Mapping Quantitative Trait Loci Using Multiple Families of Line Crosses

Shizhong Xu

Department of Botany and Plant Sciences, University of California, Riverside, California 92521

Manuscript received July 6, 1997
Accepted for publication September 29, 1997

ABSTRACT

To avoid a loss in statistical power as a result of homozygous individuals being selected as parents of a mapping population, one can use multiple families of line crosses for quantitative trait genetic linkage analysis. Two strategies of combining data are investigated: the fixed-model and the random-model strategies. The fixed-model approach estimates and tests the average effect of gene substitution for each parent, while the random-model approach treats each effect of gene substitution as a random variable and directly estimates and tests the variance of gene substitution. Extensive Monte Carlo simulations verify that the two strategies perform equally well, although the random model is preferable in combining data from a large number of families. Simulations also show that there may be an optimal sampling strategy (number of families *vs.* number of individuals per family) in which QTL mapping reaches its maximum power and minimum estimation error. Deviation from the optimal strategy reduces the efficiency of the method.

LINE crossing is a common experimental design for mapping quantitative trait loci (QTLs) in plants and laboratory animals. Statistical methods are well developed for QTL mapping using line-crossing data (Lander and Botstein 1989; Haley and Knott 1992; Martínez and Curnow 1992; Jansen 1993, 1994; Zeng 1994). Methods developed by these authors are mainly designed to handle a single cross, *e.g.*, a single F_2 family. Under these methods, the effects of gene substitution (the first moments) are tested and estimated. Because of this, the methods are classified by Xu and Atchley (1995) as the fixed-model approach. The sampling strategy (using a single family) and the statistical methodology (the fixed model) consequently restrain the inference space of the parameter estimation to the particular cross. This is undesirable if the two lines initiating the cross are not segregating at a QTL, for then no matter how many offspring are sampled in the F_2 or backcross population, the QTL cannot be detected. If a QTL is present, but is not detected because of fixation to the same allele in both lines, then a type of type II error has occurred. This type II error, referred to as genetic drift error by Xu (1996a), has largely been ignored in the QTL mapping literature.

A type II error of this kind can be reduced or even prevented by using multiple families of line crosses. At the low end, Muranty (1996) claims that QTL detection in a population derived from two parents is often less powerful than one derived from more parents. He then demonstrates that if QTL heterozygote frequency in the base population is high enough, a mating design with six parents should give a good sample of variance and allow the detection of QTL with reasonable power. Muranty (1996) introduced the idea of multiple-family QTL mapping by using an ideal situation in which the genotype of the QTL is known without error. In actuality, the QTL genotype cannot be observed, so the statistical method demonstrated by Muranty should be modified before it is applied to genome scanning using real data.

In this paper, I propose two strategies for combining data from multiple families of line crosses: the fixed-model and the random-model approaches. I then conduct Monte Carlo simulations to show that both the fixed- and the random-model approaches work as expected.

METHODOLOGY

Linear model: Consider *t* independent F_2 families each derived from cross of two inbred lines (a total of 2*t* independent inbred lines are involved). The phenotypic value of a quantitative character can be described by the following linear model:

$$y_{ij} = \mu + \beta_i + z_{ij}\alpha_i + w_{ij}\delta_i + \varepsilon_{ij} \qquad (1)$$

where y_{ij} is the phenotypic value of a trait under consideration for the *j*-th individual in the *i*-th family, μ is the overall mean, β_i is an unknown family-specific effect, α_i and δ_i are the respective effect of allelic substitution and the dominance deviation at the QTL of interest, and ε_{ij} is the residual error distributed as $N(0,\sigma^2_\varepsilon)$. The variables z_{ij} and w_{ij} are defined as follows:

$$z_{ij} = \begin{cases} +1 & \text{if } Q_1Q_1 \\ 0 & \text{if } Q_1Q_2 \\ -1 & \text{if } Q_2Q_2 \end{cases} \qquad (2)$$

and

Author e-mail: xu@genetics.ucr.edu

$$w_{ij} = \begin{cases} +1 \text{ if } Q_1Q_2 \\ -1 \text{ if } Q_1Q_1 \text{ or } Q_2Q_2 \end{cases} \tag{3}$$

where Q_1 Q_1, Q_2 Q_2 and Q_1 Q_2 represent genotypes of the two parental lines and the F_1 hybrid, respectively, for the *i*-th family at the candidate QTL. The maximum number of alleles at each locus (QTL or markers) is two in each F_2 family, but this number can be arbitrary in the whole population where the inbred lines are sampled. Since the genotype of a QTL is not observable, z_{ij} and w_{ij} are missing. Let $p_{(kl)j}$ be the conditional probability that the individual is of genotype Q_k Q_l given information of marker genotypes. This conditional probability is derived based on genotypes of the nearest flanking markers (Haley and Knott 1992).

Let $E(z_{ij}|I_M)$ and $E(w_{ij}|I_M)$ be the conditional expectations of z_{ij} and w_{ij} given marker information (I_M). The linear model can be approximated by substituting z_{ij} and w_{ij} by their conditional expectations (Haley and Knott 1992; Martínez and Curnow 1992),

$$y_{ij} = \mu + \beta_i + E(z_{ij}|I_M)\alpha_i + E(w_{ij}|I_M)\delta_i + e_{ij} \tag{4}$$

where $E(z_{ij}|I_M) = (+1)\,p_{(11)j} + (0)\,p_{(12)j} + (-1)\,p_{(22)j} = p_{(11)j} - p_{(22)j}$ and $E(w_{ij}|I_M) = (+1)\,p_{(12)j} + (-1)\,[p_{(11)j} + p_{(22)j}]$. The residual variance is

$$\begin{aligned} \mathrm{Var}(e_{ij}) = {} & \mathrm{Var}(z_{ij}|I_M)\alpha_i^2 + \mathrm{Var}(w_{ij}|I_M)\delta_i^2 + \\ & 2\mathrm{Cov}(z_{ij}w_{ij}|I_M)\alpha_i\delta_i + \sigma_\varepsilon^2 \end{aligned} \tag{5}$$

where $\mathrm{Var}(z_{ij}|I_M)\alpha_i^2$ is the variance of the QTL effect that is not explained because of uncertainty in z_{ij}, $\mathrm{Var}(w_{ij}|I_M)\delta_i^2$ is the variance of the QTL effect that is not explained because of uncertainty in w_{ij}, and $\mathrm{Cov}(z_{ij}w_{ij}|I_M)\alpha_i\delta_i$ is the covariance because of uncertainty in both z_{ij} and w_{ij}. All three additional components in the residual variance will vanish if the genotype of the QTL is actually observed, *i.e.*, $\mathrm{Var}(z_{ij}|I_M) = \mathrm{Var}(w_{ij}|I_M) = \mathrm{Cov}(z_{ij}\ w_{ij}||I_M) = 0$. Otherwise, $\mathrm{Var}(z_{ij}|I_M) = p_{(11)j}[1 - p_{(11)j}] + p_{(22)j}[1 - p_{(22)j}] + 2p_{(11)j}p_{(22)j}$, $\mathrm{Var}(w_{ij}|I_M) = 4p_{(12)j}[1 - p_{(12)j}]$ and $\mathrm{Cov}(z_{ij}\ w_{ij}||I_M) = p_{(22)j}[1 - p_{(22)j}] - p_{(11)j}[1 - p_{(11)j}] + p_{(12)j}p_{(22)}j - p_{(12)j}p_{(11)j}$.

Fixed model strategy: The first strategy of combining several different line crosses is to estimate and test $\{\alpha_i\ \delta_i\}$ for $i = 1, \ldots, t$. The null hypothesis is H_0: $\alpha_i = \delta_i = 0\ \forall i$. This approach is analogous to the nested design for multiple-family analysis (Weller *et al.* 1990; Knott *et al.* 1996); *i.e.* it treats QTL effects as nested within families. Because the first moments are estimated, the method is called the fixed-model strategy. Let n_i be the number of individuals in the *i*-th family and

$$N = \sum_{i=1}^{t} n_i$$

be the overall sample size and define y as an $N \times 1$ vector of the data. The model can be expressed in matrix notation by

$$\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{Z}\alpha + \boldsymbol{W}\delta + \boldsymbol{e} \tag{6}$$

where $\boldsymbol{X}$ is an $N \times (t + 1)$ known design matrix, $\beta = [\mu\ \beta_1, \ldots, \beta_t]^T$ are non-QTL effects, $\boldsymbol{Z}$ is an $N \times t$ design matrix filled by $E(z_{ij}|I_M)$ at the appropriate positions, $\boldsymbol{\alpha} = [\alpha_1 \ldots \alpha_t]^T$ is a vector of gene substitution effects, $\boldsymbol{W}$ is an $N \times t$ design matrix filled by $E(w_{ij}|I_M)$ at the appropriate positions, $\boldsymbol{\delta} = [\delta_1 \ldots \delta_t]^T$ is a vector of dominance deviations and $\boldsymbol{e}$ is an $N \times 1$ vector of residuals. Under a fixed model, the expectation and variance matrix of $\boldsymbol{y}$ conditional on the marker information are $E(\boldsymbol{y}|I_M) = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\alpha} + \boldsymbol{W}\boldsymbol{\delta}$ and $\mathrm{Var}(\boldsymbol{y}|I_M) = \mathrm{Var}(\boldsymbol{e}) = \boldsymbol{R}\sigma_\varepsilon^2$, where $\boldsymbol{R}$ is a diagonal matrix with the element corresponding to y_{ij} being

$$\begin{aligned} R_{ij} = {} & \mathrm{Var}(z_{ij}|I_M)\lambda_{\alpha i} + \mathrm{Var}(w_{ij}|I_M)\lambda_{\delta i} + \\ & 2\mathrm{Cov}(z_{ij}w_{ij}|I_M)\lambda_{\alpha i\delta i} + 1, \end{aligned} \tag{7}$$

where $\lambda_{\alpha i} = \alpha_i^2/\sigma_\varepsilon^2$, $\lambda_{\delta i} = \delta_i^2\ /\sigma_\varepsilon^2$ and $\lambda_{\alpha i\delta i} = \alpha_i\delta_i/\sigma_\varepsilon^2$.

Under the fixed model, parameters are estimated via an iteratively reweighted least-squares algorithm described below. Given an initial guess of $\lambda_{\alpha i}$, $\lambda_{\delta i}$ and $\lambda_{\alpha i\delta i}$, matrix $\boldsymbol{R}$ is considered as known. Under the pretense of known $\boldsymbol{R}$, the solutions of $\theta = \{\beta, \alpha, \delta\}$ and σ_ε^2 are obtained via the following equations:

$$\begin{aligned} \begin{bmatrix} \hat\beta \\ \hat\alpha \\ \hat\delta \end{bmatrix} &= \begin{bmatrix} \boldsymbol{X}^T\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{X}^T\boldsymbol{R}^{-1}\boldsymbol{Z} & \boldsymbol{X}^T\boldsymbol{R}^{-1}\boldsymbol{W} \\ \boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{Z} & \boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{W} \\ \boldsymbol{W}^T\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{W}^T\boldsymbol{R}^{-1}\boldsymbol{Z} & \boldsymbol{W}^T\boldsymbol{R}^{-1}\boldsymbol{W} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{X}^T\boldsymbol{R}^{-1}\boldsymbol{y} \\ \boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{y} \\ \boldsymbol{W}^T\boldsymbol{R}^{-1}\boldsymbol{y} \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{C}_\beta & \boldsymbol{C}_{\beta\alpha} & \boldsymbol{C}_{\beta\delta} \\ \boldsymbol{C}_{\alpha\beta} & \boldsymbol{C}_\alpha & \boldsymbol{C}_{\alpha\delta} \\ \boldsymbol{C}_{\delta\beta} & \boldsymbol{C}_{\delta\alpha} & \boldsymbol{C}_\delta \end{bmatrix} \begin{bmatrix} \boldsymbol{X}^T\boldsymbol{R}^{-1}\boldsymbol{y} \\ \boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{y} \\ \boldsymbol{W}^T\boldsymbol{R}^{-1}\boldsymbol{y} \end{bmatrix} \end{aligned} \tag{8}$$

and

$$\hat\sigma_\varepsilon^2 = \frac{1}{N}(\boldsymbol{y} - \boldsymbol{X}\hat\beta - \boldsymbol{Z}\hat\alpha - \boldsymbol{W}\hat\delta)^T\boldsymbol{R}^{-1}(\boldsymbol{y} - \boldsymbol{X}\hat\beta - \boldsymbol{Z}\hat\alpha - \boldsymbol{W}\hat\delta). \tag{9}$$

Note that these solutions are maximum likelihood estimations (MLEs) under Model 6. If N in the denominator of Equation 9 had been replaced by $N - 3t$, as done in regression analysis, the solutions would no longer be MLEs. Whether N or $N - 3t$ is used to estimate σ_ε^2 does not affect the test statistic. Because $\boldsymbol{R}$ depends on the unknown parameters, it must be updated by the estimates of α, δ and σ_ε^2, and the estimation is then repeated until convergence.

The least-squares method of Knott *et al.* (1996) simply ignores the correction for the residual variance, *i.e.*, assuming $\boldsymbol{R} = \boldsymbol{I}$. When densed markers are used or the QTL effect is small or both, this assumption will have little effect on the results (Xu 1995).

Compared with the EM algorithm under a mixture model, this algorithm is extremely fast—only two to three cycles of iteration are required. However, three additional parameters are added to the model for each additional family, so the number of parameters to estimate grows quickly as the number of families increases.

Under the null hypothesis, H_0: $\boldsymbol{\alpha} = \boldsymbol{\delta} = \boldsymbol{0}$, the maximum likelihood is

$$L_0 = (\hat\sigma_\varepsilon^2)^{-N/2}|\boldsymbol{R}|^{-1/2}\mathrm{Exp}\left\{-\frac{1}{2\hat\sigma_\varepsilon^2}(\boldsymbol{y} - \boldsymbol{X}\hat\beta)^T\boldsymbol{R}^{-1}(\boldsymbol{y} - \boldsymbol{X}\hat\beta)\right\}.$$

Under the alternative hypothesis, H_1: at least one of α_i or δ_i is not zero $\forall$ *i*, the maximum likelihood is

$$\begin{aligned} L_1 = {} & (\hat\sigma_\varepsilon^2)^{-N/2}|\boldsymbol{R}|^{-1/2} \\ & \mathrm{Exp}\left\{-\frac{1}{2\hat\sigma_\varepsilon^2}(\boldsymbol{y} - \boldsymbol{X}\hat\beta - \boldsymbol{Z}\hat\alpha - \boldsymbol{W}\hat\delta)^T\ \boldsymbol{R}^{-1}(\boldsymbol{y} - \boldsymbol{X}\hat\beta - \boldsymbol{Z}\hat\alpha - \boldsymbol{W}\hat\delta)\right\}. \end{aligned}$$

The test statistic is taken as

$$\Lambda = -2[\ln(L_0) - \ln(L_1)]. \qquad (10)$$

In QTL mapping with multiple families, effort is directed from a single family into multiple families. As a consequence, one is no longer interested in the α and δ of any particular family, but rather in α and δ from all families. However, $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\delta}}$ are first moment estimations and their magnitudes are only meaningful when reported relative to the size of the residual variance. As a result, they must be converted into variances before being considered for publication. In traditional QTL analysis for a single F_2 family, the variance explained by a QTL is reported with the F_2 family as the reference population. The QTL analysis with multiple families, however, is interested in the variance of QTL allelic effects among different F_2 families. This variance differs from the within family variance by one generation. The additive QTL variance among the sampled families is expressed by

$$\sigma_\alpha^2 = \frac{1}{t-1}\sum_{i=1}^{t}(\alpha_i - \bar{\alpha})^2 \qquad (11)$$

where

$$\bar{\alpha} = \frac{1}{t}\sum_{i=1}^{t}\alpha_i.$$

Similarly, the dominance variance is

$$\sigma_\delta^2 = \frac{1}{t-1}\sum_{i=1}^{t}(\delta_i - \bar{\delta})^2 \qquad (12)$$

where

$$\bar{\delta} = \frac{1}{t}\sum_{i=1}^{t}\delta_{i.}$$

To estimate σ_α^2 from $\hat{\alpha}$, assume that the estimation is unbiased so that $E(\hat{\alpha}) = \alpha$ and denote $\mathrm{Var}(\hat{\alpha})$ by $\boldsymbol{V}_{\hat{\alpha}}$. Rewrite Equation 11 in matrix notation by $\sigma_\alpha^{\mathbf{2}} = \alpha^T\boldsymbol{A}\alpha$ and define the observed variance of the estimated α's among families as $\sigma_{\hat{\alpha}}^2 = \hat{\alpha}^T\boldsymbol{A}\hat{\boldsymbol{\alpha}}$, where

$$A_{ii} = \frac{1}{t} \text{ and } A_{ij} = \frac{-1}{(t-1)t}.$$

It is known (Seber 1977) that $E(\hat{\alpha}^T\boldsymbol{A}\hat{\boldsymbol{\alpha}}) = \mathrm{Tr}(\boldsymbol{A}\boldsymbol{V}_{\hat{\alpha}}) + \alpha^T\boldsymbol{A}\alpha$, where Tr() represents the trace matrix operation (the sum of all diagonal elements). Therefore, an unbiased estimator of $\sigma_\alpha^{\mathbf{2}} = \alpha^TA\alpha$ is

$$\hat{\sigma}_\alpha^2 = \hat{\alpha}^TA\hat{\alpha} - \mathrm{Tr}(AV_{\hat{\alpha}}). \qquad (13)$$

The estimation is only asymptotically unbiased because $E(\hat{\alpha}) = \alpha$ is true only asymptotically. The variance covariance matrix of the estimated parameters are obtained by $\boldsymbol{V}_{\hat{\alpha}} = \boldsymbol{C}_{\boldsymbol{\alpha}}\hat{\boldsymbol{\sigma}}_\varepsilon^2$, where $\boldsymbol{C}_{\boldsymbol{\alpha}}$ is a submatrix in the coefficient matrix given in Equation 8.

An asymptotically unbiased estimate of σ_δ^2 is analogously derived:

$$\hat{\sigma}_\delta^2 = \hat{\delta}^TA\hat{\delta} - \mathrm{Tr}(AV_{\hat{\delta}}) \qquad (14)$$

where $\boldsymbol{V}_{\hat{\sigma}} = \boldsymbol{C}_{\boldsymbol{\delta}}\hat{\boldsymbol{\sigma}}_\varepsilon^2$.

Two special situations may be noted. When only a single family is analyzed (the traditional QTL mapping strategy), $\boldsymbol{A}$ is only a scalar of value 1, which leads to $\sigma_\alpha^2 = \alpha^2$ and $\sigma_\delta^2 = \delta^2$. Alternatively, when there are an infinite number of families, $\boldsymbol{A}$ approaches

$$\frac{1}{t}\boldsymbol{I}, \text{ leading to } \sigma_\alpha^2 = \frac{1}{t}\sum_{i=1}^{t}\alpha_i^2 \text{ and } \sigma_\delta^2 = \frac{1}{t}\sum_{i=1}^{t}\delta_i^2.$$

Random-model strategy: The second strategy of QTL mapping is to directly test and estimate the variances of the QTL effects, and because of this it is called the random model approach. Consider each α_i and δ_i as randomly sampled from a large hypothetical population with a means of zero and variances of σ_α^2 and σ_δ^2, respectively. Under the null hypothesis that there is no QTL segregating, $\sigma_\alpha^2 = \sigma_\delta^2 = 0$. The model stays the same as (6), although now with different expectation and variance matrices. Under the random model, $E(\boldsymbol{y}\,|I_M) = \boldsymbol{X\beta}$ and $\mathrm{Var}(\boldsymbol{y}\,|\boldsymbol{I}_M) = \boldsymbol{V} = \boldsymbol{ZZ}^T\sigma_\alpha^2 + \boldsymbol{WW}^T\sigma_\delta^2 + \boldsymbol{R}\sigma_\varepsilon^2$ where $\boldsymbol{R}$ is a diagonal matrix with the element corresponding to y_{ij} equal to

$$R_{ij} = \mathrm{Var}(z_{ij}|I_M)\lambda_\alpha + \mathrm{Var}(w_{ij}|I_M)\lambda_\delta + 1 \qquad (15)$$

where $\lambda_\alpha = \sigma_\alpha^2/\sigma_\varepsilon^2$ and $\lambda_\delta = \sigma_\delta^2/\sigma_\varepsilon^2$.

Derivation of (15) is based on the assumption that α and δ are uncorrelated. If the indicator variables, $\boldsymbol{Z}$ and $\boldsymbol{W}$, were observed, then the variance of y would be $\mathrm{Var}(\boldsymbol{y}\,|\boldsymbol{ZW}) = \boldsymbol{V} = \boldsymbol{ZZ}^T\sigma_\alpha^2 + \boldsymbol{WW}^T\sigma_\delta^2 + I\sigma_\varepsilon^2$. When $\boldsymbol{Z}$ and $\boldsymbol{W}$ are replaced by their conditional expectations given marker information, this variance matrix becomes

$$\begin{aligned}
\mathrm{Var}(\boldsymbol{y}|I_M) &= E[\mathrm{Var}(\boldsymbol{y}\backslash\boldsymbol{ZW})] \\
&= E(\boldsymbol{ZZ}^T)\sigma_\alpha^2 + E(\boldsymbol{WW}^T)\sigma_\delta^2 + \boldsymbol{I}\sigma_\varepsilon^2 \\
&= [E(\boldsymbol{ZZ}^T) - E(\boldsymbol{Z})E(\boldsymbol{Z}^T) + E(\boldsymbol{Z})E(\boldsymbol{Z}^T)]\sigma_\alpha^2 + \\
&\quad [E(\boldsymbol{WW}^T) - E(\boldsymbol{W})E(\boldsymbol{W}^T) + E(\boldsymbol{W})E(\boldsymbol{W}^T)]\sigma_\delta^2 + \\
&\quad \boldsymbol{I}\sigma_\varepsilon^2 \\
&= [\mathrm{Var}(\boldsymbol{Z}) + E(\boldsymbol{Z})E(\boldsymbol{Z}^T)]\sigma_\alpha^2 + [\mathrm{Var}(\boldsymbol{W}) + \\
&\quad E(\boldsymbol{W})E(\boldsymbol{W}^T)]\,\sigma_\delta^2 + \boldsymbol{I}\,\sigma_\varepsilon^2 \\
&= E(\boldsymbol{Z})E(\boldsymbol{Z}^T)\sigma_\alpha^2 + E(\boldsymbol{W})E(\boldsymbol{W}^T)\sigma_\delta^2 + [\mathrm{Var}(\boldsymbol{Z})\sigma_\alpha^2 + \\
&\quad \mathrm{Var}(\boldsymbol{W})\sigma_\delta^2 + \boldsymbol{I}\sigma_\varepsilon^2] \\
&= E(\boldsymbol{Z})E(\boldsymbol{Z}^T)\sigma_\alpha^2 + E(\boldsymbol{W})E(\boldsymbol{W}^T)\sigma_\delta^2 + [\mathrm{Var}(\boldsymbol{Z})\lambda_\alpha + \\
&\quad \mathrm{Var}(\boldsymbol{W})\lambda_\delta + \boldsymbol{I}]\sigma_\varepsilon^2 \\
&= E(\boldsymbol{Z})E(\boldsymbol{Z}^T)\sigma_\alpha^2 + E(\boldsymbol{W})E(\boldsymbol{W}^T)\sigma_\delta^2 + \boldsymbol{R}\sigma_\varepsilon^2
\end{aligned}$$

where $\boldsymbol{R} = \mathrm{Var}(\boldsymbol{Z}\lambda_\alpha + \mathrm{Var}(\boldsymbol{W})\lambda_\delta + \boldsymbol{I}$. Recall that $E(\mathbf{Z})$ is in fact $E(\boldsymbol{Z}|I_M)$ and has been denoted by $\boldsymbol{Z}$ for notational convenience.

Note that the definitions of λ_α and λ_δ are different from those of the fixed model. It should be noticed that the family-specific effects, βs, have been treated as fixed effects, although they can be considered as random effects with a mean of zero and a common variance σ_β^2.

Given the expectation and the variance of the model and under the pretense of a normal distribution of $\boldsymbol{y}$, we have the following likelihood function:

$$L(\theta|\boldsymbol{y}) = |\boldsymbol{V}|^{-1/2}\,\mathrm{Exp}\left\{\left(-\frac{1}{2}\right)(\boldsymbol{y} - \boldsymbol{X}\beta)^T\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X}\beta)\right\}. \qquad (16)$$

The MLE of $\theta = [\beta^T\ \sigma_\alpha^2\sigma_\delta^2\ \sigma_\varepsilon^2]^T$ is solved using any convenient numerical algorithm. The log likelihood ratio is used as the test statistic.

The random-model strategy involves inverting and determinating $\boldsymbol{V}$, an $N \times N$ block diagonal matrix, which can be time consuming for large blocks (each block is of $n \times n$ dimension). A simple algorithm developed in random mating designs (S. Xu , unpublished data) can be adopted here. The algorithm provides the following matrix equivalencies:

$$\boldsymbol{V}^{-1} = \sigma_{\varepsilon}^{-2}[\boldsymbol{H} - \boldsymbol{HZ}(\boldsymbol{B}^{-1})\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{H}\lambda_{\alpha}] \qquad (17)$$

and

$$|\boldsymbol{V}| = (\sigma_{\varepsilon}^{2})^{N}\frac{|\boldsymbol{R}|}{|\boldsymbol{B}^{-1}||\boldsymbol{U}^{-1}|} \qquad (18)$$

where $\boldsymbol{U} = \boldsymbol{W}^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{W}\lambda_{\delta} + \boldsymbol{I}$, $\boldsymbol{H} = \boldsymbol{R}^{-1} - \boldsymbol{R}^{-1}\boldsymbol{WU}^{-1}\boldsymbol{W}^{\mathrm{T}}\boldsymbol{R}^{-1}\lambda_{\delta}$ and $\boldsymbol{B} = \boldsymbol{Z}^{\mathrm{T}}\boldsymbol{HZ}\lambda_{\alpha} + \boldsymbol{I}$. Matrices $\boldsymbol{R}$ and $\boldsymbol{U}$ are diagonal while matrix $\boldsymbol{B}$ can be expanded as

$$\begin{aligned}\boldsymbol{B} &= \boldsymbol{Z}^{\mathrm{T}}\boldsymbol{HZ}\lambda_{\alpha} + \boldsymbol{I}\\ &= \boldsymbol{Z}^{\mathrm{T}}(\boldsymbol{R}^{-1} - \boldsymbol{R}^{-1}\boldsymbol{WU}^{-1}\boldsymbol{W}^{\mathrm{T}}\boldsymbol{R}^{-1}\lambda_{\delta})\boldsymbol{Z}\lambda_{\alpha} + \boldsymbol{I}\\ &= \boldsymbol{Z}^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{Z}\lambda_{\alpha} - \boldsymbol{Z}^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{WU}^{-1}\boldsymbol{W}^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{Z}\lambda_{\delta}\lambda_{\alpha} + \boldsymbol{I}.\end{aligned}$$

Since each of $\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{Z}$, $\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{W}$ and $\boldsymbol{U}^{-1}$ is diagonal, $\boldsymbol{B}$ must also be diagonal. Solving for the inverses and determinants of diagonal matrices is trivial.

NUMERICAL COMPARISON

Design of simulations: In this section, the two statistical methods are verified and compared numerically via Monte Carlo simulations. The criteria of verification are standard errors of the parameter estimation and the statistical powers. Factors considered include (1) marker heterozygosity; (2) relative position of QTL; (3) mode of QTL inheritance; (4) QTL variances; (5) distribution of the QTL allelic effect and (6) sampling strategy (family number *vs.* family size). Only a single chromosome segment of length 100 cM covered by 11 evenly spaced codominant markers is simulated. The total number of individuals [N = family number (t) × family size (n)] is set at ≈500 in all simulations. Under each condition, the simulation is repeated for 100 times. The standard deviation of an estimated parameter among the 100 replicates provides a measure of the standard error of parameter estimation. The statistical power is determined by counting the number of runs (over the 100 replicates) that have test statistics greater than an empirical threshold. The empirical threshold value under each condition is obtained by choosing the 95th percentile of the highest test statistic over 1000 additional runs under the null model (no QTL is segregating).

Marker heterozygosity in the population in which the inbred lines are sampled is simulated at three levels: (1) two alleles, (2) four alleles and (3) eight alleles. All alleles are equally frequent so that the marker heterozygosities represented by the three situations are one half, three quarters and seven eighths, respectively.

A single QTL is located at one of the three possible positions (measured from the left end of the chromosome): 0 cM (overlapping with the first marker), 25 cM (between markers 3 and 4) and 50 cM (in the middle of the chromosome). The estimated QTL location takes the point of the chromosome segment that has the highest test statistic value.

The mode of QTL inheritance is determined by the ratio of σ_{α}^{2} to σ_{δ}^{2}: additive mode (σ_{α}^{2}:σ_{δ}^{2} = 1:0); mixed mode (σ_{α}^{2}:σ_{δ}^{2} = 1:1); dominance mode (σ_{α}^{2}:σ_{δ}^{2} = 0:1).

The variance explained by the QTL is $\sigma_{q}^{2} = \sigma_{\alpha}^{2} + \sigma_{\delta}^{2}$, which is simulated at three levels: (1) $\sigma_{q}^{2} = 0.11$, corresponding to $h_{q}^{2} = \sigma_{q}^{2}/(\sigma_{q}^{2} + \sigma_{\varepsilon}^{2}) = 0.10$; (2) $\sigma_{q}^{2} = 0.25$, corresponding to $h_{q}^{2} = 0.20$; and (3) $\sigma_{q}^{2} = 0.43$, corresponding to $h_{q}^{2} = 0.30$. In all simulations, the parametric value of σ_{ε}^{2} is set at 1.

Three distributions of the allelic effect of the QTL are considered. The first is uniform distribution with 10 equally frequent alleles. Each allele is assigned a value between 0 and 9. The F_1 hybrid of each family is generated by randomly sampling two from the 10 alleles with replacement. F_2 individuals are then generated by selfing the F_1 hybrid. The additive value of an F_2 individual is the sum of effects of the two alleles. The dominance effect takes the product of the two parental alleles. These genetic values (additive and dominance) are finally rescaled so that they have a mean of zero and the assigned variances. The second is normal distribution with infinite number of alleles. An F_1 hybrid is made of two random alleles, each being assigned a value sampled from $N(0,1)$ distribution. The dominance effect between any two sampled alleles takes the product of the two allelic effects. When F_2 individuals are generated, their genetic values at the QTL are rescaled so that they have the appropriate assigned variances. The third distribution is 10 alleles, each having a value between 0 and 9. The frequency of an allele, however, scales exponentially with its assigned effect. Let p_j be the frequency of the j-th allele for $j = 0, \ldots, 9$, then

$$p_j = c^{j}/\sum_{k=0}^{9} c^{k}$$

where $c = 0.5$. Again, the genetic values of an F_2 individual are rescaled. Note that the first distribution is a special case of the third distribution with $c = 1$.

The last but most important factor considered in the simulations is the sampling strategy: family number *vs.* family size ($N = t \times n = 500$). Eight levels are considered: (1) $t \times n = 1 \times 500$; (2) $t \times n = 3 \times 167$; (3) $t \times n = 6 \times 83$; (4) $t \times n = 10 \times 50$; (5) $t \times n = 15 \times 33$; (6) $t \times n = 20 \times 25$; (7) $t \times n = 50 \times 10$; and (8) $t \times n = 100 \times 5$.

Instead of performing simulations under all possible cases, I simulated a situation in which the central level is chosen for each factor considered. This particular situation is then referred to as the "standard," which is described as follows: (1) four equally frequent alleles for each marker locus; (2) the QTL located at 25 cM; (3) mixed mode of QTL inheritance, *i.e.*, $\sigma_{\alpha}^{2} = \sigma_{\delta}^{2} = 0.125$; (4) the total QTL variance of $\sigma_{q}^{2} = \sigma_{\alpha}^{2} + \sigma_{\delta}^{2} = 0.125 + 0.125 = 0.25$, corresponding to $h_{q}^{2} = 0.20$; (5)

TABLE 1

Empirical threshold values for significance test at α = 0.05, where α is the type I error rate

Method		Fixed model	Random model
Marker alleles	Two	38.47	14.68
	Four	41.01	15.05
	Eight	41.68	16.28
Sampling strategy	1 × 500[a]	9.347	11.21
	3 × 167	18.65	15.47
	6 × 83	27.53	14.19
	10 × 50	41.01	15.05
	15 × 33	56.90	16.47
	20 × 25	71.94	17.74
	50 × 10	—[b]	21.02
	100 × 5	—	24.62

[a] Number of families × number of individuals per family.

[b] Simulations are not conducted in these two cases.

TABLE 2

Estimates of QTL parameters and empirical powers ($\alpha = 0.05$) under different levels of marker polymorphism

Method	Marker alleles	cM_A	σ^2_α	σ^2_δ	σ^2_ε	Power (%)
Fixed model	Two	27.84 (10.32)	0.130 (0.099)	0.071 (0.202)	0.918 (0.061)	89
	Four	24.99 (8.65)	0.130 (0.074)	0.142 (0.144)	0.931 (0.057)	94
	Eight	25.76 (9.08)	0.123 (0.078)	0.143 (0.147)	0.922 (0.064)	93
Random model	Two	27.08 (12.76)	0.128 (0.091)	0.104 (0.088)	0.979 (0.066)	89
	Four	25.26 (8.46)	0.123 (0.067)	0.144 (0.154)	0.977 (0.063)	94
	Eight	25.64 (8.88)	0.120 (0.075)	0.150 (0.144)	0.963 (0.068)	94

Standard errors of the estimates, given in parentheses, are calculated by the standard deviations among 100 replicated simulations.

cM_A, estimated QTL position in cM; σ^2_α, estimated additive variance of the QTL, the true σ^2_α being 0.125; σ^2_δ, estimated dominance variance of the QTL, the true σ^2_δ being 0.125; σ^2_ε, estimated residual variance, the true σ^2_ε being 1.0.

the QTL allelic effect normally distributed; and (6) 10 families, each having 50 individuals, *i.e.*, $t \times n = 10 \times 50$. When the influence of different levels of a factor on the performances of the two statistical methods are examined, all other factors are set to the standard levels.

Results of simulations: The empirical threshold values at a type I error rate of 0.05 are given in Table 1. The number of alleles per marker locus does not seem to have an influence on the threshold values. As the number of families increases, the threshold value increases under the fixed-model strategy considerably more than it does under the random-model strategy. This is expected because increasing the number of families increases the number of parameters tested under the fixed model while the number of parameters tested does not change under the random-model strategy.

When each marker has two equally frequent alleles in the population in which the parental lines are sampled, the two models have similar estimation errors and statistical powers (Table 2). The estimation of the QTL position, however, is biased and with large error in both methods. The statistical powers are also low, with two marker alleles relative to more marker alleles. The fixed-model strategy generally provides a biased estimate for the residual variance, as shown in this and subsequent tables.

The proportion of the phenotypic variance explained by the QTL (h^2_q) does not have an impact on the comparison of the two methods (Table 3). Both methods produce biased estimates of the QTL position and low statistical powers when h^2_q is low.

Table 4 shows that when the QTL is located at one end of the chromosome segment, estimation of the QTL position is biased toward the center and also with large error in both methods. There is little change in the power to detect a QTL as the true QTL position varies.

Mixed mode of QTL inheritance (additive and dominance) seems to have a higher statistical power than either of the additive or the dominance mode of inheritance. The estimation of the QTL position is biased and with large error under the dominance mode of inheritance. Again, the two methods do not show any major difference (see Table 5).

Distribution of the QTL allelic effect does not affect the comparison of the two methods (Table 6). It does, however, have an effect on the statistical power and the estimation errors of QTL parameters. The uniform distribution produces results similar to (in fact, slightly better than) the normal distribution. The exponential distribution decreases the statistical power and increases errors of parameter estimation.

Finally, the sampling strategy has a major impact on the performance (Table 7). First, there seems to be an optimal sampling strategy (10 × 50) that leads to the highest statistical power and smallest estimation errors of QTL parameters. Second, the sampling strategy of a single family causes a severe loss in power and huge biases and errors of QTL parameter estimation. Third, the residual variance is underestimated as

TABLE 3

Estimates of QTL parameters and empirical powers ($\alpha = 0.05$) under different levels of heritability of the QTL

Method	h_q2	cM_A	σ^2_α	σ^2_δ	σ^2_ε	Power (%)
Fixed model	0.10	28.06 (15.23)	0.066 (0.061)	0.060 (0.049)	0.925 (0.057)	73
	0.20	24.99 (4.65)	0.130 (0.074)	0.142 (0.144)	0.931 (0.057)	97
	0.30	25.52 (5.11)	0.220 (0.128)	0.199 (0.224)	0.925 (0.060)	98
Random model	0.10	28.30 (15.91)	0.064 (0.057)	0.059 (0.057)	0.971 (0.061)	73
	0.20	25.26 (4.46)	0.123 (0.067)	0.144 (0.154)	0.977 (0.063)	96
	0.30	25.52 (5.23)	0.231 (0.132)	0.267 (0.614)	0.969 (0.067)	100

Standard errors of the estimates, given in parentheses, are calculated by the standard deviations among 100 replicated simulations. h^2_q, proportion of total phenotypic variance explained by the QTL.

TABLE 4

Estimates of QTL parameters and empirical powers ($\alpha = 0.05$) under three different locations of the QTL

Method	cM_T	cM_A	σ^2_α	σ^2_δ	σ^2_ε	Power (%)
Fixed model	0	4.93 (15.28)	0.125 (0.080)	0.137 (0.115)	0.917 (0.063)	94
	25	24.99 (4.65)	0.130 (0.074)	0.142 (0.144)	0.931 (0.057)	97
	53	52.38 (5.57)	0.138 (0.080)	0.128 (0.104)	0.933 (0.066)	93
Random model	0	3.09 (9.83)	0.124 (0.074)	0.125 (0.096)	0.967 (0.067)	95
	25	25.26 (4.46)	0.123 (0.067)	0.144 (0.154)	0.977 (0.063)	96
	53	52.15 (5.25)	0.133 (0.079)	0.128 (0.104)	0.979 (0.072)	95

Standard errors of the estimates, given in parentheses, are calculated by the standard deviations among 100 replicated simulations. cM_T, true position of the QTL measured in centimorgans from the left end of the chromosome.

the number of families increases. This is especially so for the fixed model. Overall, the two strategies of QTL mapping perform equally well, except that the fixed-model approach is difficult to implement for large number of families.

DISCUSSION

Unless it is known that the parents are heterozygous at most QTLs for a trait of interest, it is generally recommended to use at least a few independent families for QTL analysis. Using more than a single family for QTL mapping may reduce a type II error caused by homogeneous parents being sampled. In traditional QTL mapping using a single-line cross, little attention has been paid to the type II error of this kind. This is because the two parental lines involved are not randomly selected from a pool of available strains; instead, they are selected to be at the opposite extremes for the trait of interest. As a consequence, it is almost guaranteed that most QTLs are heterozygous in the F_1 parents, and thus a type II error of this kind is likely avoided. A nonrandom selection of parental lines can increase the statistical power for detecting QTLs responsible for the trait used as the selection criterion, but it may not be helpful in detecting QTLs responsible for other traits. In addition, one must be careful about the statistical inference space of the parameter estimation: because of the nonrandom selection, estimation of the QTL effect is biased and can only be inferred upon the two parental lines, not the pool of available strains where the two lines were selected.

Although the two strategies of consensus QTL mapping appear to perform equally well, the fixed-model approach is generally less preferable for the following reasons. With multiple-family QTL mapping, one is no longer interested in the effect of gene substitution in any particular family, but rather is interested in the variance of the substitution effect among different families. In other words, the average effect of gene substitution is considered to be a random variable with variance σ^2_α. Rather than estimating and testing σ^2_α, the fixed-model approach estimates and tests each observation of the random variable. It is conceptually incorrect to estimate and test values of a random variable. Further-

TABLE 5

Estimates of QTL parameters and empirical powers ($\alpha = 0.05$) under different modes of inheritance of the QTL

Method	Mode of inheritance	cM_A	σ^2_α	σ^2_δ	σ^2_ε	Power (%)
Fixed model	A	25.27 (8.02)	0.242 (0.135)	0.007 (0.017)	0.930 (0.064)	91
	A + D	24.99 (4.65)	0.130 (0.074)	0.142 (0.144)	0.931 (0.057)	97
	D	27.20 (12.49)	0.009 (0.028)	0.230 (0.208)	0.922 (0.060)	92
Random model	A	25.23 (5.96)	0.250 (0.135)	0.007 (0.010)	0.973 (0.066)	92
	A + D	25.26 (4.46)	0.123 (0.067)	0.144 (0.154)	0.977 (0.063)	96
	D	27.49 (13.35)	0.012 (0.018)	0.250 (0.313)	0.965 (0.063)	93

Standard errors of the estimates, given in parentheses, are calculated by the standard deviations among 100 replicated simulations.

Abbreviations: A, additive; D, dominance; A + D, both.

TABLE 6

Estimates of QTL parameters and empirical powers ($\alpha = 0.05$) under different allelic distributions of the QTL

Method	Distribution[a]	cM_A	σ^2_α	σ^2_δ	σ^2_ε	Power (%)
Fixed model	Uniform	24.87 (3.66)	0.125 (0.075)	0.131 (0.084)	0.946 (0.065)	98
	Normal	24.99 (4.65)	0.130 (0.074)	0.142 (0.144)	0.931 (0.057)	97
	Exponential	26.86 (9.68)	0.117 (0.088)	0.154 (0.255)	0.926 (0.064)	89
Random model	Uniform	24.88 (3.78)	0.119 (0.066)	0.131 (0.084)	0.992 (0.067)	99
	Normal	25.26 (4.46)	0.123 (0.067)	0.144 (0.154)	0.977 (0.063)	96
	Exponential	26.18 (9.93)	0.115 (0.081)	0.163 (0.255)	0.969 (0.070)	88

Standard errors of the estimates, given in parentheses, are calculated by the standard deviations among 100 replicated simulations.

[a] Distribution of the QTL allelic effect.

more, the fixed-model approach involves two steps: (1) estimating the effects and (2) converting the effects into a variance. Because of this, the fixed-model approach is computationally inferior to the random-model approach when the number of families is large.

The random-model approach to QTL mapping was originally developed in human genetic linkage analysis in which a large number of small families are often involved (Haseman and Elston 1972; Goldgar 1990; Schork 1993; Olson and Wijsman 1993; Fulker and Cardon 1994; Kruglyak and Lander 1995; Xu and Atchley 1995). Because linkage phases of markers in the parents are generally not known in small pedigrees, the random-model approach is often implemented through an identical-by-descent (IBD) based variance component analysis. The IBD-based method does not depend on information about linkage phases of the parents; rather, it utilizes information on the number of alleles IBD shared by two siblings. The random-model approach proposed in this paper is closely related to the IBD-based method. Recall that the variance–covariance matrix of the data is Var($\boldsymbol{y}|I_M$) = $\boldsymbol{V}$ = $\boldsymbol{ZZ}^T\sigma^2_\alpha + \boldsymbol{WW}^T\sigma^2_\delta + \boldsymbol{R}\sigma^2_\varepsilon$, which can be reformulated as $\boldsymbol{V} = (\boldsymbol{ZZ}^T + \boldsymbol{D}_\alpha)\sigma^2_\alpha + (\boldsymbol{WW}^T + \boldsymbol{D}_\delta)\sigma^2_\delta + \boldsymbol{I}\sigma^2_\varepsilon = \Pi\,\sigma^2_\alpha + \Delta\,\sigma^2_\delta + \boldsymbol{I}\sigma^2_\varepsilon$, where $\boldsymbol{D}_\alpha = \mathrm{diag}\{\mathrm{Var}(z_{ij}|\boldsymbol{I}_M)\}$, $\boldsymbol{D}_\delta = \mathrm{diag}\{\mathrm{Var}(w_{ij}|\boldsymbol{I}_M)\}$, $\Pi = \boldsymbol{ZZ}^T + \boldsymbol{D}_\alpha$ and $\Delta = \boldsymbol{WW}^T + \boldsymbol{D}_\delta$. Matrices Π and Δ have been referred to as the IBD and double IBD matrices, respectively, by Xu (1996b). Here, one is able to partition the IBD matrix Π into two components, $\boldsymbol{ZZ}^T$ and $\boldsymbol{D}_\alpha$, because one knows the linkage phases of the markers in the parents. Decomposition of the IBD matrices allows one to apply the special algorithms of matrix inversion (Equation 17) and determinant calculation (Equation 18). As a consequence, this

TABLE 7

Estimates of QTL parameters and empirical powers ($\alpha = 0.05$) under different sampling strategies (number of families × number of individuals per family)

Method	Sampling strategy	cM_A	σ^2_α	σ^2_δ	σ^2_ε	Power (%)
Fixed model	1 × 500	31.87 (20.04)	0.141 (0.240)	0.189 (0.489)	1.221 (2.415)	77
	3 × 167	26.24 (9.85)	0.143 (0.151)	0.141 (0.329)	0.977 (0.060)	92
	6 × 83	25.87 (10.23)	0.123 (0.092)	0.121 (0.126)	0.976 (0.064)	91
	10 × 50	24.99 (4.65)	0.130 (0.074)	0.142 (0.144)	0.931 (0.057)	97
	15 × 33	25.16 (9.51)	0.130 (0.074)	0.125 (0.074)	0.899 (0.059)	94
	20 × 25	26.82 (12.58)	0.133 (0.069)	0.133 (0.085)	0.851 (0.058)	94
Random model	1 × 500	30.42 (19.83)	0.199 (0.456)	0.208 (0.518)	0.973 (0.084)	76
	3 × 167	25.90 (10.51)	0.159 (0.170)	0.166 (0.508)	0.989 (0.063)	91
	6 × 83	27.72 (13.54)	0.133 (0.095)	0.152 (0.228)	0.972 (0.074)	95
	10 × 50	25.26 (4.46)	0.123 (0.067)	0.144 (0.154)	0.977 (0.063)	96
	15 × 33	26.59 (6.48)	0.138 (0.083)	0.141 (0.132)	0.955 (0.078)	96
	20 × 25	26.36 (11.39)	0.125 (0.067)	0.134 (0.100)	0.945 (0.062)	92
	50 × 10	27.20 (10.41)	0.124 (0.073)	0.136 (0.072)	0.889 (0.081)	85
	100 × 5	35.10 (21.00)	0.136 (0.113)	0.135 (0.074)	0.770 (0.084)	61

Standard errors of the estimates, given in parentheses, are calculated by the standard deviations among 100 replicated simulations.

implementation of the random-model approach is computationally much faster than the fixed-model approach, especially when the number of families is large.

In the random-model strategy, the family-specific effects, β, have been treated as fixed effects. When the number of families is large, however, it is desirable to treat β as random effects. By doing so, one only estimates a single parameter, σ^2_β, instead of a large array of parameters. Assume that β are random effects so that the expectation and variance matrices of the data are $E(\boldsymbol{y}|I_M) = 1\mu$ and $\mathrm{Var}(\boldsymbol{y}|I_M) = \boldsymbol{V} = \boldsymbol{XX}^T\sigma^2_\beta + \boldsymbol{ZZ}^T\sigma^2_\alpha + \boldsymbol{WW}^T\sigma^2_\delta + \boldsymbol{R}\sigma^2_\varepsilon$, respectively. The variance of family-specific effects, σ^2_β, is contributed by both genetic and nongenetic factors. Genetic factors include polygenic effects and heritable maternal or paternal effects. Nongenetic factors include common environmental effects shared by members of the same families. Note that β or σ^2_β are nuisances because they are not QTL parameters. Therefore, they can be removed from the model using the restricted maximum likelihood method (Patterson and Thompson 1971). Such a treatment will significantly reduce the large bias observed in the estimate of the residual variance (see the last two rows of Table 7).

This paper demonstrates the algorithm of QTL mapping combining multiple F_2 families as an example. With the random-model approach, it is easy to extend the algorithm to combine all types of line cross data, *e.g.*, backcrosses, double haploids, open pollinated progenies. It is also not difficult to combine data from multiple full-sib and half-sib families. The method provides a general tool for data updating; *i.e.*, QTL linkage analysis can be constantly updated as new data become available.

I thank Damian Gessler for helpful comments on the manuscript. This research was supported by the National Institutes of Health grant GM55321-01 and the National Research Initiative Competitive grants program / USDA 95-37205-2313.

LITERATURE CITED

Fulker, D. W., and L. R. Cardon, 1994 A sib-pair approach to interval mapping of quantitative trait loci. Am. J. Hum. Genet. **54:** 1092–1103.

Goldgar, D. E., 1990 Multipoint analysis of human quantitative genetic variation. Am. J. Hum. Genet. **47:** 957–967.

Haley, C. S., and S. A. Knott, 1992 A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity **69:** 315–324.

Haseman, J. K., and R. C. Elston, 1972 The investigation of linkage between a quantitative trait and a marker locus. Behav. Genet. **2:** 3–19.

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

Jansen, R. C., 1994 Controlling the type I and type II errors in mapping quantitative trait loci. Genetics **138:** 871–881.

Knott, S. A., J. M. Elsen and C. S. Haley, 1996 Methods for multiple-marker mapping of quantitative trait loci in half-sib populations. Theor. Appl. Genet. **93:** 71–80.

Kruglyak, L., and E. S. Lander, 1995 Complete multipoint sib-pair analysis of qualitative and quantitative traits. Am. J. Hum. Genet. **57:** 439–454.

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

Martínez, O., and R. N. Curnow, 1992 Estimating the locations and the sizes of the effects of quantitative trait loci using flanking markers. Theor. Appl. Genet. **85:** 480–488.

Muranty, H., 1996 Power of tests for quantitative trait loci detection using full-sib families in different schemes. Heredity **76:** 156–165.

Olson, J. M., and E. M. Wijsman, 1993 Linkage between quantitative trait and markers loci: methods using all relative pairs. Genet. Epidemiol. **10:** 87–102.

Patterson, H. D., and R. Thompson, 1971 Recovery of inter-block information when block sizes are unequal. Biometrika **58:** 545–554.

Schork, N. J., 1993 Extended multipoint identitify-by-descent analysis of human quantitative traits: efficiency, power, and modeling considerations. Am. J. Hum. Genet. **53:** 1306–1313.

Seber, G. A. F., 1977 *Linear Regression Analysis*. John Wiley & Sons, New York.

Weller, J. I., Y. Kashi and M. Soller, 1990 Power of daughter and granddaughter designs for determining linkage between marker loci and quantitative trait loci in dairy cattle. J. Dairy Sci. **73:** 2525–2537.

Xu, S., 1995 A comment on the simple regression method for interval mapping. Genetics **141:** 1657–1659.

Xu, S., 1996a Mapping quantitative trait loci using four-way crosses. Genet. Res. Camb. **68:** 175–181.

Xu, S., 1996b Computation of the full likelihood function for estimating variance at a quantitative trait locus. Genetics **144:** 1951–1960.

Xu, S., and W. R. Atchley, 1995 A random model approach to interval mapping of quantitative trait loci. Genetics **141:** 1189–1197.

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

Communicating editor: C.-I Wu