Joint Linkage and Linkage Disequilibrium Mapping of Quantitative Trait Loci in Natural Populations

Rongling Wu,*,[1] Chang-Xing Ma*,† and George Casella*

**Department of Statistics, University of Florida, Gainesville, Florida 32611 and †Department of Statistics, Nankai University, Tianjin 300071, People's Republic of China*

Manuscript received June 15, 2001
Accepted for publication November 20, 2001

ABSTRACT

Linkage analysis and allelic association (also referred to as linkage disequilibrium) studies are two major approaches for mapping genes that control simple or complex traits in plants, animals, and humans. But these two approaches have limited utility when used alone, because they use only part of the information that is available for a mapping population. More recently, a new mapping strategy has been designed to integrate the advantages of linkage analysis and linkage disequilibrium analysis for genome mapping in outcrossing populations. The new strategy makes use of a random sample from a panmictic population and the open-pollinated progeny of the sample. In this article, we extend the new strategy to map quantitative trait loci (QTL), using molecular markers within the EM-implemented maximum-likelihood framework. The most significant advantage of this extension is that both linkage and linkage disequilibrium between a marker and QTL can be estimated simultaneously, thus increasing the efficiency and effectiveness of genome mapping for recalcitrant outcrossing species. Simulation studies are performed to test the statistical properties of the MLEs of genetic and genomic parameters including QTL allele frequency, QTL effects, QTL position, and the linkage disequilibrium of the QTL and a marker. The potential utility of our mapping strategy is discussed.

GENETIC mapping of quantitative trait loci (QTL) has become a routine tool for the genetic study of plants, animals, and humans. With such a tool, many fundamental genetic questions including the inheritance mode of a quantitative trait, genotype × environment interaction, and the genetic basis of heterosis can be addressed (reviewed by Tanksley 1993; Templeton 1999; Wu *et al.* 2000; Mackay 2001). Genetic mapping also has potential to reshape our understanding of complex biological phenomena, such as human diseases and adaptive plasticity (the capacity of a given individual to change its phenotype across different environments). Most of these phenomena are now viewed as having some genetic components and, therefore, can be modified or changed genetically for a feature beneficial to humans. It can be anticipated that genetic mapping will play an increasingly important role in unraveling the genetic basis of quantitative variation in the next decade with the advent of novel DNA-based molecular marker technologies, such as single-nucleotide polymorphisms (SNPs; Wang *et al.* 1998).

Because of differences in biological properties of study materials, considerable effort is being made to develop statistical genetic mapping methods for specific species or populations. In terms of the genetic principles behind mapping, the methodology of genetic mapping includes two main areas: linkage analysis and association studies (reviewed by Olson *et al.* 1999). Linkage analysis is based on the recombination of nonalleles at a marker and QTL during meiosis and, thus, can directly estimate the map distance (measured by recombination fraction) between the two syntenic loci. However, it is difficult to detect recombination events between closely spaced (<1 cM) loci when there are a limited number of meioses occurring in a mapping population (*e.g.*, Hästbacka *et al.* 1992, 1994; Darvasi *et al.* 1993; Long *et al.* 1995). Association studies, on the other hand, use all recombinations generated since nonrandom association of nonalleles at a marker and QTL (commonly referred to as linkage disequilibrium) was introduced into a population, thus increasing the precision of the estimate of the QTL location (Risch and Merikangas 1996; Rabinowitz 1997; Xiong and Guo 1997). Yet, the localization of QTL using linkage disequilibrium mapping is ineffective when the significant linkage disequilibrium detected between a marker and QTL results from the recent occurrence of disequilibrium rather than from a tight linkage between the loci. Such a spurious association detected even when the marker is not physically linked to any causative loci may be due to population subdivision and admixture. Current population- (Gordon *et al.* 2000; Luo *et al.* 2000) or family-based analyses [*e.g.*, the transmission/disequilibrium test (TDT); Spielman and Ewens 1996; Allison 1997] of linkage disequilibrium cannot distinguish strong disequilibrium and

[1]*Corresponding author:* Department of Statistics, 533 McCarty Hall C, University of Florida, Gainesville, FL 32611. E-mail: rwu@stat.ufl.edu

loose linkage from weak disequilibrium and tight linkage (Whittaker *et al.* 2000).

The limits of linkage analysis and linkage disequilibrium mapping when they are used alone can be overcome by a new strategy for taking advantage of each approach in genetic mapping. Such a joint linkage and linkage disequilibrium mapping strategy has been recently devised by Wu and Zeng (2001) in that a random sample from a natural population and the open-pollinated progeny of the sample were analyzed jointly. This strategy was established on the principle of gene transmission from the parental to progeny generation during which the linkage between a marker and QTL is broken down due to meiotic recombination. It can therefore divide the composite measure of linkage disequilibrium from traditional population- or family-based association tests relying on recombinations in a single generation into two components: the linkage between the marker and QTL and their linkage disequilibrium created at a historic time. With the measures of these two components, one can clearly determine the mechanistic basis of a significant disequilibrium detected between a marker and QTL, which increases the feasibility for fine mapping QTL affecting a quantitative trait.

In this article, we extend the joint linkage and linkage disequilibrium mapping strategy to map QTL segregating in a natural population. The extension allows for simultaneous estimates for a number of genetic and genomic parameters including the allele frequency of QTL, its effects, its location, and its population association with a known marker locus. Our analysis is performed within the maximum-likelihood framework, implemented with the expectation-maximization (EM) algorithm. The statistical properties of the estimates for different genetic parameters are studied through extensive simulations. A comparison of the power for detecting linkage disequilibrium is made on the basis of traditional disequilibrium analyses and the joint linkage and linkage disequilibrium analysis proposed here.

STATISTICAL METHOD

Population structure theory: Outcrossing species likely have heterogeneous genomes, on which both dominant and codominant loci are distributed. For codominant loci, there are often a high but variable number of alleles from locus to locus (Weber and Wong 1993; Pfeiffer *et al.* 1997). To simplify the descriptions of our mapping model, we consider only biallelic codominant loci in this article. Although straightforward in principle, extensions to other marker types, such as dominant or missing markers and multiallelic markers, require particular mathematical manipulations.

Consider one marker (**M**) and one QTL (**Q**), both segregating in a random mating population at Hardy-Weinburg equilibrium. The two alleles are denoted by M_1 and M_2 at the marker locus and by Q_1 and Q_2 at the QTL. The frequencies of alleles M_r ($r = 1, 2$) and Q_s ($s = 1, 2$) in the population are denoted by p_r and q_s, with $\Sigma_r^2 p_r = 1$ and $\Sigma_s^2 q_s = 1$. The population frequen- of one-locus genotypes $M_{r_1}M_{r_2}$ ($r_1 \le r_2 = 1, 2$) and $Q_{s_1}Q_{s_2}$ ($s_1 \le s_2 = 1, 2$) are denoted by $P_{r_1r_2}$ and $Q_{s_1s_2}$, with $\Sigma_{r_1}^2\Sigma_{r_2}^2 P_{r_1r_2} = 1$ and $\Sigma_{s_1}^2\Sigma_{s_2}^2 Q_{s_1s_2} = 1$. The nonallelic frequencies from the marker and QTL are not independent of each other in the population, with the coefficient of gametic linkage disequilibrium denoted by D_{rs}.

If the marker and QTL are located on the same region of a chromosome, they are likely linked with recombination fraction θ. On the basis of population genetic theory (Nagylaki 1992), it is easy to derive the population frequencies of four two-locus gametes (haplotypes) M_rQ_s ($r, s = 1, 2$), which are randomly combined to form the current generation t, as

$$p_{rs}^{(t)} = p_r^{(t)}q_s^{(t)} + (-1)^{r+s}D_{rs}^{(t)}, \quad r, s = 1, 2, \qquad (1)$$

where $D_{rs}^{(t)}$ has a bound of $\max[-p_1^{(t)}q_1^{(t)}, -p_2^{(t)}q_2^{(t)}] \le D_{(rs)}^{(t)} \le \min[p_1^{(t)}q_2^{(t)}, p_2^{(t)}q_1^{(t)}]$ (Weir 1996). Through free combinations, these gametes from the maternal and paternal sides produce nine different progeny genotypes $M_{r_1}M_{r_2}Q_{s_1}Q_{s_2}$ ($r_1 \le r_2$, $s_1 \le s_2 = 1, 2$ denote the two alternative alleles of the marker and QTL), whose population frequencies $P_{r_1r_2s_1s_2}^{(t)}$ in the current generation t are calculated as products of the population frequencies of the maternal and paternal gametes (Table 1). Table 1 also gives the (conditional) frequencies of the four gametes produced by each of the nine genotypes for the next (progeny) generation $t + 1$. As shown by population genetic theory, the amount of linkage disequilibrium between any two loci is reduced at the rate of recombination frequency after the population mates at random for one generation (Nagylaki 1992). Therefore, the coefficient of gametic linkage disequilibrium in the progeny generation $t + 1$ is changed to be $D^{(t+1)} = (1 - \theta)D^{(t)}$. Thus, the population frequencies of two-locus gametes M_rQ_s ($r, s = 1, 2$), which are randomly combined to form the progeny generation $t + 1$, are expressed as

$$p_{rs}^{(t+1)} = p_r^{(t)}q_s^{(t)} + (-1)^{r+s}(1 - \theta)D_{rs}^{(t)}, \quad r, s = 1, 2. \qquad (2)$$

For plants, all genetic information about the progeny generation is contained in seeds. If there is no overlapping in reproduction between parental and progeny generations, the frequencies of the genotypes at the marker and QTL are the products of the frequencies of the corresponding gametes.

Sampling theory: Assume that we randomly select M unrelated individuals from the population and collect the seeds from the selected individuals. The seeds are germinated and grown into seedlings for a progeny test, which is a regular procedure for traditional quantitative genetic experiments (McKeand and Bridgwater 1998). Both the selected individuals and their progeny are genotyped for molecular markers. Assuming the species

TABLE 1

Probabilities of the gamete genotypes produced by mother plants with different genotypes for the marker and QTL

Mother		Gamete[b]			
Genotype	Frequency[a]	M_1Q_1	M_1Q_2	M_2Q_1	M_2Q_2
$M_1M_1Q_1Q_1$	$P_{1111}^{(t)} = p_{11}^{2(t)}$	1	0	0	0
$M_1M_1Q_1Q_2$	$P_{1112}^{(t)} = 2p_{11}^{(t)}p_{12}^{(t)}$	$\frac{1}{2}$	$\frac{1}{2}$	0	0
$M_1M_1Q_2Q_2$	$P_{1122}^{(t)} = p_{12}^{2(t)}$	0	1	0	0
$M_1M_2Q_1Q_1$	$P_{1211}^{(t)} = 2p_{11}^{(t)}p_{21}^{(t)}$	$\frac{1}{2}$	0	$\frac{1}{2}$	0
$M_1M_2Q_1Q_2$	$P_{1212}^{(t)} = 2(p_{11}^{(t)}p_{22}^{(t)} + p_{12}^{(t)}p_{21}^{(t)})$	$\frac{1}{2}(1-\theta)$	$\frac{1}{2}\theta$	$\frac{1}{2}\theta$	$\frac{1}{2}(1-\theta)$
$M_1M_2Q_2Q_2$	$P_{1222}^{(t)} = 2p_{12}^{(t)}p_{22}^{(t)}$	0	$\frac{1}{2}$	0	$\frac{1}{2}$
$M_2M_2Q_1Q_1$	$P_{2211}^{(t)} = p_{21}^{2(t)}$	0	0	1	0
$M_2M_2Q_1Q_2$	$P_{2212}^{(t)} = 2p_{21}^{(t)}p_{22}^{(t)}$	0	0	$\frac{1}{2}$	$\frac{1}{2}$
$M_2M_2Q_1Q_2$	$P_{2222}^{(t)} = p_{22}^{2(t)}$	0	0	0	1

[a] $p_{rs}^{(t)}$'s are the population frequencies of marker-QTL gametes, M_rQ_s, which are randomly combined to produce diploid plants in the mother generation t.
[b] θ is the recombination fraction between the marker and QTL.

studied is dioecious, the genotypes of the seeds from a selected individual are derived from its own (maternal) gametes each with a frequency given in Table 1 and the paternal gametes from the pollen pool each with a frequency described by Equation 2. Thus, every selected individual represents an open-pollinated (*i.e.*, half-sib) family with the common mother and different (unknown) fathers. On the basis of the sampling theory, the M selected individuals include three different marker genotypes, with the number denoted by $M_{r_1r_2}$ for genotype $M_{r_1}M_{r_2}$, and the genotypic frequencies in the sampled population are $P_{11}^{(t)} = p_1^{2(t)}$ for M_1M_1, $P_{12}^{(t)} = 2p_1^{(t)}p_2^{(t)}$ for M_1M_2, and $P_{22}^{(t)} = p_2^{2(t)}$ for M_2M_2. The progeny from the selected individuals (called mothers) with different marker genotypes are different in genotype composition and genotype frequency (Table 2). In other words, the marker genotype of an offspring (g_o) is dependent on the marker genotype of its mother (g_m):

$$g_o = \begin{cases} M_1M_1 \text{ or } M_1M_2 & \text{if } g_m = M_1M_1 \\ M_1M_1,\ M_1M_2 \text{ or } M_2M_2 & \text{if } g_m = M_1M_2 \\ M_1M_2 \text{ or } M_2M_2 & \text{if } g_m = M_2M_2\,. \end{cases} \quad (3)$$

Thus, different mother marker genotypes and different progeny marker genotypes form seven unique *two-level marker genotypes*, *i.e.*, $\{M_1M_1 - M_1M_1\}$, $\{M_1M_1 - M_1M_2\}$, $\{M_1M_2 - M_1M_1\}$, $\{M_1M_2 - M_1M_2\}$, $\{M_1M_2 - M_2M_2\}$, $\{M_2M_2 - M_1M_2\}$, and $\{M_2M_2 - M_2M_2\}$. The number of the progeny of marker genotype $M_{\gamma_1}M_{\gamma_2}$ produced by the ith mother plant of marker genotype $M_{r_1}M_{r_2}$ is denoted by $N_{r_1r_2i}^{\gamma_1\gamma_2}$, where the subscripts stand for the marker genotype of the mother and the superscripts for the marker genotypes of its progeny (r_1, r_2, γ_1, γ_2 = 1 or 2 constrained by Expression 3). The conditional probabilities of the QTL genotypes given each two-level marker genotype are given in Table 2 (see APPENDIX A for the derivations). These conditional probabilities are used to calculate the likelihood of the phenotype for the trait in an open-pollinated progeny design.

Estimation theory: Suppose there is a segregating QTL responsible for a quantitative trait in the half-sib families. The phenotypic value of offspring j in an open-pollinated progeny test at the putative QTL is described by a simple statistical model

$$y_j = \mu + \alpha x_j + \delta z_j + \varepsilon_j, \quad (4)$$

where μ is the overall mean, x_j and z_j are the indicator variables describing the genotypes of the QTL,

$$x_j = \begin{cases} 2 & \text{for QTL genotype } Q_1Q_1 \\ 1 & \text{for QTL genotype } Q_1Q_2 \\ 0 & \text{for QTL genotype } Q_2Q_2 \end{cases}$$

and

$$z_j = \begin{cases} 0 & \text{for QTL genotype } Q_1Q_1 \\ 1 & \text{for QTL genotype } Q_1Q_2, \\ 0 & \text{for QTL genotype } Q_2Q_2 \end{cases}$$

and ε_j is the random error, $\varepsilon_j \sim N(0, \sigma^2)$. The genotypic

TABLE 2

Conditional probabilities of the QTL genotypes upon the marker genotypes of a progeny produced by a mother plant with different marker genotypes (two-level marker genotypes)

Mother		Progeny		QTL genotype[c]		
Genotype	Frequency[a]	Genotype	Frequency[b]	Q_1Q_1 (2)	Q_1Q_2 (1)	Q_2Q_2 (0)
M_1M_1	$p_1^{2(t)}$	M_1M_1	$p_1^{3(t)}$	$\frac{p_{11}^{(t)}p_{11}^{(t+1)}}{p_1^{2(t)}}$	$\frac{p_{11}^{(t)}p_{12}^{(t+1)}+p_{12}^{(t)}p_{11}^{(t+1)}}{p_1^{2(t)}}$	$\frac{p_{12}^{(t)}p_{12}^{(t+1)}}{p_1^{2(t)}}$
		M_1M_2	$p_1^{2(t)}p_2^{(t)}$	$\frac{p_{11}^{(t)}p_{21}^{(t+1)}}{p_1^{(t)}p_2^{(t)}}$	$\frac{p_{11}^{(t)}p_{22}^{(t+1)}+p_{12}^{(t)}p_{21}^{(t+1)}}{p_1^{(t)}p_2^{(t)}}$	$\frac{p_{12}^{(t)}p_{22}^{(t+1)}}{p_1^{(t)}p_2^{(t)}}$
M_1M_2	$2p_1^{(t)}p_2^{(t)}$	M_1M_1	$p_1^{2(t)}p_2^{(t)}$	$\frac{p_{11}^{(t+1)}[p_{11}^{(t)}p_{21}^{(t)}+(1-\theta)]}{p_1^{2(t)}p_2^{(t)}}$	$\frac{p_{11}^{(t+1)}(p_{12}^{(t)}p_{22}^{(t)}+\theta\phi)+p_{12}^{(t+1)}[p_{11}^{(t)}p_{21}^{(t)}+(1-\theta)\phi]}{p_1^{2(t)}p_2^{(t)}}$	$\frac{p_{12}^{(t+1)}(p_{12}^{(t)}p_{22}^{(t)}+\theta\phi)}{p_1^{2(t)}p_2^{(t)}}$
		M_1M_2	$p_1^{(t)}p_2^{(t)}$	$\frac{p_{11}^{(t+1)}(p_{11}^{(t)}p_{21}^{(t)}+\theta\phi)+p_{21}^{(t+1)}[p_{11}^{(t)}p_{21}^{(t)}+(1-\theta)\phi]}{p_1^{(t)}p_2^{(t)}}$	$\frac{p_{11}^{(t+1)}[p_{12}^{(t)}p_{22}^{(t)}+(1-\theta)\phi]+p_{12}^{(t+1)}(p_{11}^{(t)}p_{21}^{(t)}+\theta\phi)}{p_1^{(t)}p_2^{(t)}}+\frac{p_{21}^{(t+1)}(p_{12}^{(t)}p_{22}^{(t)}+\theta\phi)+p_{22}^{(t+1)}[p_{11}^{(t)}p_{21}^{(t)}+(1-\theta)\phi]}{p_1^{(t)}p_2^{(t)}}$	$\frac{p_{12}^{(t+1)}[p_{12}^{(t)}p_{22}^{(t)}+(1-\theta)\phi]+p_{22}^{(t+1)}(p_{12}^{(t)}p_{22}^{(t)}+\theta\phi)}{p_1^{(t)}p_2^{(t)}}$
		M_2M_2	$p_1^{(t)}p_2^{2(t)}$	$\frac{p_{21}^{(t+1)}(p_{11}^{(t)}p_{21}^{(t)}+\theta\phi)}{p_1^{(t)}p_2^{2(t)}}$	$\frac{p_{21}^{(t+1)}[p_{12}^{(t)}p_{22}^{(t)}+(1-\theta)\phi]+p_{22}^{(t+1)}(p_{11}^{(t)}p_{21}^{(t)}+\theta\phi)}{p_1^{(t)}p_2^{2(t)}}$	$\frac{p_{22}^{(t+1)}[p_{12}^{(t)}p_{22}^{(t)}+(1+\theta)\phi]}{p_1^{(t)}p_2^{2(t)}}$
M_2M_2	$p_2^{2(t)}$	M_1M_2	$p_1^{(t)}p_2^{2(t)}$	$\frac{p_{21}^{(t)}p_{11}^{(t+1)}}{p_1^{(t)}p_2^{(t)}}$	$\frac{p_{22}^{(t)}p_{11}^{(t+1)}+p_{21}^{(t)}p_{12}^{(t+1)}}{p_1^{(t)}p_2^{(t)}}$	$\frac{p_{22}^{(t)}p_{12}^{(t+1)}}{p_1^{(t)}p_2^{(t)}}$
		M_2M_2	$p_2^{3(t)}$	$\frac{p_{21}^{(t)}p_{21}^{(t+1)}}{p_2^{2(t)}}$	$\frac{p_{21}^{(t)}p_{22}^{(t+1)}+p_{22}^{(t)}p_{21}^{(t+1)}}{p_2^{2(t)}}$	$\frac{p_{22}^{(t)}p_{22}^{(t+1)}}{p_2^{2(t)}}$

[a] $p_r^{(t)}$'s are the population frequencies of marker alleles.
[b] $p_{rs}^{(t)}$'s are the population frequencies of marker-QTL gametes, M_rQ_s, which are randomly combined to produce diploid plants in the mother generation t.
[c] θ is the recombination fraction between the marker and QTL, and $\phi = p_{11}^{(t)}p_{22}^{(t)} + p_{12}^{(t)}p_{21}^{(t)}$.

values of Q_1Q_1, Q_1Q_2, and Q_2Q_2 are denoted by $\mu + 2\alpha$, $\mu + \alpha + \delta$, and μ, respectively, where μ is the population mean and α and δ are the additive and dominant effects of the QTL. The unknown genetic parameters specifying the genetic architecture of the trait in the progeny population are included in the vector $\Omega = [p_r^{(t)}\ q_s^{(t)}\ D_{rs}^{(t)}\ \theta\ \mu\ \alpha\ \delta\ \sigma^2]^T$. The maximum-likelihood estimates (MLEs) of these parameters can be obtained by maximizing the likelihood of the phenotype ($\mathbf{y}$) and marker ($\mathbf{M}$) data. The likelihood of the phenotypic trait and the marker genotype data observed in the open-pollinated progeny can be written as a mixture model,

$$\begin{aligned}\ell(\mathbf{y}, \mathbf{M}|\Omega) &= \prod_{j=1}^{N}\left[\sum_{\kappa=0}^{2} h_j^{\kappa} f_{\kappa}(y_j)\right] \\ &= \prod_{i=1}^{M_{11}}\left[\prod_{j=1}^{N_{11(t)}^{11}}\sum_{\kappa=0}^{2} h_{11j}^{11\kappa} f_{\kappa}(y_j)\right]\left[\prod_{j=1}^{N_{11(t)}^{12}}\sum_{\kappa=0}^{2} h_{11j}^{12\kappa} f_{\kappa}(y_j)\right] \quad \text{if } g_m = M_1M_1 \\ &\quad \times \prod_{i=1}^{M_{12}}\left[\prod_{j=1}^{N_{12(t)}^{11}}\sum_{\kappa=0}^{2} h_{12j}^{11\kappa} f_{\kappa}(y_j)\right]\left[\prod_{j=1}^{N_{12(t)}^{12}}\sum_{\kappa=0}^{2} h_{12j}^{12\kappa} f_{\kappa}(y_j)\right]\left[\prod_{j=1}^{N_{12(t)}^{22}}\sum_{\kappa=0}^{2} h_{12j}^{22\kappa} f_{\kappa}(y_j)\right] \\ &\qquad \text{if } g_m = M_1M_2 \\ &\quad \times \prod_{i=1}^{M_{22}}\left[\prod_{j=1}^{N_{22(t)}^{12}}\sum_{\kappa=0}^{2} h_{22j}^{12\kappa} f_{\kappa}(y_j)\right]\left[\prod_{j=1}^{N_{22(t)}^{22}}\sum_{\kappa=0}^{2} h_{22j}^{22\kappa} f_{\kappa}(y_j)\right] \quad \text{if } g_m = M_2M_2,\end{aligned} \tag{5}$$

where N is the total number of offspring (seeds) in the open-pollinated progeny design, h_j^{κ} is the conditional probability of the κth QTL given a two-level marker genotype for the jth offspring ($\kappa = 0, 1, 2$), $h_{r_1r_2j}^{\gamma_1\gamma_2\kappa}$ is specified for the offspring marker genotype $M^{\gamma_1}M^{\gamma_2}$ and mother genotype $M_{r_1}M_{r_2}$ (Table 2), and $f_{\kappa}(y_j)$ is the normal distribution density function having the form

$$f_{\kappa}(y_j) = \frac{1}{\sqrt{2\pi\sigma}}\exp\left[-\frac{(y_j - \mu_{\kappa})^2}{2\sigma^2}\right],$$

$$\mu_{\kappa} = \mu + \kappa\alpha + (2 - \kappa)\kappa\delta, \quad \kappa = 0, 1, 2.$$

Calculating the MLEs of Ω is equivalent to differentiating the log-likelihood of Equation 5 with respect to each of the unknown genetic parameters, setting the derivatives to equal zero, and solving the log-likelihood equations. On the basis of these procedures, we can obtain the explicit ML estimator of marker allele frequency p_1:

$$\hat{p}_1 = \frac{1}{2N}\left[\sum_{i=1}^{M_{11}}(2N_{11i}^{11} + N_{11i}^{12}) + \sum_{i=1}^{M_{12}}(2N_{12i}^{11} + N_{12i}^{12}) + \sum_{i=1}^{M_{22}}N_{22i}^{12}\right]. \tag{6}$$

For the other parameters $\Omega_{-} = [q_s^{(t)}\ D_{rs}^{(t)}\ \theta\ \mu\ \alpha\ \delta\ \sigma^2]^T$, it is not possible to derive explicit ML estimators. To obtain MLEs for these parameters, the EM algorithm (Dempster *et al.* 1977) is used, which initializes from an arbitrary value of each of the parameters (appendix b).

The existence of the QTL under consideration can be tested by formulating the two hypotheses

$$\begin{aligned}&H_0: \alpha = \delta = 0 \\ &H_1: \text{at least one of them is not equal to zero.}\end{aligned} \tag{7}$$

A log-likelihood ratio (LR) test statistic for the test of these two hypotheses is calculated using

$$LR_Q = -2\log\left[\frac{\ell(\mathbf{y}, \mathbf{M}|\tilde{\Omega})_{\alpha=\delta=0}}{\ell(\mathbf{y}, \mathbf{M}|\hat{\Omega})}\right],$$

where $\hat{\Omega}$ and $\tilde{\Omega}$ denote the MLEs of the unknown vector under the full model (H_1) and reduced model (H_0), respectively, and LR_Q asymptotically follows the χ^2 distribution with 2 d.f. The hypotheses for testing the linkage disequilibrium detected in the progeny generation $t + 1$ can be formulated as

$$\begin{aligned}&H_0: (1 - \theta)D^{(t)} = 0 \\ &H_1: (1 - \theta)D^{(t)} \neq 0,\end{aligned} \tag{8}$$

with the corresponding LR_D approximately χ^2 distributed with 1 d.f. (Weir 1996). The acceptance of the null hypothesis of (8) may be due to either no linkage disequilibrium or a combination of loose linkage and weak linkage disequilibrium. The rejection of the null hypothesis of (8), on the other hand, exclusively reveals strong linkage disequilibrium with or without tight linkage. Further hypotheses for testing whether there is a significant linkage can be formulated as

$$\begin{aligned}&H_0: \theta = 0.5 \\ &H_1: \theta \neq 0.5,\end{aligned} \tag{9}$$

with the LR_R also approximately χ^2 distributed with 1 d.f. (Terwilliger 1995). If the null hypothesis of (8) is rejected and the null hypothesis of (9) is accepted, then a significant linkage disequilibrium detected between a marker and QTL in the progeny generation is not due to their strong linkage. In this case, results from pure linkage disequilibrium mapping (Luo *et al.* 2000; Meuwissen and Goddard 2000) are ineffective for genome mapping because the linkage disequilibrium detected is spurious. If a tight linkage is detected, one can further test whether such a linkage is tight enough to the fine mapping of QTL. This test can be carried out by letting θ equal a particular small value, *e.g.*, 0.01. In summary, by testing simultaneously for the significance of linkage and linkage disequilibrium, our analytical approach increases the predictability of gene mapping in a natural population.

SIMULATION

The statistical properties of the mapping method proposed in this article are examined by using simulated examples. Suppose the mother plants from which seeds

are collected and grown into seedlings are randomly sampled from a panmictic population. A biallelic marker locus and a biallelic QTL, each of which is segregating in the population, are genetically associated. A number of factors may affect the precision and power of the method to detect the putative QTL, which include sampling schemes, the degree of marker and QTL segregation, the degree of linkage and linkage disequilibrium, and the mode of QTL gene action.

The effects of sampling schemes and population heterozygosity: How the size of samples and their allocation between and within open-pollinated families affect the behavior of a statistical method in a mapping experiment is an important issue for a practitioner to examine. In this simulation, we investigate the effects of three different sampling schemes on parameter estimation. The three schemes include (1) few large families (10×100), (2) moderately sized families of a moderate number (32×32), and (3) many small families (100×10). Also, the effects of sampling schemes are affected by other factors, such as gene segregation, the degree of nonrandom association between the marker and QTL, and the QTL effect. The effect due to the interaction between sampling schemes and gene segregation is examined. Gene segregation for a gene in a population is described by the difference in the frequencies of alternative alleles at the gene. A larger difference (say 0.10 *vs.* 0.90) implies that a population is closer to fixture and has a smaller degree of segregation. Otherwise, a population of a smaller difference in allele frequency (say 0.50 *vs.* 0.50) has a larger degree of segregation. Table 3 gives the parameter values used to simulate the effects of sampling schemes and gene segregation. Assuming each of the M selected open-pollinated families has an equal size, the phenotype and marker data are generated using the following steps:

Step 1. Randomly assign three marker genotypes to the M hypothesized mother plants according to probabilities $p_1^{2(t)}(M_1M_1)$, $2p_1^{(t)}p_2^{(t)}(M_1M_2)$, and $p_2^{2(t)}(M_2M_2)$.

Step 2. Randomly assign three marker genotypes to the progeny within a mother plant of a particular marker genotype according to probabilities of the marker genotypes of the progeny (Table 2).

Step 3. Randomly sample joint genotypes at both the marker and QTL for an offspring derived from each mother plant from a multinomial distribution with the probabilities calculated from Table 2.

Step 4. Determine the phenotypic value for an individual with a given marker-QTL joint genotype by its genotypic value of the QTL plus a random number sampled from a normal distribution of mean zero and variance $\sigma^2 = 1$.

The mean and standard error of the MLE for each of the unknowns over 100 replicates of simulation are given in Table 1. The MLE of the marker allele frequency is estimated directly, using Equation 6. The estimation for the other parameters is viewed as a missing data problem. In general, the EM algorithm derived in this article can provide the unknown parameters with consistent MLEs compared to their actual values. Yet, the precisions of parameter estimations in terms of the standard errors estimated from multiple simulation runs are greater when using a sampling scheme of few large families (10×100) than of many small families (100×10). Such precision improvement due to the use of a better sampling scheme is much more remarkable when the population sampled is closer to fixture. For example, when the difference in allele frequency for both marker and QTL is 0.80, the standard error for the allele frequency of the QTL is 0.0151 for many small families and 0.0087 for few large families. But the corresponding values are 0.0105 and 0.0081 for a population having an equal frequency for the alternative alleles at the same locus.

The power of detecting a significant linkage disequilibrium using our method is also investigated. For a less segregating population, the power is strongly dependent on the sampling scheme used, with 0.79 for many small families and 0.95 for few large families (Table 1).

The effects of linkage and linkage disequilibrium: Because missing information about the QTL is inferred from the marker genotype, the relationship between the marker and QTL affects the estimates for genetic parameters. Here, four different relationship patterns are compared on the basis of a sampling scheme 32×32: (1) tight linkage and weak disequilibrium, (2) tight linkage and strong disequilibrium, (3) loose linkage and weak disequilibrium, and (4) loose linkage and strong disequilibrium (Table 4). In these four patterns, all parameters except recombination fraction and linkage disequilibrium are set equal. As expected, the marker allele frequency can be very well estimated. Given the same linkage between the marker and QTL, a more associated marker tends to provide more precise estimates for both the population genetic (allele frequency) and quantitative genetic parameters of the QTL (the overall mean, additive and dominant effect, and residual variance) than a less associated marker. Also, as shown in our simulation example, there is significantly greater power to detect a QTL using a more associated marker [$D_{rs}^{(t)} = 0.20$] than a less associated marker [$D_{rs}^{(t)} = 0.02$]. Similarly, given the same disequilibrium, a more linked marker displays greater precision and greater power for estimating a QTL than a less linked marker. When the marker has a loose linkage and weak disequilibrium with the QTL, the marker information provides little information about the genotype at the QTL. Under this circumstance, the MLEs for the QTL parameters are biased with lower precision compared to the other patterns. The power to detect an existing QTL based on the information of a marker with loose linkage ($\theta =$

TABLE 3

Means and standard errors (in parentheses) of the MLEs of the genetic parameters for different sampling schemes and different heterogeneity in allelic frequency from 100 simulation replicates

Sampling scheme	Allele segregation	Genetic parameter								
		$p_1^{(t)} = 0.1$	$q_1^{(t)} = 0.1$	$\theta = 0.10$	$D_{ns}^{(t)} = 0.15$	$\mu = 0$	$\alpha = 1$	$\delta = 1$	$\sigma^2 = 1$	Power
100 × 10	Low	0.1082	0.1174	0.1079	0.1483	0.0056	1.0691	1.0298	1.0441	0.79
		(0.0019)	(0.0151)	(0.0088)	(0.0071)	(0.0087)	(0.0227)	(0.0281)	(0.0393)	
32 × 32	Low	0.1042	0.1103	0.1056	0.1542	0.0054	0.9758	1.0212	1.0329	0.89
		(0.0018)	(0.0106)	(0.0047)	(0.0032)	(0.0015)	(0.0143)	(0.0192)	(0.0265)	
10 × 100	Low	0.1025	0.1082	0.1009	0.1485	0.0032	0.9874	1.0106	0.9861	0.95
		(0.0015)	(0.0087)	(0.0045)	(0.0009)	(0.0011)	(0.0115)	(0.0187)	(0.0239)	
Sampling scheme	Allele segregation	$p_1 = 0.5$	$q_1 = 0.5$	$\theta = 0.10$	$D = 0.20$	$\mu = 0$	$\alpha = 1$	$\delta = 1$	$\sigma^2 = 1$	Power
100 × 10	High	0.1022	0.1087	0.1023	0.1431	0.0045	0.9726	1.0203	1.0216	0.91
		(0.0014)	(0.0105)	(0.0052)	(0.0044)	(0.0048)	(0.0121)	(0.0216)	(0.0235)	
32 × 32	High	0.1021	0.1082	0.1008	0.1439	0.0032	0.9821	1.0189	0.9823	0.94
		(0.0009)	(0.0091)	(0.0039)	(0.0021)	(0.0020)	(0.0099)	(0.0197)	(0.0223)	
10 × 100	High	0.1009	0.1075	0.1012	0.1471	0.0031	0.9923	1.0125	1.0231	0.96
		(0.0010)	(0.0081)	(0.0042)	(0.0008)	(0.0012)	(0.0098)	(0.0180)	(0.0210)	

The sampling scheme is indicated by family number × family size, $p_1^{(t)}$ and $q_1^{(t)}$ are the frequencies of marker and QTL alleles, θ and $D_{ns}^{(t)}$ are the recombination fraction and linkage disequilibrium between the marker and QTL, μ and σ^2 are the mean and residual variance for the trait, and α and δ are the additive and dominant effects of the QTL. The power is indicated as the probability of detecting a significant linkage disequilibrium between the marker and QTL.

TABLE 4

Means and standard errors (in parentheses) of the MLEs of genetic parameters

			Disequilibrium ($D_{rs}^{(t)}$)	
		Parameter	Weak (0.02)	Strong (0.20)
Linkage (θ)	Tight (0.02)	$p_1^{(t)} = 0.5$	0.5008 (0.0017)	0.5011 (0.0016)
		$q_1^{(t)} = 0.5$	0.5053 (0.0128)	0.5077 (0.0134)
		θ	0.0195 (0.0026)	0.0208 (0.0031)
		$D_{rs}^{(t)}$	0.0094 (0.0028)	0.1920 (0.0051)
		$\mu = 0$	0.0081 (0.0074)	0.0059 (0.0046)
		$\alpha = 1$	1.0086 (0.0298)	1.0021 (0.0137)
		$\delta = 1$	0.9656 (0.0345)	1.0083 (0.0287)
		$\sigma^2 = 1$	0.9394 (0.0347)	1.0294 (0.0210)
		Power	0.59	0.99
	Loose (0.02)	$p_1^{(t)} = 0.5$	0.5009 (0.0018)	0.5089 (0.0041)
		$q_1^{(t)} = 0.5$	0.5095 (0.0210)	0.5120 (0.0239)
		θ	0.2023 (0.0211)	0.2035 (0.0127)
		$D_{rs}^{(t)}$	0.0090 (0.0014)	0.1934 (0.0157)
		$\mu = 0$	0.0090 (0.0085)	0.0099 (0.0082)
		$\alpha = 1$	1.0354 (0.0438)	1.0125 (0.0313)
		$\delta = 1$	1.0387 (0.0498)	1.0347 (0.0432)
		$\sigma^2 = 1$	0.9135 (0.0459)	0.9782 (0.0291)
		Power	0.41	0.84

The power is expressed as the probability for detecting a QTL with significant additive or dominant effects among 100 simulation replicates (the sampling strategy used is 32 × 32).

0.20) and weak disequilibrium [$D_{rs}^{(t)} = 0.02$] is typically low (Table 5).

The effects of linkage and linkage disequilibrium on parameter estimation vary among different parameters. Generally, these effects are larger on the estimates of the dominant effect of the QTL and residual variance than the additive effect and overall mean (Table 4).

The effects of QTL gene action: It has been well demonstrated that the magnitude of QTL effect affects parameter estimation, with a QTL of large effect being estimated more precisely than a QTL of small effect. Similar results have also been observed in the linkage disequilibrium-based mapping of QTL (Luo and Suhai 1999; Luo *et al.* 2000). However, it is unclear how different modes of gene action affect the precision and power of parameter estimation in linkage disequilibrium mapping. A simulation here is designed to investigate the effect of gene action of the estimates of QTL parameters.

Our simulation on gene action includes four different patterns: (1) purely additive ($\delta = 0$), (2) partially dominant ($0 < \delta/\alpha < 1$), (3) dominant ($\delta/\alpha = 1$), and (4) overdominant ($\delta/\alpha > 1$). Except for the marker allele frequency, all other parameters have a consistent trend in the precision and power of parameter estimation over gene action (Table 5). As shown by the estimates of standard error, a dominant QTL can be estimated more precisely than an additive QTL. Also, an overdominant QTL can be estimated more precisely than a dominant or partially dominant QTL. However, the power to detect a significant linkage disequilibrium between the marker and QTL is greater for an additive QTL than for a dominant QTL as well as for a partially dominant than for an overdominant QTL (Table 5).

Comparison between traditional disequilibrium mapping an our joint mapping: We conduct an additional simulation study to compare the power for detecting linkage disequilibrium on the basis of the traditional disequilibrium mapping approach (Allison 1997; Luo *et al.* 2000) and our joint linkage and linkage disequilibrium mapping approach. For comparison, the same sets of genetic parameters are hypothesized between the two approaches, each allowing for different combinations between linkage and disequilibrium (Table 6). For both approaches a sample size of 1000 is assumed. For the pure disequilibrium mapping approach, this sample is randomly derived from a natural population, representing the same generation. But for our joint linkage and linkage disequilibrium mapping approach, this sample is allocated between the parental generation and the open-pollinated progeny generation. Here, the sampling scheme of 32 × 32 is simulated.

Table 6 shows the observed power for detecting linkage disequilibrium using the two mapping approaches. Generally, greater power is observed for the joint linkage and linkage disequilibrium analysis than for the pure disequilibrium analysis. However, the increase of the power by using the joint analysis depends on the degrees of linkage and linkage disequilibrium between

TABLE 5

Means and standard errors (in parentheses) of the MLEs of the genetic parameters for different effects of the QTL from 100 simulation replicates

Gene action	Additive (α)	Dominant (δ)	Genetic parameter							Power
			$p_1^{(l)} = 0.5$	$q_1^{(l)} = 0.5$	$\theta = 0.10$	$D_{rs}^{(l)} = 0.10$	$\mu = 0$	α	δ	$\sigma^2 = 1$	
Additive	1	0	0.5015	0.5126	0.1082	0.1082	0.0089	1.0376	0.0092	1.0892	0.82
			(−0.0035)	(−0.0234)	(−0.0091)	(−0.0098)	(−0.0075)	(−0.0195)	(−0.0026)	(−0.0451)	
Partially dominant	1	0.5	0.5009	0.5103	0.1076	0.1047	0.0071	1.0202	0.4871	1.07	0.86
			(−0.0028)	(−0.021)	(−0.0093)	(−0.0081)	(−0.0042)	(−0.0103)	(−0.0125)	(−0.0342)	
Dominant	1	1	0.5013	0.5092	0.103	0.1029	0.0064	1.0156	1.0787	1.0679	0.91
			(−0.0023)	(−0.0123)	(−0.0087)	(−0.0076)	(−0.0039)	(−0.0089)	(−0.0193)	(−0.0311)	
Overdominant	1	2	0.5008	0.5089	0.1023	0.1036	0.0052	1.0081	2.1982	1.0569	0.98
			(−0.0022)	(−0.0118)	(−0.0085)	(−0.0083)	(−0.0021)	(−0.0056)	(−0.0677)	(−0.029)	

The power is indicated as the probability for detecting a significant linkage disequilibrium between the marker and QTL. The sampling strategy used is 32×32.

TABLE 6

Difference in power to detect the linkage disequilibrium between a marker and QTL using the existing (top) and our (bottom) approaches

	Disequilibrium ($D_{rs}^{(l)}$)	
Linkage (θ)	Weak (0.02)	Strong (0.20)
Tight (0.02)	0.20	0.95
	0.64	0.99
Loose (0.20)	0.12	0.78
	0.53	0.84

The values of the other genetic parameters are hypothesized as $p_1^{(l)} = 0.5$, $q_1^{(l)} = 0.5$, $\theta = 0.10$, $\mu = 0$, $\alpha = 1$, $\delta = 1$, and $\sigma^2 = 1$.

a marker and QTL. In the situations where the linkage is loose or the disequilibrium is weak, the joint mapping approach has significantly increased power compared to the traditional disequilibrium mapping approach.

DISCUSSION

We have provided a unifying framework for the fine-scale mapping of QTL affecting a quantitative trait in a natural population on the basis of a joint linkage and linkage disequilibrium mapping strategy proposed by Wu and Zeng (2001). We model marker-QTL association on the basis of a random sample (mothers) drawn from a natural population and marker-QTL linkage on the basis of the open-pollinated progeny of the sample in which recombination events happen between different loci. Such a two-stage (mother and progeny) hierarchical modeling provides a simultaneous estimate of the linkage and linkage disequilibrium between the marker and QTL, which is thus beyond many existing composite linkage disequilibrium mapping methods that cannot distinguish strong association and loose linkage from weak association and tight linkage (Terwilliger 1995; Xiong and Guo 1997; Collins and Morton 1998; Terwilliger and Weiss 1998; Luo *et al.* 2000). Moreover, by unifying the information about linkage and linkage disequilibrium, the joint mapping method displays increased power to detect linkage disequilibrium, compared to the traditional linkage disequilibrium analyses.

As an example, we used a simpler one-biallelic codominant marker/one-biallelic QTL model to demonstrate the statistical properties of the joint linkage and linkage disequilibrium analysis in the precise mapping of individual QTL for complex trait. Linkage analysis requiring informative meioses in a pedigree can rarely detect a target gene that is within 1 cM of markers, but it should be useful for a genome-wide scan for QTL because a high-density map covering the entire genome can be constructed in a single pedigree. Thus, through a genome-wide scan for QTL using linkage analysis, genomic regions containing QTL can first be identified. These identified

regions are then saturated by more markers and are further narrowed around QTL, using the joint linkage and linkage disequilibrium mapping strategy. We employ the maximum-likelihood method implemented with the EM algorithm to obtain the MLEs for model parameters including the allele frequency of QTL, its effects, its location, and its linkage disequilibrium with a marker. Extensive simulation studies show that the method can provide reasonable estimates for these genetic and genomic parameters for a wide range of parameter values.

In the current modeling, we have not considered the phenotypes of the genotyped mothers sampled from a natural population and used to supply the next progeny (contained in seeds). Yet, this would not affect the efficiency and utility of the model because we have integrated mothers' marker genotypes and progeny's marker genotypes into a two-level marker genotype framework. Thus, the phenotypes of the progeny population can be directly associated with the two-level marker genotypes. The strategy with no need of mothers' phenotypes is practically advantageous in at least two aspects. First, for species like forest trees, sample mothers from a natural population are easily genotyped, but their phenotypes are difficult to measure. Second, the mothers sampled cannot be compared to their progeny in phenotypes because of different ages and growth environments. However, for some species that can be vegetatively propagated, a field trial can be established with clonal replicates of both mothers and their progeny. In this case, mothers and their progeny with the same age can be simultaneously measured and compared. A further simulation study is needed to examine the advantage of the model implemented with mothers' phenotypes.

Although the codominant marker assumption used can be valid by genotyping markers like SNPs, there are many dominant markers derived from rapid amplified polymorphic DNAs and amplified fragment length polymorphisms in real data analyses for natural outcrossing populations. Also, with no doubt, our one marker-one QTL model is too simplistic for a quantitative trait that may be controlled by multiple genes each with a different effect. For these two practical reasons, the joint linkage and linkage disequilibrium mapping approach needs extension to allow for multiple markers including dominant and multiallelic markers. Linkage analysis in a pedigree using dominant markers is often biased and has low precision especially when a sample size is small (Maliepaard *et al.* 1997). But these problems can be overcome if they are combined with codominant markers through a Markov chain (Jiang and Zeng 1997). For the linkage disequilibrium analysis of dominant markers, a similar improvement in the precision of parameter estimate can also be expected from their combined use with codominant markers. For multiple alleles and/or loci, basic extension of the single-marker disequilibrium measures presented above has been developed in the current literature. Like linkage analysis, multipoint disequilibrium can be more efficient than single-marker analysis. For example, Hill and Weir (1994) showed that the variance of the linkage disequilibrium between a closely linked marker and a QTL is large, such that the disequilibrium cannot be used for the precise mapping of the QTL. When the disequilibria between all markers and the QTL are analyzed simultaneously, the problem of a high variability of a single linkage disequilibrium is avoided (Meuwissen and Goddard 2000). A likelihood-based multipoint approach to linkage disequilibrium mapping loci can be found in Terwilliger (1995), McPeek and Strahs (1999), Meuwissen and Goddard (2000), and Morris *et al.* (2000). When a narrow region is being considered for linkage disequilibrium fine-scale mapping, conditioning on the distances between markers allows the use of a composite likelihood to extract information from multiple markers. Xiong and Guo (1997) give a general likelihood framework for linkage disequilibrium mapping that incorporates multiallelic markers, multiple loci, and mutational processes at the disease and marker alleles.

For a multi-QTL model, a number of genetic parameters are treated as unknown. These include the number of QTL, the additive and dominant effect of each QTL, different kinds of epistatic effect between each pair of QTL, the chromosomal location of each QTL (determined by the recombination fraction between each QTL and its flanking markers), the linkage disequilibrium between each pair of QTL, and the linkage disequilibrium between each QTL and each marker. The maximum-likelihood method that works in a one-marker/one-QTL case may be insufficient for handling such a high dimension of unknowns. Markov chain Monte Carlo (MCMC) methods within a Bayesian framework may be a better solution for our multi-QTL linkage and linkage disequilibrium mapping. Unlike the traditional maximum-likelihood method, MCMC methods provide estimates for unobservables by analyzing their posterior distributions (Robert and Casella 1999). In the MCMC paradigm, we are able to incorporate prior information for model parameters including the number of QTL, where appropriate, which is thus advantageous over the maximum-likelihood method. Given the impressive applications of the Bayesian approach in QTL linkage mapping (see Satagopan *et al.* 1996; Sillanpaa and Arjas 1996, 1999; Uimari *et al.* 1996; Heath 1997; Stephens and Fisch 1998), we are confident of developing a powerful Bayesian approach for joint linkage and linkage disequilibrium mapping of multiple QTL through the entire genome.

We are grateful to Dr. Bruce Weir, Dr. Shizhong Xu, Dr. Nengjun Yi, and Dr. Zhao-Bang Zeng for stimulating discussions about this study, and to the associate editor and two anonymous referees for their helpful comments on this manuscript. This manuscript was approved for publication as journal series no. R-07961 by the Florida Agricultural Experiment Station.

LITERATURE CITED

Allison, D. B., 1997 Transmission disequilibrium tests for quantitative traits. Am. J. Hum. Genet. **60:** 676–690.

Collins, A., and N. E. Morton, 1998 Mapping a disease locus by allelic association. Proc. Natl. Acad. Sci. USA **95:** 1741–1745.

Darvasi, A., A. Weinreb, V. Minke, J. I. Weller and M. Soller, 1993 Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. Genetics **134:** 943–951.

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

Gordon, D., I. Simonic and J. Ott, 2000 Significant evidence for linkage disequilibrium over a 5-cM region among Afrikaners. Genomics **66:** 87–92.

Hästbacka, J., A. de la Chapelle, I. Kaitila, P. Sistonen, A. Weaver *et al.*, 1992 Linkage disequilibrium mapping in isolated founder populations: diastropic dysplasia in Finland. Nat. Genet. **2:** 204–221.

Hästbacka, J., A. de la Chapelle, M. Mahtani, G. Clines, M. P. Reeve-Daly *et al.*, 1994 The diastropic dysplasia gene encodes a novel sulfate transporter: positional cloning by fine-structure linkage disequilibrium mapping. Cell **78:** 1073–1087.

Heath, S. C., 1997 Markov chain Monte Carlo segregation and linkage analysis for oligogenic models. Am. J. Hum. Genet. **61:** 748–760.

Hill, W. G., and B. S. Weir, 1994 Maximum-likelihood estimation of gene location by linkage disequilibrium. Am. J. Hum. Genet. **54:** 705–714.

Long, A. D., S. L. Mullaney, L. A. Reid, J. D. Fry, C. H. Langley *et al.*, 1995 High resolution mapping of genetic factors affecting abdominal bristle number in *Drosophila melanogaster*. Genetics **139:** 1273–1291.

Luo, Z. W., and S. Suhai, 1999 Estimating linkage disequilibrium between a polymorphic marker locus and a trait locus in natural populations. Genetics **151:** 359–371.

Luo, Z. W., S. H. Tao and Z-B. Zeng, 2000 Inferring linkage disequilibrium between a polymorphic marker locus and a trait locus in natural populations. Genetics **156:** 457–467.

Mackay, T. F. C., 2001 Quantitative trait loci in Drosophila. Nat. Rev. Genet. **2:** 11–20.

Maliepaard, C., J. Jansen and J. W. van Ooijen, 1997 Linkage analysis in a full-sib family of an outbreeding plant species: overview and consequences for applications. Genet. Res. **70:** 237–250.

McKeand, S. E., and F. E. Bridgwater, 1998 A strategy for the third breeding cycle of loblolly pine in the Southeastern US. Silvae Genet. **47:** 223–234.

McPeek, M. S., and A. Strahs, 1999 Assessment of linkage disequilibrium by the decay of haplotype sharing, with application to fine-scale genetic mapping. Am. J. Hum. Genet. **65:** 858–875.

Meuwissen, T. H. E., and M. E. Goddard, 2000 Fine mapping of quantitative trait loci using linkage disequilibria with closely linked marker loci. Genetics **155:** 421–430.

Morris, A. P., J. C. Whittaker and D. J. Balding, 2000 Bayesian fine-scale mapping of disease loci, by hidden Markov models. Am. J. Hum. Genet. **67:** 155–169.

Nagylaki, T., 1992 *Introduction to Theoretical Population Genetics*. Springer-Verlag, Berlin.

Olson, J. M., J. S. Witte and R. C. Elston, 1999 Tutorial in biostatistics genetic mapping of complex traits. Stat. Med. **18:** 2961–2981.

Pfeiffer, A., A. M. Olivieri and M. Morgante, 1997 Identification and characterization of microsatellites in Norway spruce (*Picea abies* K.). Genome **40:** 411–419.

Press, W. H., S. A. Teukolsky, W. T. Vetterling and B. P. Flannery, 1992 *Numerical Recipes: The Art of Scientific computing*. Cambridge University Press, New York.

Rabinowitz, D., 1997 A transmission disequilibrium test for quantitative trait loci. Hum. Hered. **47:** 342–350.

Risch, N., and K. Merikangas, 1996 The future of genetic studies of complex human diseases. Science **273:** 1516–1517.

Robert, C. P., and G. Casella, 1999 *Monte Carlo Statistical Methods*. Springer, New York.

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

Sillanpaa, M. J., and E. Arjas, 1996 Bayesian mapping of multiple quantitative trait loci from incomplete inbred line cross data. Genetics **148:** 1373–1388.

Sillanpaa, M. J., and E. Arjas, 1999 Bayesian mapping of multiple quantitative trait loci from incomplete outbred offspring data. Genetics **151:** 1605–1619.

Spielman, R. S., and W. J. Ewens, 1996 The TDT and other family-based tests for linkage disequilibrium and association. Am. J. Hum. Genet. **59:** 983–989.

Stephens, D. A., and R. D. Fisch, 1998 Bayesian analysis of quantitative trait locus data using reversible jump Markov chain Monte Carlo. Biometrics **54:** 1334–1347.

Tanksley, S. D., 1993 Mapping genes. Annu. Rev. Genet. **27:** 205–233.

Templeton, A. R., 1999 Uses of evolutionary theory in the human genome project. Annu. Rev. Ecol. Syst. **30:** 23–49.

Terwilliger, J. D., 1995 A powerful likelihood method for the analysis of linkage disequilibrium between trait loci and one or more polymorphic marker loci. Am. J. Hum. Genet. **56:** 777–787.

Terwilliger, J. D., and K. M. Weiss, 1998 Linkage disequilibrium mapping of complex disease: Fantasy or reality? Curr. Opin. Biotechnol. **9:** 578–594.

Uimari, P., G. Thaller and I. Hoeschele, 1996 The use of multiple markers in a Bayesian method for mapping quantitative trait loci. Genetics **143:** 1831–1842.

Wang, D. G., J. B. Fan, C. J. Siao, A. Berno, P. Young *et al.*, 1998 Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. Science **280:** 1077–1082.

Weber, J. L., and C. Wong, 1993 Mutation of human short tandem repeats. Hum. Mol. Genet. **2:** 1123–1128.

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

Whittaker, J. C., M. C. Denham and A. P. Morris, 2000 The problems of using transmission/disequilibrium test to infer tight linkage. Am. J. Hum. Genet. **67:** 523–526.

Wu, R. L., and Z-B. Zeng, 2001 Joint linkage and linkage disequilibrium mapping in natural populations. Genetics **157:** 899–909.

Wu, R. L., Z-B. Zeng, S. E. McKeand and D. M. O'Malley, 2000 The case for molecular mapping in forest tree breeding. Plant Breed. Rev. **19:** 41–68.

Xiong, M. M., and S. W. Guo, 1997 Fine-scale genetic mapping based on linkage disequilibrium: theory and applications. Am. J. Hum. Genet. **60:** 1513–1531.

Communicating editor: Y.-X. Fu

APPENDIX A

We describe a procedure for deriving the conditional probabilities of QTL genotypes given two-level (mother and progeny) marker genotypes. Let us first consider the mothers with marker genotype M_1M_1. This mother marker genotype is composed of three joint marker-QTL genotypes $M_1M_1Q_1Q_1$, $M_1M_1Q_1Q_2$, and $M_1M_1Q_2Q_2$, with respective population frequencies in the mother generation (t) as $P_{1111}^{(t)} = p_{11}^{2(t)}$, $P_{1112}^{(t)} = 2p_{11}^{(t)}p_{12}^{(t)}$, and $P_{1122}^{(t)} = p_{12}^{2(t)}$ (Table 1). Each of these three mother two-locus genotypes generates either a two-locus gamete M_1Q_1 or M_1Q_2, or both, which are combined with all four possible two-locus gametes M_1Q_1, M_1Q_2, M_2Q_1, and M_2Q_2 from the pollen pool, with population frequencies $p_{11}^{(t+1)}$, $p_{12}^{(t+1)}$, $p_{21}^{(t+1)}$, and $p_{22}^{(t+1)}$, respectively, to produce the progeny generation ($t + 1$) (contained in seeds). Here it is not difficult to calculate the probabilities of different joint marker-QTL genotypes in the progeny population. For example, the probability of progeny joint genotype $M_1M_1Q_1Q_1$ derived from mother genotype M_1M_1 is the sum of $p_{11}^{2(t)} \cdot p_{11}^{(t+1)} + p_{11}^{(t)}p_{12}^{(t)} \cdot p_{11}^{(t+1)} = p_1^{(t)}p_{11}^{(t)}p_{11}^{(t+1)}$, where the first part results from the combi-

nation of the same gamete genotype M_1Q_1 from mother genotype $M_1M_1Q_1Q_1$ and the pollen pool and the second part from the combination of the same gamete genotype M_1Q_1 from mother genotype $M_1M_1Q_1Q_2$ and the pollen pool. Thus, according to Bayes' theorem, the conditional probability of the QTL genotype Q_1Q_1, given the mother's marker genotype M_1M_1 and progeny's marker genotype M_1M_1, is

$$\begin{aligned}\Pr(Q_1Q_1|\{M_1M_1 - M_1M_1\}) &= \frac{\Pr(\{M_1M_1 - M_1M_1Q_1Q_1\})}{\Pr(\{M_1M_1 - M_1M_1\})}\\ &= \frac{p_1^{(t)}p_{11}^{(t)}p_{11}^{(t+1)}}{p_1^{3(t)}}\\ &= \frac{p_{11}^{(t)}p_{11}^{(t+1)}}{p_1^{2(t)}}\end{aligned}$$

The probability of progeny joint genotype $M_1M_1Q_1Q_2$ derived from mother genotype M_1M_1 includes two components: (1) $p_{11}^{2(t)} \cdot p_{12}^{(t+1)} + p_{11}^{(t)}p_{12}^{(t)} \cdot p_{12}^{(t+1)} = p_1^{(t)}p_{11}^{(t)}p_{12}^{(t+1)}$ from the mating of mother gamete genotype M_1Q_1 and father gamete genotype M_1Q_2 from the pollen pool and (2) $p_{12}^{2(t)} \cdot p_{11}^{(t+1)} + p_{11}^{(t)}p_{12}^{(t)} \cdot p_{11}^{(t+1)} = p_1^{(t)}p_{12}^{(t)}p_{11}^{(t+1)}$ from the mating of mother gamete genotype M_1Q_2 and father gamete genotype M_1Q_1 from the pollen pool. The conditional probability of the QTL genotype Q_1Q_2 given the mother's marker genotype M_1M_1 and progeny's marker genotype M_1M_1 is thus calculated as $[p_{11}^{(t)}p_{12}^{(t+1)} + p_{12}^{(t)}p_{11}^{(t+1)}]/p_1^{2(t)}$. The rest of the conditional probabilities of the QTL genotypes given the mother's marker genotype M_1M_1 can also be calculated (see Table 2).

When the marker genotype of a sampled mother is M_1M_2, three possible joint marker-QTL genotypes are $M_1M_2Q_1Q_1$, $M_1M_2Q_1Q_2$, and $M_1M_2Q_2Q_2$, with population frequencies as $P_{1211}^{(t)} = 2p_{11}p_{21}$, $P_{1212}^{(t)} = 2[p_{11}^{(t)}p_{22}^{(t)} + p_{12}^{(t)}p_{21}^{(t)}]$, and $P_{1222}^{(t)} = 2p_{12}^{(t)}p_{22}^{(t)}$ in the generation t, respectively. The probabilities of four joint marker-QTL gamete genotypes generated by each of these three joint genotypes are given in Table 1. Thus, the probability of progeny joint genotype $M_1M_1Q_1Q_1$ derived from mother marker genotype M_1M_2 is the sum of $p_{11}^{(t)}p_{21}^{(t)} \cdot p_{11}^{(t+1)}$ and $(1 - \theta)\,[p_{11}^{(t)}p_{22}^{(t)} + p_{12}^{(t)}p_{21}^{(t)}] \cdot p_{11}^{(t+1)}$. The conditional probability of the QTL genotype given the mother's marker genotype M_1M_2 and progeny's marker genotype M_1M_1 can be calculated accordingly. The probabilities of all QTL genotypes conditional upon different progeny marker genotypes derived from the mother's marker genotype M_1M_2 are derived in Table 2.

A similar procedure can be described to derive the conditional probabilities of different QTL genotypes when the mother's marker genotype is M_2M_2 (Table 2).

APPENDIX B

The MLEs of the unknown parameters $\Omega_- = (q_1^{(t)}\ D_{rs}^{(t)}\ \theta\ \mu\ \alpha\ \delta\ \sigma^2)^T$ can be computed by implementing an EM algorithm (Dempster *et al.* 1977; Meng and Rubin 1993). The log-likelihood is given by

$$\log L(\Omega_-) = \sum_{j=1}^{N} \log\left[\sum_{\kappa=0}^{2} h_j^{\kappa} f_{\kappa}(y_j)\right]$$

with derivatives

$$\begin{aligned}\frac{\partial}{\partial\Omega_-}\log L(\Omega_-) &= \sum_{j=1}^{N}\sum_{\kappa=0}^{2}\frac{h_j^{\kappa}(\partial/\partial\Omega_-)f_{\kappa}(y_j)}{\sum_{\kappa=0}^{2}h_j^{\kappa}f_{\kappa}(y_j)}\\ &= \sum_{i=1}^{N}\sum_{\kappa=0}^{2}\frac{h_j^{\kappa}f_{\kappa}(y_j)}{\sum_{\kappa=0}^{2}h_j^{\kappa}f_{\kappa}(y_j)}\frac{\partial}{\partial\Omega_-}\log f_{\kappa}(y_j)\\ &= \sum_{i=1}^{N}\sum_{\kappa=0}^{2} h_j^{\kappa}\frac{\partial}{\partial\Omega_-}\log f_{\kappa}(y_j),\end{aligned}$$

where we define

$$H_j^{\kappa} = \frac{h_j^{\kappa}f_{\kappa}(y_j)}{\sum_{\kappa=0}^{2}h_j^{\kappa}f_{\kappa}(y_j)}, \tag{B1}$$

which could be thought of as a posterior probability that progeny j have QTL genotype κ. We then implement the EM algorithm with the expanded parameter set $\{\Omega_-, \mathbf{H}\}$, where $\mathbf{H} = \{H_j^{\kappa}, \kappa = 0, 1, 2; j = 1, \ldots, N\}$. Conditional on $\mathbf{H}$, we solve for the zeros of $\partial/\partial\Omega_- \log L(\Omega_-)$ to get our estimates of Ω_- (the M step). In the M step, the quantitative genetic parameters, μ, α, β, and σ^2, of the QTL detected are solved using

$$\mu = \frac{1}{N}\sum_{j=1}^{N}\left[y_j - 2\alpha H_j^2 - (\alpha + \delta)H_j^1\right], \tag{B2}$$

$$\alpha = \frac{1}{\sum_{j=1}^{N}(4H_j^2 + H_j^1)} \times \sum_{j=1}^{N}\left[(2H_j^2 + H_j^1)(y_j - \mu) - \delta H_j^1\right], \tag{B3}$$

$$\delta = \frac{1}{\sum_{j=1}^{N}H_j^1} \times \sum_{j=1}^{N}(y_j - \mu - \alpha H_j^1), \tag{B4}$$

$$\sigma^2 = \frac{1}{N}\sum_{j=1}^{N}\left[(y_j - \mu - 2\alpha)^2H_j^2 + (y_j - \mu - \alpha - \delta)^2H_j^1 + (y_j - \mu)^2H_j^0\right]. \tag{B5}$$

The population genetic parameters $q_s^{(t)}$ and $D_{rs}^{(t)}$ and genomic parameter θ are estimated by using a numerical subroutine approach (Press *et al.* 1992) because closed forms for the solutions of these parameters cannot be derived. The estimates of these parameters are obtained by solving Equations B6–B8 in Scheme 1.

$$
\begin{aligned}
\frac{\partial}{\partial q_1^{(t)}}\left[\log \ell(\mathbf{y}, \mathbf{M}|q_1^{(t)})\right] = {} & \frac{1}{p_1^{2(t)}}\left\{\frac{1}{p_1^{(t)}}\left[(p_{11}^{(t)} + p_{11}^{(t+1)})F_{11}^{11(2)} + (p_{12}^{(t)} + p_{12}^{(t+1)} - p_{11}^{(t)} - p_{11}^{(t+1)})F_{11}^{11(1)} - (p_{12}^{(t)} + p_{12}^{(t+1)})F_{11}^{11(0)}\right]\right. \\
& \left. + \frac{1}{p_2^{(t)}}\left[(p_1^{(t)}p_{21}^{(t+1)} + p_2^{(t)}p_{11}^{(t)})F_{11}^{12(2)} + (p_1^{(t)}p_{22}^{(t+1)} + p_2^{(t)}p_{12}^{(t)} - p_1^{(t)}p_{21}^{(t+1)} - p_2^{(t)}p_{11}^{(t)})F_{11}^{12(1)} - (p_1^{(t)}p_{22}^{(t+1)} + p_2^{(t)}p_{12}^{(t)})F_{11}^{12(0)}\right]\right\} \\
& + \frac{1}{p_1^{(t)}p_2^{(t)}}\left\{\frac{1}{p_1^{(t)}}\left[\left(p_1^{(t)}[p_{11}^{(t)}p_{21}^{(t)} + (1-\theta)\phi] + p_{11}^{(t+1)}[p_1^{(t)}p_{21}^{(t)} + p_2^{(t)}p_{11}^{(t)} + (1-\theta)\phi]\right)F_{12}^{11(2)}\right.\right. \\
& \quad + \left(p_1^{(t)}[(p_{21}^{(t)}p_{22}^{(t)} - p_{11}^{(t)}p_{21}^{(t)}) - (1-2\theta)\phi] + p_1^{(t)}[p_{21}^{(t)}p_{12}^{(t+1)} - p_{22}^{(t)}p_{11}^{(t+1)}]\right. \\
& \quad\quad \left. + p_2^{(t)}[p_{11}^{(t)}p_{12}^{(t+1)} - p_{12}^{(t)}p_{11}^{(t+1)}] + [\theta p_{11}^{(t+1)} + (1-\theta)p_{12}^{(t+1)}]\varphi\right)F_{12}^{11(1)} \\
& \quad - \left(p_1^{(t)}[p_{12}^{(t)}p_{22}^{(t)} + \theta\phi] + p_{12}^{(t+1)}[p_1^{(t)}p_{22}^{(t)} + p_2^{(t)}p_{12}^{(t)} - \theta\varphi]\right)F_{12}^{11(0)}\Big] \\
& + \left[\left(p_{11}^{(t)}p_{21}^{(t)} + [p_{11}^{(t+1)}p_{21}^{(t+1)} + p_2^{(t)}p_{11}^{(t)}] + [\theta p_1^{(t)} + (1-\theta)p_2^{(t)}]\phi + [\theta p_{11}^{(t+1)} + (1-\theta)p_{21}^{(t+1)}]\varphi\right)F_{12}^{12(2)}\right. \\
& \quad + \left([p_{12}^{(t)}p_{22}^{(t)} - p_{11}^{(t)}p_{21}^{(t)}] + [p_{12}^{(t+1)} + p_{22}^{(t+1)}][p_1^{(t)}p_{21}^{(t)} + p_2^{(t)}p_{11}^{(t)}] - [p_{11}^{(t+1)} + p_{21}^{(t+1)}][p_1^{(t)}p_{22}^{(t)} + p_2^{(t)}p_{12}^{(t)}]\right. \\
& \quad\quad \left. + [1-2\theta][p_1^{(t)} - p_2^{(t)}]\phi + [\theta(p_{12}^{(t+1)} + p_{21}^{(t+1)}) + (1-\theta)(p_{11}^{(t+1)} + p_{22}^{(t+1)})]\varphi\right)F_{12}^{12(1)} \\
& \quad - \left(p_{12}^{(t)}p_{22}^{(t)} + [p_{12}^{(t+1)} + p_{22}^{(t+1)}]\phi + [\theta p_2^{(t)} + (1-\theta)p_1^{(t)}]\phi - [\theta p_{22}^{(t+1)} + (1-\theta)p_{12}^{(t+1)}]\varphi\right)F_{12}^{12(0)}\Big] \\
& + \frac{1}{p_2^{(t)}}\left[\left(p_2^{(t)}[p_{11}^{(t)}p_{21}^{(t)} + \theta\phi] + p_{21}^{(t+1)}[p_1^{(t)}p_{21}^{(t)} + p_2^{(t)}p_{11}^{(t)} + \theta\varphi]\right)F_{12}^{22(2)}\right. \\
& \quad + \left(p_2^{(t)}[(p_{11}^{(t)}p_{22}^{(t)} - p_{11}^{(t)}p_{21}^{(t)}) - (1-2\theta)\phi] + p_1^{(t)}(p_{21}^{(t)}p_{22}^{(t+1)} - p_{22}^{(t)}p_{21}^{(t+1)})\right. \\
& \quad\quad \left. + p_2^{(t)}(p_{11}^{(t)}p_{22}^{(t+1)} - p_{12}^{(t)}p_{21}^{(t+1)}) + [\theta p_{22}^{(t+1)} + (1-\theta)p_{21}^{(t+1)}]\varphi\right)F_{12}^{22(1)} \\
& \quad - \left.\left(p_2^{(t)}[p_{12}^{(t)}p_{22}^{(t)} + (1-\theta)\phi] + p_{22}^{(t+1)}[p_1^{(t)}p_{22}^{(t)} + p_2^{(t)}p_{12}^{(t)} - (1-\theta)\varphi]\right)F_{12}^{22(0)}\right]\Big\} \\
& + \frac{1}{p_2^{2(t)}}\left\{\frac{1}{p_1^{(t)}}\left[(p_1^{(t)}p_{21}^{(t)} + p_2^{(t)}p_{11}^{(t+1)})F_{22}^{12(2)} + (p_1^{(t)}p_{22}^{(t)} + p_2^{(t)}p_{12}^{(t+1)} - p_1^{(t)}p_{21}^{(t)} - p_2^{(t)}p_{11}^{(t+1)})F_{22}^{12(1)} - (p_1^{(t)}p_{22}^{(t)} + p_2^{(t)}p_{12}^{(t+1)})F_{22}^{22(0)}\right]\right. \\
& \left. + \frac{1}{p_2^{(t)}}\left[(p_{21}^{(t)} + p_{21}^{(t+1)})F_{22}^{22(2)} + (p_{22}^{(t)} + p_{22}^{(t+1)} - p_{21}^{(t)} - p_{21}^{(t+1)})F_{22}^{22(1)} - (p_{22}^{(t)} + p_{22}^{(t+1)})F_{22}^{22(0)}\right]\right\} \\
= {} & 0, \qquad \text{(B6)}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial}{\partial\theta}\left[\log \ell(\mathbf{y}, \mathbf{M}|\theta)\right] = {} & \frac{D_{rs}^{(t)}}{p_1^{2(t)}}\left\{\frac{1}{p_1^{(t)}}[p_{11}^{(t)}F_{11}^{11(2)} + (p_{12}^{(t)} - p_{11}^{(t)})F_{11}^{11(1)} - p_{12}^{(t)}F_{11}^{11(0)}] - \frac{1}{p_2^{(t)}}[p_{11}^{(t)}F_{11}^{12(2)} + (p_{12}^{(t)} - p_{11}^{(t)})F_{11}^{12(1)} - p_{12}^{(t)}F_{11}^{12(0)}]\right\} \\
& - \frac{1}{p_1^{(t)}p_2^{(t)}}\left\{\frac{1}{p_1^{(t)}}\left[\left(p_{11}^{(t)}p_{21}^{(t)}D_{rs}^{(t)} + [(1-\theta)D_{rs}^{(t)} + p_{11}^{(t+1)}]\phi\right)F_{12}^{11(2)} - \left([p_{11}^{(t)}p_{21}^{(t)} - p_{12}^{(t)}p_{22}^{(t)}]D_{rs}^{(t)} + [(1-2\theta)D_{rs}^{(t)} + p_{11}^{(t+1)} - p_{12}^{(t+1)}]\phi\right)F_{12}^{11(1)}\right.\right. \\
& \quad \left.\left. - \left(p_{12}^{(t)}p_{22}^{(t)}D_{rs}^{(t)} + [\theta D_{rs}^{(t)} + p_{12}^{(t+1)}]\phi\right)F_{12}^{11(0)}\right] \\
& \quad + \phi\left[\left([1-2\theta]D_{rs}^{(t)} + p_{11}^{(t+1)} - p_{21}^{(t+1)}\right)F_{12}^{12(2)} - \left(2[1-2\theta]D_{rs}^{(t)} + [p_1^{(t)} - p_2^{(t)}][q_1 - q_2]\right)F_{12}^{12(1)}\right. \\
& \quad\quad + \left([1-2\theta]D_{rs}^{(t)} + p_{22}^{(t+1)} - p_{12}^{(t+1)}\right)F_{12}^{12(0)}\Big] \\
& \quad\quad + \frac{1}{p_2^{(t)}}\left[\left(p_{11}^{(t)}p_{21}^{(t)}D_{rs}^{(t)} + [\theta D_{rs}^{(t)} + p_{21}^{(t+1)}]\phi\right)F_{12}^{22(2)}\right. \\
& \quad\quad\quad + \left([p_{12}^{(t)}p_{22}^{(t)} - p_{11}^{(t)}p_{21}^{(t)}]D_{rs}^{(t)} + [(1-2\theta)D_{rs}^{(t)} + p_{22}^{(t+1)} - p_{21}^{(t+1)}]\phi\right)F_{12}^{22(1)} \\
& \quad\quad \left. + \left(p_{12}^{(t)}p_{22}^{(t)}D_{rs}^{(t)} + [(1-\theta)D_{rs}^{(t)} + p_{22}^{(t+1)}]\phi\right)F_{12}^{22(0)}\right]\Big\} \\
& - \frac{D_{rs}^{(t)}}{p_2^{2(t)}}\left\{\frac{1}{p_1^{(t)}}\left[p_{21}^{(t)}F_{22}^{12(2)} - (p_{21}^{(t)} - p_{22}^{(t)})F_{22}^{12(1)} - p_{22}^{(t)}F_{22}^{12(0)}\right] - \frac{1}{p_2^{(t)}}\left[p_{21}^{(t)}F_{22}^{22(2)} - (p_{21}^{(t)} - p_{22}^{(t)})F_{22}^{22(1)} - p_{22}^{(t)}F_{22}^{22(2)}\right]\right\} \\
= {} & 0, \qquad \text{(B7)}
\end{aligned}
$$

Scheme 1.—Equations B6–B8.

and

$$\begin{aligned}
\frac{\partial}{\partial D_{rs}^{(t)}}\left[\log\,\ell(\mathbf{y},\,\mathbf{M}|D_{rs}^{(t)})\right] = {} & \frac{1}{p_1^{2(t)}}\left\{\frac{1}{p_1^{(t)}}\left[\left(p_{11}^{(t+1)}+[1-\theta]p_{11}^{(t)}\right)F_{11}^{11(2)}+\left(p_{12}^{(t+1)}-p_{11}^{(t+1)}+[1-\theta][p_{12}^{(t)}-p_{11}^{(t)}]\right)F_{11}^{11(1)}\right.\right.\\
& \left.-\left(p_{12}^{(t+1)}+[1-\theta]p_{12}^{(t)}\right)F_{11}^{11(0)}\right]\\
& +\frac{1}{p_2^{(t)}}\left[\left(p_{21}^{(t+1)}+[1-\theta]p_{11}^{(t)}\right)F_{11}^{12(2)}-\left(p_{21}^{(t+1)}+p_{22}^{(t+1)}+[1-\theta][p_{12}^{(t)}-p_{11}^{(t)}]\right)F_{11}^{12(1)}-\left(p_{22}^{(t+1)}+[1-\theta]p_{12}^{(t)}\right)F_{11}^{12(0)}\right]\Bigg\}\\
& +\frac{1}{p_1^{(t)}p_2^{(t)}}\left\{\frac{1}{p_1^{(t)}}\left[\left(p_{11}^{(t+1)}[p_{21}^{(t)}-p_{11}^{(t)}]+[1-\theta][p_{11}^{(t)}p_{21}^{(t)}+(1-\theta)\phi]+p_{11}^{(t+1)}(1-\theta)\psi\right)F_{12}^{11(2)}\right.\right.\\
& +\left(p_{11}^{(t+1)}[p_{12}^{(t)}-p_{22}^{(t)}]+p_{12}^{(t+1)}[p_{21}^{(t)}-p_{11}^{(t)}]+(1-\theta)[p_{12}^{(t)}p_{22}^{(t)}-p_{11}^{(t)}p_{21}^{(t)}-(1-2\theta)\phi]\right.\\
& \left.+[p_{11}^{(t+1)}\theta+p_{12}^{(t+1)}(1-\theta)]\psi\right)F_{12}^{11(1)}\\
& +\left(p_{12}^{(t+1)}[p_{12}^{(t)}-p_{22}^{(t)}]-[1-\theta][p_{12}^{(t)}p_{22}^{(t)}+\theta\phi]+p_{12}^{(t+1)}\theta\psi\right)F_{12}^{11(0)}\Big]\\
& +\left[\left([p_{21}^{(t)}-p_{11}^{(t)}][p_{11}^{(t+1)}+p_{21}^{(t+1)}]+[\theta p_{11}^{(t+1)}+(1-\theta)p_{21}^{(t+1)}]\psi-[1-\theta][1-2\theta]\phi\right)F_{12}^{12(2)}\right.\\
& +\left([p_{21}^{(t)}-p_{11}^{(t)}][p_{12}^{(t+1)}+p_{22}^{(t)}]+[p_{12}^{(t)}-p_{22}^{(t)}][p_{11}^{(t+1)}+p_{21}^{(t)}]+[\theta(p_{12}^{(t+1)}+p_{21}^{(t+1)})\right.\\
& \left.+(1-\theta)(p_{11}^{(t+1)}+p_{22}^{(t+1)})]+2(1-\theta)(1-2\theta)\phi\right)F_{12}^{12(1)}\\
& \left.+\left([p_{12}^{(t)}-p_{22}^{(t)}][p_{12}^{(t+1)}+p_{22}^{(t+1)}]+[\theta p_{22}^{(t+1)}+(1-\theta)p_{12}^{(t+1)}]\psi-[1-\theta][1-2\theta]\phi\right)F_{12}^{12(0)}\right]\\
& +\frac{1}{p_2^{(t)}}\left[\left(p_{21}^{(t+1)}[p_{21}^{(t)}-p_{11}^{(t)}]-[1-\theta][p_{11}^{(t)}p_{21}^{(t)}+\theta\phi]+p_{21}^{(t+1)}\theta\psi\right)F_{12}^{22(2)}\right.\\
& +\left(p_{21}^{(t+1)}[p_{12}^{(t)}-p_{22}^{(t)}]+p_{22}^{(t+1)}[p_{21}^{(t)}-p_{11}^{(t)}]+(1-\theta)[p_{11}^{(t)}p_{21}^{(t)}-p_{12}^{(t)}p_{22}^{(t)}-(1-2\theta)\phi]\right.\\
& \left.+[p_{22}^{(t+1)}\theta+p_{21}^{(t+1)}(1-\theta)]\psi\right)F_{12}^{22(1)}\\
& \left.+\left(p_{22}^{(t+1)}[p_{12}^{(t)}-p_{22}^{(t)}]+[1-\theta][p_{12}^{(t)}p_{22}^{(t)}+(1-\theta)\phi]+p_{22}^{(t+1)}(1-\theta)\psi\right)F_{12}^{22(0)}\right]\Bigg\}\\
& -\frac{1}{p_2^{3(t)}}\left\{\frac{1}{p_1^{(t)}}\left[\left(p_{11}^{(t+1)}-[1-\theta]p_{21}^{(t)}\right)F_{22}^{12(2)}+\left(p_{12}^{(t+1)}-p_{11}^{(t+1)}+[1-\theta][p_{21}^{(t)}-p_{22}^{(t)}]\right)F_{22}^{12(1)}\right.\right.\\
& \left.-\left(p_{12}^{(t+1)}-[1-\theta]p_{22}^{(t)}\right)F_{22}^{12(0)}\right]\\
& -\frac{1}{p_2^{(t)}}\left[\left(p_{21}^{(t+1)}+[1-\theta]p_{21}^{(t)}\right)F_{22}^{22(2)}-\left(p_{21}^{(t+1)}-p_{22}^{(t+1)}+[1-\theta][p_{21}^{(t)}-p_{22}^{(t)}]\right)F_{22}^{22(1)}\right.\\
& \left.\left.-\left(p_{22}^{(t+1)}+[1-\theta]p_{22}^{(t)}\right)F_{22}^{22(0)}\right]\right\}\\
= {} & 0, \qquad \text{(B8)}
\end{aligned}$$

where $F_{r_1r_2}^{\gamma_2\gamma_2(\kappa)}\,\Sigma N_{\substack{\gamma_1\gamma_2\\ r_1r_2i\\ j=1}}^{\gamma_1\gamma_2} = H_{r_1r_2(j)}^{\gamma_1\gamma_2(\kappa)}/h_{r_1r_2(j)}^{\gamma_1\gamma_2(\kappa)}$ for the κth QTL conditional on a two-level marker genotype with the subscripts and superscripts given by Equation 5, $\phi = p_{11}^{(t)}p_{22}^{(t)}+p_{12}^{(t)}p_{21}^{(t)}$, $\varphi = p_1^{(t)}p_{22}^{(t)}+p_2^{(t)}p_{12}^{(t)}-p_1^{(t)}p_{21}^{(t)}-p_2^{(t)}p_{11}^{(t)}$, and $\psi = p_{11}^{(t)}+p_{22}^{(t)}-p_{12}^{(t)}-p_{21}^{(t)}$.

Scheme 1.—*Continued.*

The estimates obtained from Equations B2–B8 in Scheme 1 are then used to update **H** (the E step). In the E step, the posterior probability of progeny *j* to have QTL genotype κ is calculated using Equation B1. The iteration between the E and M steps is repeated until convergence. The values at convergence are the MLEs of the parameters.