theta = (a_1, d_1, a_2, d_2, \ldots, a_m, d_m, \sigma^2)$ is

$$L(\theta \mid \bar{\mathbf{Y}}, \mathbf{X}) = \prod_{i=1}^{n} \left[\sum_{j=1}^{[(k+1)(k+2)/2]^m} p_{ij} N\left(\bar{g}_j, \frac{\sigma^2}{k}\right) \right], \qquad (7)$$

where $\bar{g}_j$'s are the $[(k+1)(k+2)/2]^m$ genotypic means, and the mixing proportions, p_{ij}'s, are the conditional probabilities of the corresponding genotypic means given the marker genotype. The density of each individual is a mixture of $[(k+1)(k+2)/2]^m$ possible normals with different means, $\bar{g}_j$'s, and mixing proportions, p_{ij}'s. Note that the mixing proportions in the likelihoods can be obtained by Equation 9 and need not to be estimated at the tested positions. The EM algorithm (Dempster *et al.* 1977) is used for the estimation of the parameters in Equations 7 by treating the trait means and markers, $\bar{y}_i$'s and X_i's, as *observed data* and the coded variables of mean genotypic values, x_{ij}^*'s and z_{ij}^*'s, as *missing data.*

The EM algorithm and maximum-likelihood estimate: The coded variables, x_{ij}^* and z_{ij}^*, associated with the additive and dominance effects in the mean genotypic values of $Q_j^{[v]}$ are determined by K_{j2}, K_{j1}, and K_{j0}. Therefore, inferring the distribution of x_{ij}^* and z_{ij}^* is equivalent to inferring the distribution of K_{j2}, K_{j1}, and K_{j0}. To infer the distribution of K_{j2}, K_{j1}, and K_{j0} using the marker information from the F_u individuals, one may first infer the distribution of the $Q_j^{[u]}$ genotype given the marker information and then infer the distribution of K_{j2}, K_{j1}, and K_{j0} given the QTL genotype of $Q_j^{[u]}$. That is,

$$\begin{aligned}&\mathrm{Prob}(K_{j2}, K_{j1}, K_{j0} \mid X_i)\\ &\quad = \mathrm{Prob}(Q_j^{[u]} \mid X_i) \times \mathrm{Prob}(K_{j2}, K_{j1}, K_{j0} \mid Q_j^{[u]}, X_i). \qquad (8)\end{aligned}$$

For the flanking marker interval, I_j, in the RI populations, there are nine possible flanking marker genotypes. Given each of the nine marker genotypes, the conditional probabilities of the QTL genotypes Q_jQ_j, Q_jq_j, and q_jq_j for the within $Q_j^{[u]}$ are different in different RI populations, and they depend on their population structure. If an F_2 population ($u = 2$) is genotyped for markers, these conditional probabilities of $Q_j^{[2]}$ genotypes given the nine flanking marker genotypes have been provided by several researchers (see, *e.g.*, Table 2 of Kao and Zeng 1997). If an F_u, $u > 2$, population is genotyped, the conditional probabilities are similar to those for the F_2 population with r_t substituted for r_2 (see Haldane and Waddington 1931, for the derivation of r_t). Due to segregation, $Q_j^{[v]}$ and $Q_j^{[u]}$ may have the same or different genotypes. If $Q_j^{[u]}$ is Q_jQ_j (q_jq_j), $Q_j^{[v]}$ of each progeny is sure to be Q_jQ_j (q_jq_j). That is, $\mathrm{Prob}(K_{j2} = k, K_{j1} = 0, K_{j0} = 0 \mid Q_j^{[u]} = Q_jQ_j) = 1$ $[\mathrm{Prob}(K_{j2} = 0, K_{j1} = 0, K_{j0} = k \mid Q_j^{[u]} = q_jq_j) = 1]$. If $Q_j^{[u]}$ is Q_jq_j, $Q^{[v]}$ among the

k progeny can be Q_jQ_j, Q_jq_j, or q_jq_j, and it will follow a multinomial distribution (see *Genetic model*); *i.e.*,

$$\begin{aligned}&\mathrm{Prob}(K_{j2}=k_2, K_{j1}=k_1, K_{j0}=k_0 \mid Q_j^{[u]}=Q_jq_j, X_i)\\ &\quad = p(k_2,k_1,k_0)\\ &\quad = \frac{k!}{k_2!k_1!k_0!}\left(\frac{1}{2}-\frac{1}{2^{v-u+1}}\right)^{k_2}\left(\frac{1}{2^{v-u}}\right)^{k_1}\left(\frac{1}{2}-\frac{1}{2^{v-u+1}}\right)^{k_0}.\end{aligned}$$

Taking all three genotypes of $Q_j^{[u]}$ into consideration, it is straightforward to obtain

$$\begin{aligned}&\mathrm{Prob}(K_{j2}=k_2, K_{j1}=k_1, K_{j0}=k_0 \mid X_i)\\ &\quad = \sum_{Q_j^{[u]}=Q_jQ_j,Q_jq_j,q_jq_j} \mathrm{Prob}(K_{j2}=k_2, K_{j1}=k_1, K_{j0}=k_0 \mid Q_j^{[u]}, X_i)\\ &\qquad \times \mathrm{Prob}(Q_j^{[u]} \mid X_i).\end{aligned}$$

The possible number of allocations for each set of K_{j2}, K_{j1}, and K_{j0} is $(k+1)(k+2)/2$. If all the m QTL are considered at a time, there are 9^m possible flanking marker genotypes, and, for each marker genotype, there are totally $[(k+1)(k+2)/2]^m$ possible allocations for K_{j2}'s, K_{j1}'s, and K_{j0}'s, $j = 1, 2, \ldots, m$ (genotypic means). The joint distribution of K_{j2}'s, K_{j1}'s, and K_{j0}'s, $j = 1, 2, \ldots, m$, is simply the product of the m individual multinomial distributions:

$$\begin{aligned}&\mathrm{Prob}(K_{12}, K_{11}, K_{10}, K_{22}, K_{21}, K_{20}, \ldots, K_{m2}, K_{m1}, K_{m0} \mid X_i)\\ &\quad = \prod_{j=1}^{m} \mathrm{Prob}(K_{j2}, K_{j1}, K_{j0} \mid X_i). \qquad (9)\end{aligned}$$

Under such a setting, the proposed model can be statistically formulated as a two-stage hierarchical model for the use of the EM algorithm. First the random variables x_{ij}^*'s and z_{ij}^*'s, $j = 1, 2, \ldots, m$, are sampled from a multinomial experiment

$$\begin{aligned}&(x_{i1}^*, z_{i1}^*, x_{i2}^*, z_{i2}^*, \ldots, x_{im}^*, z_{im}^*)\\ &\quad \sim h_i(x_i^*, z_i^*, x_{i2}^*, z_{i2}^*, \ldots, x_{im}^*, z_{im}^* \mid X_i)\\ &\quad = \mathrm{Prob}(K_{12}, K_{11}, K_{10}, K_{22}, K_{21}, K_{20}, \ldots,\\ &\qquad K_{m2}, K_{m1}, K_{m0} \mid X_i)\end{aligned}$$

to determine the genotypic mean $\overline{G}_i$, and then a normal variable for that genotypic mean is generated from

$$\bar{y}_i \mid (\theta, X_i, x_{i1}^*, z_{i2}^*, x_{i2}^*, z_{i2}^*, \ldots, x_{im}^*, z_{im}^*) \sim N(\overline{G}_i, \sigma^2/k),$$

where $\overline{G}_i = \mu + a_1x_{i1}^* + d_1z_{i2}^* + a_2x_{i2}^* + d_2z_{i2}^* + \cdots + a_mx_{im}^* + d_mz_{im}^*$, belonging to one of $\overline{g}_j$'s, $j = 1, 2, \ldots, [(k+1)(k+2)/2]^m$, to produce the mean trait value, $\bar{y}_i$. Following the definition of the EM algorithm, the complete-data likelihood function can be written as

$$L(\theta \mid \mathbf{Y}_{\mathrm{com}}) = \prod_{i=1}^{n}\prod_{j=1}^{[(k+1)(k+2)/2]^m}\left[p_{ij}N\left(g_j, \frac{\sigma^2}{k}\right)\right]^{I(\overline{G}_i=\overline{g}_j)}, \qquad (10)$$

where $\mathbf{Y}_{\mathrm{com}}$ contains the missing and observed data to denote the complete data. Note that the mixing proportions, p_{ij}'s, are not for estimation and can be determined by Equation 9. Following the definition of the EM algorithm. In the E-step, the conditional expected complete-data log-likelihood with respect to the conditional distribution of missing data given observed data and the current estimated parameter is computed. The M-step is to find θ to maximize the conditional expected complete-data log-likelihood. The maximization will become complicated as k or m increases. Although the derivations of the solutions in the M-step are complicated (not shown), these solutions can be regularized together in the form of the general formulas by Kao and Zeng (1997). The general formulas were originally devised to obtain the maximum-likelihood estimate (MLE) for the backcross and F_2/F_2 designs by constructing a genetic design matrix $\mathbf{D}$ to systematize the solutions into tidy formulations, and the elements in the $\mathbf{D}$ matrix are the coded variables associated with the genetic effects in all the possible genotypic values. For the (F_u/F_v, $u < v$) designs, the complicated solutions in the maximization step after regularization have the same forms of general formulas by assigning the coded variables associated with the genetic effects in the genotypic means to the elements of $\mathbf{D}$. For example, when considering $m = 1$ and $k = 3$, there are two coded variables, one for the additive effect and another for the dominance effect, and 10 possible genotypic means. The solutions are equivalent to the general formulas by constructing $\mathbf{D}$ with dimension 10×2 as

$$\mathbf{D}' = \begin{bmatrix} 1 & \frac{2}{3} & \frac{1}{3} & \frac{1}{3} & -\frac{1}{3} & 0 & 0 & -\frac{1}{3} & -\frac{2}{3} & -1 \\ -\frac{1}{2} & -\frac{1}{6} & -\frac{1}{2} & \frac{1}{6} & -\frac{1}{2} & -\frac{1}{6} & -\frac{1}{2} & \frac{1}{6} & -\frac{1}{6} & -\frac{1}{2} \end{bmatrix},$$

where the first column with elements 1 and $-\frac{1}{2}$ corresponds to the coded variables, x_i^* and z_i^*, in the first genotypic mean, $\overline{g}_1 = \mu + a - d/2$, for all the progeny with $Q^{[v]} = QQ$ ($K_2 = 3$, $K_1 = 0$, $K_0 = 0$), and the second row with elements $\frac{2}{3}$ and $-\frac{1}{6}$ is for the coded variables in the second genotypic mean, $\overline{g}_2 = \mu + 2a/3 - d/6$, for the progeny, two with QQ and one with Qq ($K_2 = 2$, $K_1 = 1$, $K_0 = 0$). The remaining eight rows are for the other possible genotypic means, $\overline{g}_3, \overline{g}_4, \ldots, \overline{g}_{10}$, corresponding with the allocations of different genotypes among progeny. For m QTL in the model, there are $[(k+1)(k+2)/2]^m$ possible genotypic means and 2^m parameters (ignoring epistasis), and the genetic design matrix has a dimension of $[(k+1)(k+2)/2]^m \times 2^m$. Each row of $\mathbf{D}$ is assigned to the values of the coded variables for the m QTL in each genotypic mean. The construction of the genetic design matrix for different m and k as well as for considering epistasis is straightforward, although the dimension expands dramatically as m or k becomes large. The E- and M-steps are iterated until convergence, and the converged values of the parameters are the MLE.

The problems if epistasis is present and ignored: Many current methods ignore epistasis in the analysis of QTL for simplicity. It is important to check the problems if epistasis is present and ignored and in addition to solve the problems in QTL mapping. Without loss of generality, consider that the quantitative trait is controlled by two unlinked epistatic QTL, Q_A and Q_B. If the trait value is regressed on Q_A (Q_B), the estimates of the additive and dominance effects can be found to be

$$a_A = a_1 - \frac{1}{2}\left[1 - \frac{1}{2^{(u-2)}}\right] i_{ad} \quad \text{and} \quad d_A = d_1 - \frac{1}{2}\left[1 - \frac{1}{2^{(u-2)}}\right] i_{dd} \tag{11}$$

in the F_u/F_u designs, where a_1 (d_1) is the additive (dominance) effect of Q_A, and i_{ad} and i_{dd} are their epistatic effects (Equations 11 can be obtained from Equations A2 and A3 by setting $u = v$ in the APPENDIX). It shows that the estimate of the a_A can be confounded by a_1 and i_{ad}, and d_A is confounded by d_1 and i_{dd}. Also, it is important to note that the epistatic effect i_{aa} is not confounding in the estimation of the marginal effects. Therefore, in the F_2/F_2 design, $a_A = a_1$ and $d_A = d_1$, and there is no confounding problem as the model has the orthogonal property in this design. In the (F_u/F_u, $u > 2$) designs, the problem of confounding occurs. For example, $a_A = a_1 - \frac{1}{4}i_{ad}$ and $d_A = d_1 - \frac{1}{4}i_{dd}$ in the F_3/F_3 design, $a_A = a_1 - \frac{3}{8}i_{ad}$ and $d_A = d_1 - \frac{3}{8}i_{dd}$ in the F_4/F_4 design, and $a_A = a_1 - \frac{7}{16}i_{ad}$ and $d_A = d_1 - \frac{7}{16}i_{dd}$ in the F_5/F_5 design. The fractions associated with the confounding epistatic effects are $\frac{1}{4}$, $\frac{3}{8}$, and $\frac{7}{16}$ in the F_3/F_3, F_4/F_4, and F_5/F_5 designs, respectively, and the confounding problem is found to become more serious in the later RI populations. This fraction approaches $\frac{1}{2}$ for the designs with large u. For large u, the dominance component may become diminished and hard to estimate, and the additive component plays the major role in estimation, due to the loss of heterozygotes and increase of homozygotes in the later F_u populations. But, for the early F_u populations, say the F_2, F_3, and F_4 populations, the dominance components may not be negligible and should be considered in the model. In addition, ignoring epistasis can inflate the sampling variances of QTL effects and will reduce the power of QTL detection. Among the epistatic variance components, the component contributed by i_{aa} is relatively larger than the components by other epistatic effects. According to Equation 5, the components contributed by i_{aa} are $\frac{1}{4}i_{aa}^2$, $\frac{9}{16}i_{aa}^2$, and $\frac{49}{64}i_{aa}^2$ in the F_2/F_2, F_3/F_3, and F_4/F_4 populations, respectively. This component becomes greater for the later RI populations and approaches i_{aa}^2 for $u \to \infty$. The above implies that QTL mapping could be problematic, such as biasing the estimation of QTL parameters and reducing the power of QTL detection, if epistasis is ignored in QTL analysis. By taking epistasis into account, the variance components contributed by the epistatic effects can be controlled to enhance the power of QTL detection and the confounding problem can be avoided to improve QTL detection.

The traditional (approximate) method and its problems: The approximate method is to model the relation between the mean trait of the F_v progeny and the QTL in their ancestral F_u individuals, $Q^{[u]}$'s. It implicitly assumes that the traits measured on the F_v progeny are controlled by the genomes of the F_u individuals, rather than by those of the progeny. This assumption overlooks the differences between population structures, as the genotypic frequencies, heterozygosity, and linkage disequilibrium between the genotyped (ancestral) and phenotyped (progeny) populations are different. Consequently, some problems, such as less power and bias in estimation, will occur (see the APPENDIX). By the approximate method, the estimate of the additive effect is confounded by the additive effect a_1 and the epistatic effect i_{ad}, and the estimate of the dominance effect is confounded by d_1 and i_{dd}. The confounding depends on u and v. For example, $b_a = a_1 - i_{ad}/4$ and $b_d = d_1/2 - i_{dd}/8$ in the F_2/F_3 design, $b_a = a_1 - \frac{3}{8}i_{ad}$ and $b_d = d_1/4 - \frac{3}{32}i_{dd}$ in the F_2/F_4 design, $b_a = a_1 - 3i_{ad}/8$ and $b_d = d_1/2 - 3i_{dd}/16$ in the F_3/F_4 design, and $b_a = a_1 - 7i_{ad}/16$ and $b_d = d_1/4 - 7i_{dd}/64$ in the F_3/F_5 design (see the APPENDIX). The estimated additive effect is an unbiased estimate of a_1, and the estimated dominance effect is only a fraction of d_1. Using the approximate method, the confounding of i_{ad} and i_{dd} in the estimation of additive and dominance effects becomes more severe for the designs with a larger difference between u and v; moreover, the confounding problems remain unsolved if epistasis is taken into account (see the APPENDIX). In addition to the confounding problem in estimation, the uncontrolled genetic variance will become a part of the genetic residual, causing loss of power in QTL detection. In general, the application of the approximate method to the QTL mapping in the (F_u/F_v, $u < v$) designs has the problems of confounding, estimating dominance effects, and controlling the genetic variances. To avoid the problems and to increase the power, it is desirable to consider the genome structure in the F_v population by using the proposed method for QTL mapping in the systems of the (F_u/F_v, $u < v$) designs.

SIMULATION STUDIES

Simulations were performed to achieve three purposes: (1) to verify the derived mapping properties of the proposed and current methods, (2) to compare the performance of the proposed and current methods in different (F_u/F_v, $v \leq v$) designs, and (3) to evaluate the relative mapping efficiency of different experimental designs. Two 100-cM chromosomes each with 11 equally spaced markers and one QTL were simulated. The two unlinked epistatic QTL, Q_A and Q_B, are assumed to be located at 25 cM on their chromosomes. The additive effects of Q_A and Q_B are assumed to be $a_1 = 2$ and $a_2 = 2$, respectively, and there is no dominance effect. Their additive-by-dominance effect is assumed to be $i_{ad} = 2$, and the other three epistatic effects are assumed to be zero. With these parameter settings, the marginal effects of the two QTL contribute 44.44% and 44.44% to the total genetic variance, respectively, and epistasis contributes 11.11% to the total genetic variance in the F_2/F_2 design. The environmental variance is assumed to be 85.5 (the heritability, h^2, is 0.05 in the F_2/F_2 design). Also, according to Equation 11, when ignoring epistasis in the estimation of a_1 and a_2, the estimate of a_1 will be confounded by i_{ad}, and the estimate of a_2 will not be confounded. The QTL, Q_A, will be referred to as the confounded QTL, and Q_B will be referred to as the unconfounded QTL. The sample size is 200, and the number of replicates is 100. The simulations include three parts. The first part is for the (F_u/F_v, $u = v$) designs. Six such designs, F_2/F_2, F_3/F_3, F_4/F_4, F_5/F_5, F_6/F_6, and F_{10}/F_{10}, are simulated. The second part is for the F_2/F_3 designs, and four different numbers of phenotyping progeny, $k = 1$, $k = 3$, $k = 5$, and $k = 10$, are assumed. The third part considers designs with other genotyping and phenotyping populations, including the F_2/F_4, F_3/F_4, F_3/F_5, and F_4/F_5 designs. The number of progeny for phenotyping is assumed to be $k = 5$. In each part, a stepwise selection procedure (Kao *et al.* 1999; Zeng *et al.* 1999) is adopted to detect QTL. Both the proposed and approximate interval mapping (IM)-based (one-QTL) and MIM-based (multiple-QTL) methods are used in the analysis. The critical value for claiming significance was chosen as $\chi^2_{k,0.05/20}$, where k is the number of parameters in testing (see discussion). The simulation results are shown in Tables 1–3.

Table 1 shows the results of the first part of the simulation. When the IM-based method is used to detect QTL, one can consider one (additive or dominance) effect or two effects in the search. The model considering dominance effect only did not detect any QTL, and the performance of the two-effect model is inferior as compared to that of the additive-effect model. Table 1 presents the QTL mapping result of the model considering the additive effect only. The powers for detecting the confounded Q_A are 30, 14, 17, 18, 11, and 17% in the F_2/F_2, F_3/F_3, F_4/F_4, F_5/F_5, F_6/F_6, and F_{10}/F_{10} designs, respectively, and the powers for detecting the unconfounded Q_B are 33, 39, 44, 50, 53, and 54% in the six different designs, respectively (critical value $\chi^2_{1,0.05/20} = 9.14$). As compared to the QTL detection in the F_2/F_2 design, the confounded Q_A was detected with decreasing power, and the unconfounded Q_B was detected with increasing power by using the later RI populations. The reasons are that the estimation of the additive effect of Q_A is confounded by i_{ad} due to ignoring epistasis and such confounding becomes more severe in the F_3, F_4, F_5, F_6, and F_{10} populations and that the estimation of the additive effect of Q_B is not confounded so that the power can be increased due to the accumulation of homozygotes in the later RI populations (Equation 11). The estimated additive effects of the confounded Q_A by the IM method are 2.13 (SD 1.40), 1.43 (SD 01.25), 1.40 (SD 1.11), 1.10 (SD 1.24), 1.08 (SD 1.17), and 1.02 (SD 1.25), respectively, in the six designs (the predicted confounded estimates by the IM method are 2.0, 1.5, 1.25, 1.125, 1.0625, and 1.00394, respectively, according to Equation 11). Except for the F_2/F_2 design, the estimates of a_1 by the IM method are poorly estimated and very far away from the true value $a_1 = 2$ due to the confounding of $i_{ad} = 2$. The confounding problem is more severe for the confounded Q_A in the designs using the later RI populations if epistasis is ignored. The estimated effects of the unconfounded Q_B are 2.20 (SD 1.46), 2.09 (SD 1.06), 2.06 (SD 0.88), 2.07 (SD 0.97), 2.23 (SD 0.65), and 2.14 (SD 0.75), respectively. These estimates by the IM method are very close to the true value $a_2 = 2$ as expected, because they are not confounded. The advantages of the MIM method include that the detected QTL can be fitted into the model for further QTL search and the epistasis between QTL can be considered. When the MIM method considers only one QTL in the model ($m = 1$), the mapping results are identical to those of the IM method. If the detected Q_A (Q_B) is fitted into the MIM model ($m = 2$ without epistasis) in the search for other QTL, both the powers of detecting Q_A and Q_B increase 1–6%. For example, the powers increase from 6% (1%) to 45% (15%) by using the MIM method without epistasis in the F_3/F_3 design. If epistasis is taken into account in the search, four different types of epistatic effects can be considered in the model. Among the four possible epistatic effects, only the model taking i_{ad} into account improves QTL detection, and the models fitting other epistasis become inferior as a higher critical value is used for claiming significance (critical value $\chi^2_{2,0.05/20} = 11.98$). By considering epistasis, the values of the average partial LRT statistic increase by 1–2 and the powers of detecting Q_A and Q_B also increase as compared to the MIM method without epistasis. For example, the powers of detecting Q_A and Q_B increase 4% (5%) and 7% (7%) to 35% (20%) and 41% (52%) in the F_2/F_2 (F_3/F_3) design after taking epistasis into account. The increase in powers of detecting Q_A for the other designs is less notable. Also, by considering epistasis, the confounding problem in the estimation of a_1 seems to be relieved by considering i_{ad} in the model. The means of the estimated a_1 are 2.06 (SD 1.33), 1.74 (SD 1.26), 1.63 (SD 1.74), 1.51 (SD 2.11), 1.52 (SD 2.22) and 1.04 (SD 2.42) in the designs, respectively, and the means of the estimated i_{ad} are 1.33 (SD 2.63), 1.50 (SD 3.01), 0.78 (SD 3.42), 1.15 (SD 4.39), 0.97 (SD 4.67), and 0.10 (SD 4.76), respectively. These estimates of a_1 and i_{ad} in the later RI populations seem to be unsatisfactory, especially for the F_{10}/F_{10} design, but they can be improved by increasing the sample size or as the heritability becomes higher (not shown) or by using the

TABLE 1

Simulation results of using different QTL mapping methods under different (F_u/F_v, $u = v$) designs

		Q_A				Q_B						
Method	Design	Posi = 25	$a_1 = 2$	LRT	Power (%)	Posi = 25	$a_2 = 2$	LRT	Power (%)	$i_{ad} = 2$	$\sigma^2 = 85.5$	LRT
IM	F_2	33.11	2.13	7.09	30	37.81	2.20	7.67	33			
		(24.07)	(1.40)	(4.84)		(26.86)	(1.46)	(5.19)				
	F_3	42.35	1.43	5.66	14	33.78	2.09	8.64	39			
		(31.10)	(1.25)	(4.11)		(21.60)	(1.06)	(4.74)				
	F_4	38.73	1.40	5.74	17	30.34	2.06	9.27	44			
		(26.19)	(1.11)	(3.7)		(17.84)	(0.88)	(5.18)				
	F_5	42.35	1.10	5.26	18	32.37	2.07	9.96	50			
		(30.26)	(1.24)	(3.84)		(22.16)	(0.97)	(5.39)				
	F_6	43.04	1.08	4.95	11	30.63	2.23	10.25	53			
		(28.33)	(1.1)	(3.27)		(16.20)	(0.65)	(5.39)				
	F_{10}	43.52	1.02	5.25	17	32.76	2.14	10.16	54			
		(27.0)	(1.25)	(3.52)		(19.43)	(0.75)	(5.41)				
MIM (without epistasis)	F_2	32.20	2.09	7.31	31	39.81	2.20	7.94	33		81.04	14.98
		(24.05)	(1.45)	(5.08)		(26.86)	(1.43)	(5.52)			(10.39)	(7.44)
	F_3	42.96	1.40	5.92	15	33.78	2.12	9.02	45		3.04	14.56
		(30.45)	(1.30)	(4.15)		(21.60)	(1.06)	(4.90)			(9.72)	(6.42)
	F_4	37.40	1.47	5.98	20	30.34	2.06	9.57	46		82.90	15.25
		(26.71)	(1.00)	(3.69)		(17.84)	(0.89)	(5.49)			(4.86)	(6.17)
	F_5	39.84	1.16	5.28	18	32.37	2.06	10.09	50		83.88	15.24
		(28.29)	(1.15)	(3.79)		(22.16)	(0.98)	(5.51)			(10.21)	(6.05)
	F_6	41.60	1.09	5.08	14	30.63	2.23	10.47	56		84.88	15.33
		(28.02)	(1.14)	(3.48)		(16.20)	(0.65)	(5.57)			(8.67)	(6.60)
	F_{10}	44.04	1.02	5.39	18	32.76	2.14	10.42	58		83.26	15.54
		(28.30)	(1.24)	(3.61)		(19.42)	(0.75)	(5.46)			(9.53)	(5.81)
MIM (with epistasis)	F_2	33.88	2.06	9.17	35	37.81	2.20	10.32	41	1.33	80.23	16.24
		(25.32)	(1.33)	(5.10)		(26.86)	(1.44)	(5.82)		(2.63)	(10.44)	(7.22)
	F_3	39.62	1.74	8.30	20	33.78	2.10	11.90	52	1.50	81.75	16.94
		(27.43)	(1.26)	(4.39)		(21.60)	(1.05)	(5.72)		(3.01)	(9.74)	(6.73)
	F_4	39.06	1.63	7.87	20	30.34	2.08	11.89	56	0.78	81.83	17.14
		(26.99)	(1.74)	(4.06)		(17.84)	(0.88)	(5.53)		(3.42)	(7.67)	(6.02)
	F_5	42.58	1.51	7.61	18	32.37	2.04	12.39	52	1.15	82.77	16.98
		(29.25)	(2.11)	(3.79)		(22.16)	(0.97)	(6.17)		(4.39)	(10.24)	(6.14)
	F_6	38.61	1.52	6.61	15	30.63	2.23	12.61	60	0.97	83.76	16.87
		(27.58)	(2.22)	(3.77)		(16.20)	(0.65)	(6.10)		(4.67)	(8.63)	(6.80)
	F_{10}	44.24	1.04	6.57	18	32.76	2.13	11.87	59	0.10	82.25	16.73
		(28.43)	(2.42)	(3.75)		(19.42)	(0.75)	(6.18)		(4.76)	(9.60)	(6.13)

A total of 100 replicates, each with sample size 200, were analyzed with two unlinked epistatic QTL, Q_A and Q_B, controlling the trait variation. The heritability is 0.05 in the F_2 population. The critical value for the methods of IM and MIM without epistasis is $\chi^2_{1,0.05/20} = 9.14$, and the value for MIM with epistasis is $\chi^2_{2,0.05/20} = 11.98$. Posi, position.

(F_u/F_v, $u < v$) designs with multiple phenotyping progeny (Table 3). The poor estimation of i_{ad} in the F_{10}/F_{10} design may be attributed to the lack of heterozygotes. In the estimation of the QTL position, the means of the estimated positions of Q_A are 33.88 (SD 25.32), 39.62 (SD 27.43), 39.06 (SD 26.99), 42.58 (SD 29.25), 38.64 (SD 27.58), and 44.24 (SD 28.43), respectively, in the six different designs, and the means for Q_B are 37.81 (SD 26.86), 33.78 (SD 21.60), 30.34 (SD 17.84), 32.37 (SD 22.16), 30.63 (SD 16.20), and 32.76 (SD 19.42), respectively. The estimated QTL positions are found to be biased toward the center of chromosomes. The position of unconfounded Q_B (confounded Q_A) seems to be estimated with greater (reduced) accuracy as the later RIL populations are used in the design. The above results shows that the use of the RI population after F_2 can improve the estimation of parameters and power of detection of the unconfounded Q_B, but it is difficult to improve the resolution of the confounded Q_A as compared to the use of the F_2/F_2 design.

Table 2 shows the QTL mapping results using the F_2/F_3 designs with different numbers of phenotyping progeny. If there is only 1 progeny ($k = 1$) for trait measurement, the IM and MIM (with or without epistasis) methods have less power to detect Q_A and Q_B as compared to the powers in the F_2/F_2 or F_3/F_3 designs. For example, the powers of detecting Q_B (Q_A) by the proposed MIM method are 32% (12%) in the F_2/F_3 design

TABLE 2

Simulation results of using different QTL mapping methods under the F_2/F_3 design with different numbers of phenotyping progeny

			Q_A				Q_B						
Design	Method		Posi = 25	$a_1 = 2$	LRT	Power (%)	Posi = 25	$a_2 = 2$	LRT	Power (%)	$i_{ad} = 2$	$\sigma^2 = 85.5$	LRT
F_2/F_3 $k = 1$	IM	e	43.07	1.43	4.69	11	37.73	2.06	6.80	26			
			(29.46)	(1.53)	(3.06)		(27.41)	(1.46)	(4.21)				
		a	43.55	1.45	4.67	10	36.78	2.05	6.71	24			
			(29.77)	(1.51)	(3.06)		(26.44)	(1.44)	(4.08)				
	MIM[a]	e	43.06	1.46	4.90	12	37.73	2.08	7.07	27		81.63	11.70
			(30.34)	(1.52)	(3018)		(27.41)	(1.46)	(4.48)			(9.99)	(5.69)
		a	43.00	1.48	4.84	11	36.78	2.06	6.94	25		84.42	11.58
			(29.91)	(1.50)	(3.26)		(26.44)	(1.45)	(4.36)			(9.90)	(5.52)
	MIM[b]	e	43.06	1.62	6.72	12	37.73	2.03	9.43	32	1.25	77.70	13.52
			(29.23)	(2.01)	(3.39)		(27.41)	(1.44)	(5.04)		(5.48)	(10.22)	(5.93)
		a	42.58	1.37	6.66	17	36.78	2.05	9.34	33	0.73	83.58	13.37
			(28.42)	(4.42)	(3.51)		(26.44)	(1.46)	(5.16)		(2.94)	(9.90)	(5.85)
F_2/F_3 $k = 3$	IM	e	30.12	1.56	8.31	37	28.16	2.04	13.89	79			
			(19.09)	(0.25)	(4.29)		(14.47)	(0.49)	(5.90)				
		a	29.68	1.57	8.30	37	26.78	2.08	13.79	79			
			(18.71)	(0.58)	(4.30)		(13.52)	(0.19)	(5.89)				
	MIM[a]	e	32.22	1.51	8.77	40	28.16	2.09	14.44	80		83.95	22.66
			(22.61)	(0.68)	(4.66)		(14.47)	(0.49)	(6.15)			(10.62)	(4.21)
		a	31.16	1.51	8.75	39	26.78	2.08	14.34	79		85.90	22.54
			(20.97)	(0.68)	(4.65)		(13.52)	(0.48)	(6.15)			(10.59)	(7.20)
	MIM[b]	e	29.60	1.82	10.42	46	28.16	2.09	15.68	83	1.48	82.2	24.31
			(22.43)	(1.04)	(4.44)		(14.47)	(0.49)	(6.81)		(2.96)	(10.59)	(4.13)
		a	28.74	1.45	10.42	47	26.78	2.09	16.42	82	0.78	85.01	24.21
			(21.36)	(0.76)	(4.39)		(13.52)	(0.49)	(6.64)		(1.49)	(10.52)	(7.10)
F_2/F_3 $k = 5$	IM	e	28.64	1.56	12.33	67	25.22	2.02	20.22	91			
			(16.09)	(0.49)	(5.89)		(8.25)	(0.47)	(8.90)				
		a	28.56	1.56	12.31	67	25.25	2.01	20.13	91			
			(16.07)	(0.48)	(5.87)		(8.23)	(0.16)	(8.86)				
	MIM[a]	e	27.12	1.51	12.92	70	25.22	2.00	21.06	92		85.99	33.14
			(13.04)	(0.49)	(6.06)		(8.25)	(0.45)	(9.09)			(10.53)	(9.67)
		a	27.04	1.52	12.97	70	25.25	2.00	20.98	92		87.68	33.14
			(12.90)	(0.49)	(6.08)		(8.23)	(0.45)	(9.01)			(10.52)	(9.66)
	MIM[b]	e	27.54	1.96	14.94	73	25.22	2.01	22.31	93	1.91	84.25	35.16
			(15.82)	(0.68)	(6.40)		(8.25)	(0345)	(9.92)		(2.18)	(10.34)	(10.07)
		a	27.10	1.48	15.00	72	25.25	2.01	23.54	93	0.98	84.25	35.16
			(16.03)	(0.49)	(6.46)		(8.23)	(0.45)	(9.79)		(1.10)	(10.36)	(10.12)
F_2/F_3 $k = 10$	IM	e	26.03	1.53	19.65	93	25.53	2.02	35.8	100			
			(8.42)	(0.40)	(8.42)		(4.49)	(0.35)	(11.62)				
		a	26.21	1.54	19.64	93	25.52	2.01	35.65	100			
			(8.31)	(0.42)	(8.03)		(4.50)	(0.35)	(11.58)				
	MIM[a]	e	26.2	1.49	22.29	99	25.53	1.99	38.64	100		85.38	58.09
			(8.27)	(0.29)	(7.93)		(4.49)	(0.33)	(12.26)			(9.24)	(13.84)
		a	26.14	1.49	22.35	99	25.52	1.99	38.25	100		54.01	58.00
			(8.34)	(0.29)	(7.92)		(4.50)	(0.33)	(12.22)			(0.92)	(13.81)
	MIM[b]	e	27.04	1.89	24.67	99	25.53	2.00	40.15	100	1.62	83.74	60.47
			(10.42)	(0.46)	(8.14)		(4.49)	(0.34)	(12.70)		(1.38)	(8.83)	(13.87)
		a	27.30	1.49	24.73	99	25.52	2.00	41.25	100	0.81	85.37	60.38
			(10.39)	(0.31)	(8.13)		(4.50)	(0.34)	(12.49)		(0.69)	(0.89)	(13.83)

A total of 100 replicates, each with sample size 200, were analyzed with two unlinked epistatic QTL, Q_A and Q_B, controlling the trait variation. The heritability is 0.05 in the F_2 population. The critical value for the methods of IM and MIM without epistasis is $\chi^2_{1,0.05/20} = 9.14$, and the value for MIM with epistasis is $\chi^2_{2,0.05/20} = 11.98$. Posi, position. *k*, the number of phenotyping progeny. e, proposed (exact) method. a, approximate method.

[a] Without epistasis.

[b] With epistasis.

TABLE 3

Simulation results of using different QTL mapping methods under different (F_u/F_v, $u < v$) designs with five phenotyping progeny

			Q_A				Q_B						
Design	Method		Posi = 25	$a_1 = 2$	LRT	Power (%)	Posi = 25	$a_2 = 2$	LRT	Power (%)	$i_{ad} = 2$	$\sigma^2 = 85.5$	LRT
F_2/F_4 $k = 5$	IM	e	30.54	1.37	9.63	47	25.23	2.02	20.38	92			
			(20.14)	(0.44)	(5.93)		(9.33)	(0.43)	(8.20)				
		a	30.37	1.37	9.62	47	25.58	2.00	20.24	92			
			(20.24)	(0.44)	(5.91)		(9.51)	(0.43)	(8.12)				
	MIM[a]	e	29.94	1.30	10.37	50	25.23	2.01	21.39	93		84.15	30.75
			(20.06)	(0.55)	(6.48)		(9.33)	(0.44)	(8.93)			(10.35)	(10.47)
		a	29.72	1.31	10.38	48	25.58	2.00	21.30	93		86.58	30.63
			(19.44)	(0.53)	(6.46)		(9.51)	(0.44)	(8.89)			(10.21)	(10.39)
	MIM[b]	e	29.02	1.80	11.77	52	25.23	2.01	22.12	94	1.36	82.25	32.15
			(19.96)	(1.71)	(6.43)		(9.33)	(0.44)	(9.47)		(4.36)	(10.45)	(10.59)
		a	29.42	1.30	11.73	50	25.58	2.00	22.82	93	0.34	85.87	31.98
			(19.84)	(0.54)	(6.42)		(9.51)	(0.44)	(9.12)		(1.11)	(10.16)	(10.47)
F_3/F_4 $k = 5$	IM	e	27.05	1.33	13.12	68	24.79	2.04	29.35	99			
			(13.44)	(0.41)	(7.63)		(4.63)	(0.36)	(9.91)				
		a	27.02	1.33	13.12	68	24.75	2.03	29.25	99			
			(13.48)	(0.41)	(7.64)		(4.68)	(0.36)	(9.89)				
	MIM[a]	e	26.64	1.28	13.93	75	24.79	2.01	30.49	99		83.67	43.29
			(14.33)	(0.39)	(7.24)		(4.63)	(0.35)	(10.15)			(8.65)	(11.72)
		a	26.72	1.28	13.98	75	24.75	2.01	30.42	99		74.49	43.23
			(14.29)	(0.39)	(7.24)		(4.68)	(0.36)	(10.16)			(8.63)	(11.73)
	MIM[b]	e	25.60	1.92	16.53	84	24.79	2.01	32.44	99	1.81	81.78	45.89
			(14.04)	(0.87)	(7.10)		(4.63)	(0.36)	(10.85)		(2.41)	(8.53)	(11.67)
		a	25.78	1.48	16.63	84	24.75	2.01	33.40	99	0.94	82.98	45.88
			(14.04)	(0.43)	(7.14)		(4.68)	(0.36)	(10.69)		(1.22)	(8.46)	(11.72)
F_3/F_5 $k = 5$	IM	e	28.45	1.14	10.69	62	24.91	1.99	28.59	99			
			(16.47)	(0.53)	(6.33)		(4.98)	(0.37)	(10.12)				
		a	28.47	1.14	10.69	61	24.95	1.99	28.53	99			
			(16.45)	(0.53)	(6.35)		(5.04)	(0.37)	(10.10)				
	MIM[a]	e	27.74	1.11	11.64	67	24.91	1.98	29.79	100		83.19	40.23
			(17.13)	(0.50)	(6.54)		(4.98)	(0.37)	(10.90)			(9.60)	(12.46)
		a	27.76	1.11	11.66	68	24.95	1.98	29.67	100		84.31	40.19
			(17.10)	(0.50)	(6.56)		(5.04)	(0.37)	(10.90)			(9.53)	(12.47)
	MIM[b]	e	28.88	1.49	13.17	71	24.99	1.98	30.72	100	0.97	81.75	41.76
			(18.90)	(1.93)	(6.74)		(4.98)	(0.37)	(11.20)		(4.08)	(9.54)	(12.56)
		a	28.38	1.17	13.36	71	24.95	1.98	31.84	100	0.34	83.35	41.89
			(18.38)	(0.63)	(6.67)		(5.04)	(0.37)	(10.98)		(1.23)	(9.42)	(12.44)
F_4/F_5 $k = 5$	IM	e	27.27	1.20	11.76	59	26.82	1.96	30.73	99			
			(12.16)	(0.33)	(6.09)		(4.27)	(0.38)	(11.22)				
		a	27.28	1.20	11.76	59	26.83	1.96	30.67	99			
			(12.14)	(0.33)	(6.08)		(4.27)	(0.38)	(11.19)				
	MIM[a]	e	26.96	1.17	13.28	70	26.82	1.95	32.53	99		85.6	44.01
			(11.08)	(0.36)	(6.99)		(4.27)	(0.39)	(12.28)			(10.73)	(13.64)
		a	27.18	1.17	13.30	69	26.83	1.96	32.53	99		85.95	43.97
			(11.13)	(0.36)	(7.00)		(4.27)	(0.39)	(12.29)			(10.72)	(13.64)
	MIM[b]	e	26.86	1.86	15.18	75	26.82	1.95	33.37	99	1.65	84.07	45.90
			(10.93)	(1.12)	(7.23)		(4.27)	(0.39)	(12.85)		(2.58)	(10.49)	(13.66)
		a	26.94	1.47	15.28	75	26.83	1.95	34.59	99	0.87	84.68	45.95
			(10.92)	(0.56)	(7.30)		(4.27)	(0.39)	(12.41)		(1.35)	(10.51)	(13.68)

A total of 100 replicates, each with sample size 200, were analyzed with two unlinked epistatic QTL, Q_A and Q_B, controlling the trait variation. The heritability is 0.05 in the F_2 population. The critical value for the methods of IM and MIM without epistasis is $\chi^2_{1,0.05/20} = 9.14$, and the value for MIM with epistasis is $\chi^2_{2,0.05/20} = 11.98$. Posi, position. *k*, the number of phenotyping progeny. e, proposed (exact) method. a, approximate method.

[a] Without epistasis.

[b] With epistasis.

with $k = 1$, and they are 41% (35%) and 52% (20%) in the F_2/F_2 and F_3/F_3 designs. By increasing the number of phenotyping progeny, the powers of detecting Q_A and Q_B are enhanced. By using 3 ($k = 3$), 5 ($k = 5$), and 10 ($k = 10$) progeny for phenotyping, the powers of detecting the unconfounded Q_B increase to 83, 93, and 100%, respectively, and the powers of detecting the confounded Q_A become 46, 73, and 99%, respectively. As compared to the results of the IM method in the F_2/F_2 and F_3/F_3 designs, the use of three progeny for phenotyping can greatly enhanced the power of detecting the unconfounded Q_B from 33% (or 39%) to 79%, but it does not greatly increase the power of detecting the confounded Q_A [the powers of detecting Q_A in the F_2/F_2, F_3/F_3, and (F_2/F_3, $k = 3$) designs are 30, 14, and 37%, respectively]. The use of more progeny for phenotyping also improves the estimation of QTL effects and positions. For example, by using the proposed MIM method with epistasis, the means of the estimated Q_B (Q_A) positions are 37.73 cM with SD 27.41 cM (43.06 cM with SD 29.23 cM) and 25.22 cM with SD 8.25 cM (27.54 cM with SD 15.82 cM) for $k = 1$ and $k = 5$. The estimated a_1, a_2, and i_{ad} are 1.62 (SD 2.01), 2.03 (SD 1.44), and 1.25 (SD 5.48) for $k = 1$, and they are 1.96 (SD 0.68), 2.01 (SD 0.45), and 1.91 (SD 2.18) for $k = 5$. In addition, if epistasis is not taken into account or the approximate method is used, there will always be confounding problems in estimation as shown in Table 2. For example, the means of the estimated a_1 and i_{ad} by the approximate MIM method are 1.48 (predicted value $a_1 - i_{ad}/4 = 1.5$) with SD 0.49 and 0.98 (predicted value $i_{ad}/2 = 1.0$) with SD 1.10 for $k = 5$. By using 10 progeny for phenotyping, both Q_A and Q_B can almost be detected with power 1 and with good precision and accuracy. In general, the estimation becomes improved as more progeny are used for phenotyping. The performance of the MIM method is also better than that of the IM method as expected.

Table 3 shows the QTL mapping results of using the (F_u/F_v, $v > u > 2$) designs with $k = 5$. The means of the estimated additive effects by the proposed IM and MIM without epistasis are 1.33 (SD 0.41) and 1.28 (SD 0.39), respectively, in the F_3/F_4 design, and they are 1.20 (SD 0.33) and 1.17 (SD 0.36), respectively, in the F_4/F_5 design (the predicted estimated a_1's by Equation 11 are 1.25 and 1.125 in the two designs). By taking epistasis into account, the confounding problem can be solved by the proposed method. The means of the estimated a_1 by considering epistasis are 1.92 (SD 0.87) and 1.86 (SD 1.12) in the two designs, respectively. The means of the estimated i_{ad}'s are 1.81 (SD 2.41) and 1.65 (SD 2.58), respectively. On the contrary, the approximate methods always have the confounding problem whether or not epistasis is taken into account. For example, the means of the estimated additive effects by the approximate MIM with epistasis are 1.48 (SD 0.43) and 1.47 (SD 0.56) for the two designs, respectively (the predicted a_1 by the approximate method is 1.5). The powers of QTL detection are also increasing and the estimation of QTL positions is also improved by using the MIM approach and by taking epistasis into account. For example, the power increases from 68% (59%) by the IM method to 84% (75%) by the MIM method in the F_3/F_4 (F_4/F_5) design, and the mean of the estimated position of the confounded Q_A is improved from 27.05 with SD 13.44 (27.27 with SD 12.16) to 25.60 with SD 14.04 (26.86 with SD 10.93). In addition, the use of the F_2/F_4 and F_3/F_5 designs does not provide a better resolution of the confounded Q_A when compared to the use of F_2/F_3 and F_3/F_4 designs as expected. For example, the powers of detecting Q_A by the proposed MIM method are 52 and 71%, respectively, in the two designs, and the means of the estimated positions are 29.02 cM (SD 19.96 cM) and 28.88 cM (SD 18.90 cM), respectively. Across all different designs with $k = 5$, the unconfounded Q_B can be well detected with high power and great accuracy and precision as compared to the confounded Q_A.

DISCUSSION

The data required in QTL mapping analysis are usually composed of two parts, phenotypic trait values and marker genotypes. In data collection using the designs of RI populations, the trait values can be obtained from the same genotyped population by using the (F_u/F_v, $u = v$) designs or from the progeny of the genotyped population using the (F_u/F_v, $u < v$) designs. The great benefit of using the (F_u/F_v, $u < v$) designs in QTL mapping is not only through reducing the cost and environmental variance by phenotyping several progeny for each genotyped individual (Lander and Botstein 1989; Knapp and Bridges 1990), but also likely through taking advantage of the changes in population structures between different RI populations. Different RI populations have different homozygosities, genotypic frequencies, and proportions of recombinant genotypes. The increase of homozygosity may help the estimation of additive effects due to the accumulation of homozygotes, but it will hinder the estimation of dominance effects due to the loss of heterozygotes. Also, in modeling, the orthogonal property of the genetic model, which holds in the F_2 population, will be lost in the other later RI populations as the genotypic frequencies have changed. Then, the confounding problem in the QTL estimation may occur if epistasis is present and ignored. Such a confounding problem cannot be relieved by enlarging the sample size or increasing heritability or using the approximate methods, and it becomes more severe for the later populations (see Equations 11 and Table 1). Therefore, the use of the later RI populations can greatly benefit the detection of unconfounded QTL with additive effects, but it may deter the detection of confounded QTL and the QTL with large dominance effects as compared to the QTL mapping using the F_2 population. By taking epistasis into account, the confounding problem can be alleviated by the proposed method. The approximate method, however, always has the confounding problems. In addition, the (F_u/F_v, $u < v$) designs also allow for phenotyping more progeny for each genotyped individual to reduce environmental variance so that the resolution of QTL can be further enhanced. The resolution of the unconfounded QTL can be easily improved, but more progeny are needed to improve the resolution of the confounded QTL as compared to the F_u/F_u designs (comparing the results of Table 1 with those of Tables 2 and 3).

In statistical modeling, the relation between the phenotype and the underlying QTL genotype is relatively simple and can be modeled by a 3^m normal mixture model for the (F_u/F_v, $u = v$) designs. However, for the (F_u/F_v, $u < v$) designs, when the phenotypic means of the k progeny from the genotyped individuals are used in QTL mapping, the relationship between the phenotypic means and the involved QTL genotypes becomes increasingly complicated and should be modeled by a $[(k+1)(k+2)/2]^m$ normal mixture model as discussed here. Such complication in statistical modeling arises mainly from the segregation of heterozygote into homozygotes and heterozygote and from the numerous possible combinations of different genotypes among the k progeny. Genetically, segregation will vary the homozygosity and linkage disequilibrium in different RI populations. It is possible to utilize different experimental designs of these RI populations to benefit QTL mapping by taking advantage of their specific population structures. To achieve this purpose, for QTL mapping in the (F_u/F_v, $u < v$) design, the proposed method is designed to take the population structures of phenotyping populations into account by modeling the relationship between the phenotypic means and the underlying QTL in the same populations. Then, the likelihood of the proposed method is a mixture of $[(k+1)(k+2)/2]^m$ normals with the number of mixture components and mixing proportions adjusted for the phenotyping population. The approximate method, however, ignores the fact of segregation and differences in population structure between different RI populations, and it relates the phenotypic traits of the progeny with the QTL in their ancestral populations. Therefore, the likelihood of the approximate method is always a mixture of 3^m normals with constant mixing proportions derived from the genotyped population. Consequently, the approximate method may have the problems of confounding and estimating the dominance effect, and the proposed method can avoid the problems to improve the QTL mapping in (F_u/F_v, $u < v$) designs as shown in this article. In addition, it is straightforward to modify the proposed method for the (F_u/F_v, $u < v$) designs with each individual progeny trait (y_{ij}'s, $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, k$) recorded. The mapping results by the approaches of using traits and trait means are similar, but the approach of using individual traits can be more computationally economical for its relatively simple likelihood (with a mixture of 3^m normals).

The proposed method has a much more complicated mixture likelihood, and the mixture likelihood will have different numbers of components with different weights (mixing proportions) for different designs. Therefore, the determination of the critical value for the proposed method is challenging in the (F_u/F_v, $u \leq v$) designs. It is well known that the critical value cannot be simply chosen from a χ^2-distribution because of violation of the standard conditions of asymptotic theory for mixture models (Self and Liang 1987; Feng and McCulloch 1994) and that the determination of the critical values for claiming QTL detection may depend on the factors, such as heritability, marker density, size of the genomes, number of (linked or unlinked) QTL, and the direction of QTL effects (Jensen 1993; Zeng *et al.* 1999; Zou *et al.* 2004). Several methods, such as the method by Piepho (2001), the permutation tests (Churchill and Doerge 1994), residual bootstrapping (Zeng *et al.* 1999), and the resampling method by Zou *et al.* (2004), have been proposed to determine the values, but they generally require additional assumptions, such as a dense map with equally spaced markers, or are applicable only to some standard designs, such as a backcross or F_2 design (the F_u/F_v, $u = v$ design), or are restricted to the model with a 2- (3-) normal mixture (see Zou *et al.* 2004 for the discussion). In addition, the concept of the false discovery rate (Benjamini and Hochberg 1995) has been introduced to deal with the problem of statistical significance by the control of type II rather than type I errors in QTL mapping. As the proposed method considers a more complicated mixture of ($[(k+1)(k+2)/2]^m$) normals with different numbers of components and mixing proportions varying with m, k, u, and v, and the different population structures may also affect the critical values, the issue of determining the critical values in the (F_u/F_v, $u \leq v$) designs will become even more complicated and still needs to be unraveled. Here, Bonferroni argument based on χ^2-distribution (Lander and Botstein 1989) is used to choose the critical values before the complicated issue is solved. Further research on the theoretical basis of determining the critical value is of great value to QTL mapping in the (F_u/F_v, $u \leq v$) designs.

The RI populations have been very important and popular in the study of QTL for a long time (Haldane and Waddington 1931; Stuber *et al.* 1992; Beavis *et al.* 1994; Veldboom *et al.* 1994; Darvasi and Soller 1995; Austin and Lee 1996; Liu *et al.* 1996; Belknap 1998; Chapman *et al.* 2003; Complex Trait Consortium 2004; Broman 2005). As compared to the F_2 population, the population structures in the later RI populations have some precious properties, such as larger additive genetic variance, higher homozygosity, and more recombinants. These properties may benefit the QTL resolution and should be well utilized in the study of QTL mapping. With the ability to consider the changes in population structures of different populations, the proposed method can serve as an effective tool to map for QTL in specific designs and evaluate the efficiency of QTL mapping among different experimental designs under the system of RI populations. Other important issues of QTL mapping by using the (F_u/F_v, $u \leq v$) designs include the consideration of endosperm traits (Wu *et al.* 2002; Xu *et al.* 2003; Kao 2004) and the extension of the methods from the system of RI populations to the system of IRI populations. The IRI populations, which are derived by randomly mating for some generations after F_2 and then followed by cycles of selfing, have the advantages of producing more recombinants as compared to the RI populations, and they can benefit the analysis of quantitative traits (Liu *et al.* 1996; Winkler *et al.* 2003). It is critical to provide adequate statistical methods for these designs by considering their specific population structures to explore their properties in the QTL mapping study.

The author thanks two anonymous reviewers for helpful comments. This study was supported by grants NSC94-2118-M-001-021 from the National Science Council, Taiwan, Republic of China.

LITERATURE CITED

Austin, D. F., and M. Lee, 1996 Comparative mapping in F2:3 and F6:7 generations of quantitative trait loci for grain yield and yield components in maize. Theor. Appl. Genet. **92:** 817–826.

Beavis, W. D., O. S. Smith, D. Grant and R. Fincher, 1994 Identification of quantitative trait loci using a small sample of topcrossed and F4 progeny from maize. Crop Sci. **34:** 882–896.

Belknap, J. K., 1998 Effect of within-strain sample size on QTL detection and mapping using recombinant inbred mouse strains. Behav. Genet. **28:** 29–37.

Benjamini, Y., and Y. Hochberg, 1995 Controlling false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B **57:** 289–300.

Broman, K. W., 2005 The genomes of recombinant inbred lines. Genetics **169:** 1133–1146.

Carlborg, O., and C. S. Haley, 2004 Epistasis: too often neglected in complex trait studies??. Nat. Rev. Genet. **5:** 618–625.

Chapman, A., V. R. Pantalone, A. Ustun, E. L. Allen, D. Landau-Ellis *et al.*, 2003 Quantitative trait loci for agronomic and seed quality traits in an F2 and F4:6 soybean population. Euphytica **129:** 387–393.

Churchill, G. A., and R. W. Doerge, 1994 Empirical threshold values for quantitative trait mapping. Genetics **138:** 967–971.

Complex Trait Consortium, 2004 The Collaborative Cross, a community resource for the genetic analysis of complex traits. Nat. Genet. **36:** 1133–1137.

Comstock, R. E., and H. F. Robinson, 1952 Estimation of average dominance of genes, pp. 495–516 in *Heterosis*, edited by J. W. Gowen. Iowa State College Press, Ames, IA.

Cowen, N. M., 1988 The use of replicated progenies in marker-based mapping of QTLs. Theor. Appl. Genet. **75:** 857–862.

Darvasi, A., and M. Soller, 1995 Advanced intercross lines, an experimental population for genetic fine mapping. Genetics **141:** 1199–1207.

Dempster, A. P., N. M. Laird and D. B. Rubin, 1977 Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc. **39:** 1–38.

Doerge, R. W., Z.-B. Zeng and B. S. Weir, 1997 Statistical issues in the search for genes affecting quantitative traits in experimental populations. Stat. Sci. **12:** 195–219.

Edwards, M. D., T. Helentjaris, S. Wright and C. W. Stuber, 1992 Molecular-marker-facilitated investigations of quantitative-trait loci in maize. IV. Analysis based on genome saturation with isozyme and restriction fragment length polymorphism markers. Theor. Appl. Genet. **83:** 765–774.

Feng, Z. D., and C. E. McCulloch, 1994 On the likelihood ratio test statistic for the number of components in a normal mixture with unequal variances. Biometrics **50:** 1158–1162.

Fisch, R. D., M. Ragot and G. Gay, 1996 A generalization of the mixture model in the mapping of quantitative trait loci for progeny from a biparental cross of inbred lines. Genetics **143:** 571–577.

Haldane, J. B. S., and C. H. Waddington, 1931 Inbreeding and linkage. Genetics **16:** 357–374.

Jensen, R. C., 1993 Interval mapping of multiple quantitative trait loci. Genetics **135:** 205–211.

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

Kao, C.-H., and Z.-B. Zeng, 1997 General formulas for obtaining the MLEs and the asymptotic variance-covariance matrix in mapping quantitative trait loci when using the EM algorithm. Biometrics **53:** 359–371.

Kao, C.-H., and Z.-B. Zeng, 2002 Modeling epistasis of quantitative trait loci using Cockerham's model. Genetics **160:** 1243–1261.

Kao, C.-H., Z.-B. Zeng and R. D. Teasdale, 1999 Multiple interval mapping for quantitative trait loci. Genetics **152:** 1203–1216.

Knapp, S. J., and W. C. Bridges, 1990 Using molecular markers to estimate quantitative trait locus parameters: power and genetic variances for unreplicated and replicated progeny. Genetics **126:** 769–777.

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

Liu, S. C., S. P. Kowalski, T. H. Lan, K. A. Feldmann and A. H. Paterson, 1996 Genome-wide high-resolution mapping by recurrent intermating using *Arabidopsis thaliana* as a model. Genetics **142:** 247–258.

Lynch, M., and B. Walsh, 1998 *Genetics and Analysis of Quantitative Traits.* Sinauer Associates, Sunderland, MA.

Mihaljevic, R., H. F. Utz and A. E. Melchinger, 2004 Congruency of quantitative trait loci detected for agronomic traits in testcrosses of five populations of European maize. Crop Sci. **44:** 114–124.

Mihaljevic, R., C. C. Schon, H. F. Utz and A. E. Melchinger, 2005 Correlations and QTL correspondence between line per se and testcross performance for agronomic traits in four populations of European maize. Crop Sci. **45:** 114–122.

Nakamichi, R., Y. Ukai and H. Kishino, 2001 Detection of closely linked multiple quantitative trait loci using a genetic algorithm. Genetics **158:** 463–475.

Piepho, H. P., 2001 A quick method for computing approximate threshold for quantitative trait loci detection. Genetics **157:** 425–432.

Sala, R. G., F. H. Andrade, E. L. Camadro and J. C. Cerono, 2006 Quantitative trait loci for grain moisture at harvest and field grain drying rate in maize (Zea mays, L.). Theor. Appl. Genet. **112:** 462–471.

Satagopan, J. M., B. S. Yandell, M. A. Newton and T. C. Osborn, 1996 A Bayesian approach to detect quantitative trait loci using Markov chain Monte Carlo. Genetics **144:** 805–816.

Self, S. G., and K. Y. Liang, 1987 Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. J. Am. Stat. Assoc. **82:** 605–610.

Sen, S., and G. A. Churchill, 2001 A statistical framework for quantitative trait mapping. Genetics **159:** 371–387.

Stuber, C. W., S. E. Linncoln, D. W. Wolff, T. Helentjaris and E. S. Lander, 1992 Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular marks. Genetics **132:** 823–839.

Veldboom, L. R., M. Lee and W. L. Woodman, 1994 Molecular marker-facilitated studies in an elite maize population: I. Linkage analysis and determination of QTL for morphological traits. Theor. Appl. Genet. **88:** 7–16.

Weir, B. S., 1996 *Genetic Data Analysis II.* Sinauer Associates, Sunderland, MA.

Winkler, C. R., N. M. Jensen, M. Cooper, D. W. Podlich and O. S. Smith, 2003 On the determination of recombination rates in intermated recombinant inbred populations. Genetics **164:** 741–745.

Wu, R.-L., C.-X. Ma, M. Gallo-Meagher, R. C. Littell and G. Casella, 2002 Statistical methods for dissecting triploid endosperm traits using molecular markers: an autogamous model. Genetics **162:** 875–892.

Xu, C., X. He and S. Xu, 2003 Mapping quantitative trait loci underlying triploid endosperm traits. Heredity **90:** 228–235.

Yi, N., S. Xu and D. B. Allison, 2003 Bayesian model choice and search strategies for mapping interacting quantitative trait loci. Genetics **165:** 867–883.

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

Zeng, Z.-B., C.-H. Kao and C. Basten, 1999 Estimating the genetic architecture of quantitative traits. Genet. Res. **74:** 279–289.

Zhang, Y. M., and S. Xu, 2004 Mapping quantitative trait loci in F_2 incorporating phenotypes of F_3 progeny. Genetics **166:** 1981–1993.

Zou, F., J. P. Fine, J. Hu and D. Y. Lin, 2004 An efficient resampling method for assessing genome-wide statistical significance in mapping quantitative trait loci. Genetics **168:** 2307–2316.

Communicating editor: A. D. Long

APPENDIX

If m QTL without epistasis are considered, the model of the traditional method for the mean trait, $\bar{y}_i$, of the k F_v progeny from each of the n genotyped F_u individuals and the QTL can be written as

$$\bar{y}_i = \mu + \sum_{j=1}^{m} b_{aj} w^*_{a_{ij}} + \sum_{j=1}^{m} b_{dj} w^*_{d_{ij}} + \bar{\epsilon}_i, \qquad \text{(A1)}$$

where $w^*_{a_{ij}}$'s and $w^*_{d_{ij}}$'s, $j = 1, 2, \ldots, m$, are the coded variables for the genotype of $Q_j^{[u]}$'s, $j = 1, 2, \ldots, m$, and they are coded as $(1, -\frac{1}{2})$, $(0, \frac{1}{2})$, and $(-1, -\frac{1}{2})$ for Q_jQ_j, Q_jq_j, and q_jq_j, respectively. The mean residual error $\bar{\epsilon}_i$ has a mean of zero and variance σ^2/k, where σ^2 is the residual variance of the trait on the basis of a single individual. As $Q^{[u]}$ may not be coincident with marker and could be Q_jQ_j, Q_jq_j, or q_jq_j, the likelihood of the model is a mixture of 3^m normals. In parameter estimation, the general formulas by Kao and Zeng (1997) derived on the basis of the EM algorithm (Dempster *et al.* 1977) can be used for obtaining their MLE.

To show the problems of less power and bias in estimation for the approximate method, without loss of generality, again assume that the quantitative trait value y_i is affected by two unlinked epistatic QTL, Q_A and Q_B. The variances of $w^*_{a_1}$, $w^*_{d_1}$, $w^*_{a_2}$, $w^*_{d_2}$, $w^*_{a_1} \times w^*_{a_2}$, $w^*_{a_1} \times w^*_{d_2}$, $w^*_{a_2} \times w^*_{d_1}$, and $w^*_{d_1} \times w^*_{d_2}$ can be found to be $1 - (\frac{1}{2})^{u-1}$, $(\frac{1}{2})^{u-1}[1 - (\frac{1}{2})^{u-1}]$, $1 - (\frac{1}{2})^{u-1}$, $(\frac{1}{2})^{u-1}[1 - (\frac{1}{2})^{u-1}]$, $[1 - (\frac{1}{2})^{u-1}]^2$, $\frac{1}{4} - (\frac{1}{2})^{(u+1)}$, $\frac{1}{4} - (\frac{1}{2})^{(u+1)}$, and $\frac{1}{16} - [\frac{1}{2} - (\frac{1}{2})^{(u-1)}]^4$, respectively. The covariances between $w^*_{a_1}$ ($w^*_{d_1}$) and the coded variables in Equation 6 are needed in the derivation. The covariances between $w^*_{a_1}$ and x_1 and between $w^*_{a_1}$ and w_{ad} are $1 - (\frac{1}{2})^{u-1}$ and $[\frac{1}{2} - (\frac{1}{2})^{u}][(\frac{1}{2})^{v-2} - 1]$, respectively. The covariances between $w^*_{d_1}$ and z_1 and between $w^*_{a_1}$ and w_{dd} are $(\frac{1}{2})^{v-1}[1 - (\frac{1}{2})^{u-1}]$ and $(\frac{1}{2})^{v}[1 - (\frac{1}{2})^{v-2}][(\frac{1}{2})^{(u-1)} - 1]$, respectively. Therefore, when y_i is regressed on Q_A with additive and dominance effects, the estimates of the additive and dominance effects by the approximate method can be found to be

$$b_a = a_1 - \frac{1}{2}\left[1 - \frac{1}{2^{(v-2)}}\right] i_{ad} \qquad \text{(A2)}$$

and

$$b_d = \frac{1}{2^{(v-u)}} d_1 - \frac{1}{2^{(v-u+1)}}\left[1 - \frac{1}{2^{(v-2)}}\right] i_{dd} \qquad \text{(A3)}$$

in the (F_u/F_v, $u < v$) design. It shows that the b_a is confounded by the the additive effect a_1 and the epistatic effect i_{ad}, and b_d is confounded by d_1 and i_{dd}. When multiple QTL and their epistasis are considered in the model, the estimates of their effects can be also derived. It can be found that the approximate methods also have the confounding problems.