Mapping Quantitative Trait Loci by Genotyping Haploid Tissues

R. L. Wu

Forest Biotechnology Group, Department of Forestry, North Carolina State University, Raleigh, North Carolina 27695-8008

Manuscript received May 20, 1998
Accepted for publication April 15, 1999

ABSTRACT

Mapping strategies based on a half- or full-sib family design have been developed to map quantitative trait loci (QTL) for outcrossing species. However, these strategies are dependent on controlled crosses where marker-allelic frequency and linkage disequilibrium between the marker and QTL may limit their application. In this article, a maximum-likelihood method is developed to map QTL segregating in an open-pollinated progeny population using dominant markers derived from haploid tissues from single meiotic events. Results from the haploid-based mapping strategy are not influenced by the allelic frequencies of markers and their linkage disequilibria with QTL, because the probabilities of QTL genotypes conditional on marker genotypes of haploid tissues are independent of these population parameters. Parameter estimation and hypothesis testing are implemented via expectation/conditional maximization algorithm. Parameters estimated include the additive effect, the dominant effect, the population mean, the chromosomal location of the QTL in the interval, and the residual variance within the QTL genotypes, plus two population parameters, outcrossing rate and QTL-allelic frequency. Simulation experiments show that the accuracy and power of parameter estimates are affected by the magnitude of QTL effects, heritability levels of a trait, and sample sizes used. The application and limitation of the method are discussed.

CURRENT statistical methods for mapping quantitative trait loci (QTL) have been well developed based on controlled crosses (Lander and Botstein 1989; Knott and Haley 1992; Zeng 1993, 1994; Jansen and Stam 1994; Xu and Atchley 1995; Kao and Zeng 1997). By these methods, molecular markers derived from diploid tissues, such as leaves, buds, or root tips, are associated with the phenotypic traits of diploid tissues. Accurate mapping of QTL using these methods depends critically on well-defined mapping pedigrees, such as F_2, F_3, or backcrosses, initiated with two inbred lines. However, the development of such pedigrees is extremely difficult in outcrossing species, especially forest trees, due to their high heterozygosity, high genetic load, and long generation intervals (O'Malley 1996). The mapping strategy based on inbred lines, therefore, may not be appropriate for these species. New strategies based on half- or full-sib families derived from controlled crosses have been proposed for outcrossing species (Knott and Haley 1992; Mackinnon and Weller 1995; Hoeschele *et al.* 1997; Uimari and Hoeschele 1997; Liu and Dekkers 1998; Xu 1998). However, their application is limited in the case where the population frequencies of marker alleles are not correctly estimated (Mackinnon and Weller 1995) or where linkage disequilibria exist between the markers and QTL of interest. Here, I propose to develop an alternative strategy for mapping QTL with molecular markers in outcrossing species.

This approach is based on the molecular characterization of a haploid nongametic tissue that is derived from the same meiotic event as the gamete. In gymnosperms, such a haploid tissue occurs naturally and is called a megagametophyte (Bierhorst 1971). The megagametophyte, with the genotype identical to the maternal gamete, surrounds the embryo in the mature seed and supplies initial nutrients during seed to germination. Because the megagametophyte is genetically equivalent to a haploid progeny, any heterozygous locus in the seed parent will always segregate 1:1 in the megagametophytes (Wilcox *et al.* 1996) regardless of the pollen contribution, if segregation distortion does not occur. As a result, dominant markers derived from haploid megagametophytes are as informative as codominant markers. Isozymic analyses using the megagametophyte have been carried out in gymnosperms for many years for estimation of genetic diversity, heterozygosity, and genetic relatedness and for studies of gene flow in natural populations (Wheeler and Guries 1982; Millar 1983; Hamrick *et al.* 1992; Huang *et al.* 1994; Rogers 1997). More recently, attempts have also been made to employ the megagametophyte to construct genetic linkage maps by collecting PCR-based dominant markers from the progeny of a heterozygous tree. A number of coniferous species have been mapped using the megagametophyte, and they include *Pinus taeda* (Grattapaglia *et al.* 1991; Wilcox *et al.* 1996; Jordan 1997),

Address for correspondence: Program in Statistical Genetics, Department of Statistics, Box 8203, North Carolina State University, Raleigh, NC 27695-8203. E-mail: rwu@statgen.ncsu.edu

P. sylvestris (Yazdani *et al.* 1995), *P. pinaster* (Plomion *et al.* 1995), *P. elliotii* (Nelson *et al.* 1993), *P. radiata* (Dale 1994), *P. massoniana* (Yan 1997), *Picea glauca* (Tulserium *et al.* 1992), and *P. abies* (Benelli and Bucci 1994).

Because the megagametophyte includes only a half of the offspring's genetic information, the strategy of QTL mapping using the megagametophyte should be based on statistical inferences about the other, unknown half from the paternal gamete. Many statistical methods have been suggested to map QTL affecting a quantitative trait in a segregating progeny population. Hoeschele *et al.* (1997) classified these methods into six groups. Group 1 includes linear regression analysis using a single or multiple linked markers. Group 2 includes maximum-likelihood (ML) analysis of a postulated biallelic QTL based on a single or multiple linked markers. Group 3 includes regression of squared phenotypic differences of pairs of relatives on the expected proportion of identity-by-descent at a locus. Group 4 includes residual (or restricted) ML analysis based on a mixed linear model incorporating normally distributed QTL allelic effects with a covariance matrix conditional on observed marker data. Group 5 includes exact Bayesian linkage analysis using single or multiple linked markers and fitting biallelic or infinite-allele QTL. Finally, group 6 includes an approximate Bayesian analysis of a postulated biallelic QTL. These methods differ in their computational requirements and statistical power. Although groups 1, 3, 4, and 6 are computationally inexpensive, these methods are less suited for genetic parameter estimation in outcrossing populations. ML and Bayesian analysis are the computationally most demanding methods but take account of the distribution of multilocus marker-QTL genotypes and permit investigators to fit different models of variation at the QTL.

In this article, I develop a statistical method to map QTL based on haploid tissues using ML. Lander and Botstein (1989) used ML to map a putative QTL lying in the interval bracketed by two flanking markers. This mapping method was further developed by Zeng (1993, 1994), who combined the principle of interval mapping and multiple regression analysis. Zeng's method, called composite interval mapping (CIM), can effectively reduce the influence of linked QTL on parameter estimation by controlling the genetic background outside a given interval. CIM has been applied to identify QTL in mice (Dragani *et al.* 1995) and Drosophila (Liu *et al.* 1996; Nuzhdin *et al.* 1997). It has also been modified to be more broadly useful for several particular circumstances, such as outbred human families (Xu and Atchley 1995) and four-way cross populations (Xu 1996). My method is developed within the framework of CIM.

THEORY

CIM: Consider an individual that is heterozygous at both molecular markers and QTL of interest in a random mating population. The open-pollinated progeny from this heterozygote will establish a mapping population. If the mapping material is a monoecious plant species, such as a conifer, the heterozygote may be pollinated by its own pollen and other unrelated plants' pollen. Thus, seeds collected from the mother tree include those from both selfing and outcrossing pollination. During selfing pollination, maternal and paternal gametes combine to form selfed seeds from the same tree; both kinds of gametes may be assumed to have the identical frequencies, which are dependent on the recombination frequencies between a given set of loci (Table 1). However, for the outcrossed progeny, although maternal gametes are the same as those for the selfed progeny, paternal gametes come from the natural population (excluding the mother tree) and their frequencies are determined by allelic frequencies at individual loci and the gametic phase disequilibria between the loci (Weir 1996). Because the markers derived from haploid tissues of open-pollinated seeds are used to identify QTL underlying a quantitative trait and because the genotype of each haploid is represented by the maternal gamete of the mother tree, it is necessary to define the probabilities of the QTL genotypes conditional on the maternal marker gametes. Assume that a putative QTL, **Q**, is located between two flanking markers, $\mathbf{M}_t$ and $\mathbf{M}_{t+1}$, with the recombination frequency of r, on a chromosome with m ordered marker loci (and therefore $m - 1$ intervals), and that the recombination frequencies between the QTL and these two markers are r_t and r_{t+1}, respectively. I use the ratio (θ) of r_t to r to describe the position of the QTL in the interval. The probabilities of QTL genotypes conditional on each of the four marker gametes of $\mathbf{M}_t$ and $\mathbf{M}_{t+1}$ are given in Table 1, separately for the selfed and outcrossed progenies. For example, in the outcrossed progeny, the conditional probability of QQ upon maternal gamete M_tM_{t+1} is given by

$$p^{\mathrm{O}}(QQ/M_tM_{t+1}) = \frac{p^{\mathrm{O}}(M_{t-t}QQM_{(t+1)-(t+1)})}{p^{\mathrm{O}}(M_tM_{t+1})},$$

where the uppercase superscript O indicates the outcrossed progeny (similarly, the selfed progeny is denoted by the uppercase superscript S; see Table 1), the denominator is the probability of an individual in the outcrossed progeny carrying maternal gamete M_tM_{t+1}, which is $\frac{1}{2}(1 - r)$, and the numerator is the probability of the individual carrying maternal gamete M_tQM_{t+1} and paternal gamete $_tQ_{t+1}$ (underscores denote alternative marker alleles M or m),

$$\begin{aligned} p^{\mathrm{O}}(M_{t-t}QQM_{(t+1)-(t+1)}) &= p^{\mathrm{m}}(M_tQM_{t+1})p^{\mathrm{p}}(_tQ_{t+1}) \\ &= \tfrac{1}{2}(1 - r_t)(1 - r_{t+1}) \\ &\quad \times [\{uwv + wD_{M_tM_{t+1}} + uD_{QM_{t+1}} \\ &\quad + vD_{M_tQ} + D_{M_tQM_{t+1}}\} \end{aligned}$$

TABLE 1

Conditional probability of QTL genotypes given the maternal gamete of the flanking markers ($\mathbf{M}_t - \mathbf{M}_{t+1}$) in the selfed and outcrossed progenies of a heterozygous individual

			Selfed			Outcrossed		
Marker genotype	Frequency	Sample size	p^{S}_{2j} QQ	p^{S}_{1j} Qq	p^{S}_{0j} qq	p^{O}_{2j} QQ	p^{O}_{1j} Qq	p^{O}_{0j} qq
M_tM_{t+1}	$\frac{1}{2}(1-r)$	n_1	$\frac{1}{2}$	$\frac{1}{2}$	0	w	$1-w$	0
M_tm_{t+1}	$\frac{1}{2}r$	n_2	$\frac{1}{2}(1-\theta)$	$\frac{1}{2}$	$\frac{1}{2}\theta$	$(1-\theta)w$	$1-\theta-w+2\theta w$	$\theta(1-w)$
m_tM_{t+1}	$\frac{1}{2}r$	n_3	$\frac{1}{2}\theta$	$\frac{1}{2}$	$\frac{1}{2}(1-\theta)$	θw	$\theta+w-2\theta w$	$(1-\theta)(1-w)$
m_tm_{t+1}	$\frac{1}{2}(1-r)$	n_4	0	$\frac{1}{2}$	$\frac{1}{2}$	0	W	$1-w$

θ describes the position of the putative QTL in the $\mathbf{M}_t - \mathbf{M}_{t+1}$ interval and can be treated either as a parameter or as a constant with $\theta = r_t/r$, where r_t is the recombination frequency between the QTL and marker $\mathbf{M}_t$ and r is the recombination frequency between the two flanking markers. Double recombination within the marker interval is ignored.

$$
\begin{aligned}
&+ \{uw(1-v) - wD_{M_tM_{t+1}} - uD_{QM_{t+1}} \\
&+ (1-v)D_{M_tQ} - D_{M_tQM_{t+1}}\} \\
&+ \{(1-u)wv - wD_{M_tM_{t+1}} \\
&+ (1-u)D_{QM_{t+1}} - vD_{M_tQ} - D_{M_tQM_{t+1}}\} \\
&+ \{(1-u)w(1-v) + wD_{M_tM_{t+1}} \\
&- (1-u)D_{QM_{t+1}} - (1-v)D_{M_tQ} \\
&+ D_{M_tQM_{t+1}}\}] \\
&= \tfrac{1}{2}(1-r)w,
\end{aligned}
$$

where double crossovers are ignored, $p^{m}(\cdot)$ and $p^{p}(\cdot)$ are the population frequencies of maternal and paternal gametes, respectively; u, w, and v are the population frequencies of alleles M_t, Q, and M_{t+1}, respectively; $D_{M_tM_{t+1}}$, $D_{QM_{t+1}}$, and D_{M_tQ} are the gametic linkage disequilibria between loci $\mathbf{M}_t$ and $\mathbf{M}_{t+1}$, $\mathbf{Q}$ and $\mathbf{M}_{t+1}$, and $\mathbf{M}_t$ and $\mathbf{Q}$, respectively; and $D_{M_tQM_{t+1}}$ is the gametic linkage disequilibrium among these three loci (Weir 1996). It is shown that the conditional probabilities in the outcrossed progeny are determined by allelic frequencies at the QTL and the linkage between the QTL and marker loci, but are independent of allelic frequencies for the markers and linkage disequilibria between the markers and QTL in the pollen pool (see Table 1).

Assuming that no epistasis exists between loci, the phenotype of the jth individual from the open-pollinated progeny (of size n) of the heterozygous maternal parent can be expressed in terms of the QTL located in the interval of two adjacent markers $\mathbf{M}_t$ and $\mathbf{M}_{t+1}$,

$$
y_j = \mu + ax_j^* + dz_j^* + \sum_{k\neq t,t+1}^{m} b_k x_{kj} + \varepsilon_j, \qquad (1)
$$

where μ is the overall mean, a and d are the additive and dominant effects of the putative QTL, respectively, and x_j^* and z_j^* are the indicator variables of the jth individual whose values are taken as

$$
x_j^* = \begin{cases} 2 & \text{if the QTL genotype is } QQ \\ 1 & \text{if the QTL genotype is } Qq \\ 0 & \text{if the QTL genotype is } qq \end{cases}
$$

$$
z_j^* = \begin{cases} 1 & \text{if the QTL genotype is } Qq \\ 0 & \text{if the QTL genotype is } QQ \text{ or } qq. \end{cases}
$$

b_k is the partial regression coefficient of the phenotype y on the kth marker conditional on all other markers, x_{kj} is the known indicator variable of the kth marker in the jth individual, taking the value of 1 or 0 depending on the type of marker allele from a maternal gamete, and ε_j is a random variable, $\varepsilon_j \sim N(\mu, \sigma^2)$. The variance, σ^2, includes both environmental variation and genetic variation at other loci affecting the quantitative trait but segregating independently of the QTL under consideration. If the probability with which a maternal gamete receives pollen from unrelated individuals, *i.e.*, outcrossing rate, is denoted by ρ, the likelihood function of the quantitative effect for a mixed selfed and outcrossed progeny population (of size n) is expressed by

$$
L = \prod_{j=1}^{n}\left\{\sum_{i=0}^{2}[(1-\rho)p_{ij}^{S} + \rho p_{ij}^{O}]\, f_i(y_j)\right\}, \qquad (2)
$$

where p_{ij}^{S} and p_{ij}^{O} are the prior probabilities of the jth individual taking $x_j^* = i$ (representing the ith QTL genotype characterized by the number of Q alleles) for the selfed and outcrossed progenies, respectively, and $f_i(y_j)$ is the density function of the phenotype of the jth individual with QTL genotype i:

$$
f_i(y_j) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(y_j-\mu_i)^2}{2\sigma^2}\right]
$$

$$
\begin{aligned}
\mu_2 &= \mathbf{x}_j\mathbf{b} + 2a \\
\mu_1 &= \mathbf{x}_j\mathbf{b} + a + d \\
\mu_0 &= \mathbf{x}_j\mathbf{b} \\
\mathbf{x}_j\mathbf{b} &= \mu + \sum_{k\neq t,t+1}^{m} b_k x_{kj}.
\end{aligned}
$$

By differentiating the likelihood function (2) with re-

spect to each of the unknown parameters, a, d, $\mathbf{b}$, and σ^2, setting the derivatives equal to zero, and then solving the equations, the ML estimates of these parameters can be obtained as

$$\hat{a} = (\mathbf{y} - \mathbf{X}\hat{\mathbf{b}})'\hat{\mathbf{P}}_2/(2\mathbf{1}'\hat{\mathbf{P}}_2) \quad (3)$$

$$\hat{d} = (\mathbf{y} - \mathbf{X}\hat{\mathbf{b}})'\hat{\mathbf{P}}_1/(\mathbf{1}'\hat{\mathbf{P}}_1) - \hat{a} \quad (4)$$

$$\hat{\mathbf{b}} = (\mathbf{X}'\mathbf{X})^{-1}[\mathbf{y} - (2\hat{\mathbf{P}}_2 + \hat{\mathbf{P}}_1)\hat{a} - \hat{\mathbf{P}}_1\hat{d}] \quad (5)$$

$$\hat{\sigma}^2 = \tfrac{1}{n}[(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}})'(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) - 4(\mathbf{1}'\hat{\mathbf{P}}_2)\hat{a}^2 - (\mathbf{1}'\hat{\mathbf{P}}_1)(\hat{a} + \hat{d})^2], \quad (6)$$

where $\mathbf{y}$ is a $(n \times 1)$ vector of y_j's, $\hat{\mathbf{b}}$ is a $[(m-2) \times 1]$ vector of the ML estimates of b_k's, $\mathbf{X}$ is an $[n \times (m-2)]$ matrix of x_{jk}'s, and $\hat{\mathbf{P}}_1$ and $\hat{\mathbf{P}}_2$ are $(n \times 1)$ vectors with elements $\hat{P}_{1j}$ and $\hat{P}_{2j}$ specifying the ML estimate of the posterior probability of $x_j^* = 2$ and 1 (Zeng 1994), respectively:

$$\hat{P}_{2j} = \frac{[(1-\rho)p_{2j}^{S} + \rho p_{2j}^{O}]\,\hat{f}_2(y_j)}{\sum_{i=0}^{2}[(1-\rho)p_{ij}^{S} + \rho p_{ij}^{O}]\,\hat{f}_i(y_j)},$$

$$\hat{P}_{1j} = \frac{[(1-\rho)p_{1j}^{S} + \rho p_{1j}^{O}]\,\hat{f}_1(y_j)}{\sum_{i=0}^{2}[(1-\rho)p_{ij}^{S} + \rho p_{ij}^{O}]\,\hat{f}_i(y_j)}.$$

Similarly, the ML estimates for outcrossing rate, ρ, and the frequency of QTL allele in the pollen pool, w, are given by

$$\hat{\rho} = (\textstyle\sum_{j=1}^{n_2+n_3}\hat{F}_{1j} + \sum_{j=1}^{n_2}\hat{F}_{2j} + \sum_{j=1}^{n_3}\hat{F}_{0j} + \theta[\sum_{j=1}^{n_2}(\hat{F}_{1j} - \hat{F}_{2j}) - \sum_{j=1}^{n_3}(\hat{F}_{0j} - \hat{F}_{1j})] - 2(n_2 + n_3))/((1-2w)[\sum_{j=1}^{n_1}(\hat{F}_{1j} - \hat{F}_{2j}) + \sum_{j=1}^{n_4}(\hat{F}_{0j} - \hat{F}_{1j})]) \quad (7)$$

$$\hat{w} = (\textstyle\sum_{j=1}^{n_1+n_2+n_3}\hat{P}_{2j} + \sum_{j=1}^{n_2+n_3+n_4}\hat{P}_{1j} + (1-\rho)\sum_{j=1}^{n_1+n_2+n_3}\hat{F}_{2j} - (1-\rho\theta)\sum_{j=1}^{n_2}\hat{F}_{1j} - (1-\rho+\rho\theta)\sum_{j=1}^{n_3}\hat{F}_{1j} - (1-\rho) \times \sum_{j=1}^{n_4}\hat{F}_{1j})/(n_2 + n_3 + n_4 + \sum_{j=1}^{n_1}(\hat{P}_{2j} + \hat{P}_{1j}) - \sum_{j=1}^{n_4}\hat{P}_{2j} - (1-\rho)\sum_{j=1}^{n_1}(\hat{F}_{2j} + \hat{F}_{1j}) - \sum_{j=1}^{n_2}[(1-\rho)\hat{F}_{2j} + (1-\rho\theta)\hat{F}_{1j} + (1-\rho)\theta\hat{F}_{0j}] - \sum_{j=1}^{n_3}[(1-\rho)\hat{F}_{2j} + (1-\rho+\tau\theta)\hat{F}_{1j} + (1-\rho)(1-\theta)\hat{F}_{0j}] - (1-\rho)\sum_{j=1}^{n_4}(\hat{F}_{1j} + \hat{F}_{0j})), \quad (8)$$

where

$$\hat{F}_{2j} = \frac{\hat{f}_2(y_j)}{\sum_{i=0}^{2}[(1-\rho)p_{ij}^{S} + \rho p_{ij}^{O}]\,\hat{f}_i(y_j)},$$

$$\hat{F}_{1j} = \frac{\hat{f}_1(y_j)}{\sum_{i=0}^{2}[(1-\rho)p_{ij}^{S} + \rho p_{ij}^{O}]\,\hat{f}_i(y_j)},$$

$$\hat{F}_{0j} = \frac{\hat{f}_0(y_j)}{\sum_{i=0}^{2}[(1-\rho)p_{ij}^{S} + \rho p_{ij}^{O}]\,\hat{f}_i(y_j)},$$

The parameter describing the position of the QTL, θ, can be treated as either a parameter or a constant. If it is a parameter, then its ML estimate is the solution of

$$\hat{\theta} = (2\textstyle\sum_{j=1}^{n_2}(1-\hat{P}_{2j}) + 2\sum_{j=1}^{n_3}(\hat{P}_{1j} + \hat{P}_{2j}) - [(1+\rho-2\rho w)\sum_{j=1}^{n_2}\hat{F}_{1j} + (1-\rho+2\rho w) \times \sum_{j=1}^{n_3}\hat{F}_{1j}])/(2(n_2 + n_3) - [(1-\rho-2\rho w)\sum_{j=1}^{n_2}\hat{F}_{1j} + (1-\rho+2\rho w)\sum_{j=1}^{n_3}\hat{F}_{1j}]). \quad (9)$$

The solutions of the unknown parameters are not in closed form, and each estimate depends on estimates of other parameters. Zeng (1994) suggested that the expectation/conditional maximization (ECM) algorithm, developed by Meng and Rubin (1993), could be used to find the ML estimates of these parameters by iterating the above equations beginning with the initial estimates $\hat{a}$, $\hat{d} = 0$ or the least-squares estimates of a, d, and $\mathbf{b}$ using x_j^*, $z_j^* = p_{1j}^{S}$, or p_{1j}^{O}.

Formulation of hypothesis: The null hypotheses about the additive (a) and dominant effects (d) of the QTL can be tested with χ^2 statistics. The likelihood function under the null hypothesis can be calculated by substituting the expressions of this null hypothesis into Equation 2. The hypothesis for testing the presence of a putative QTL in the interval is H_0, $a = d = 0$ *vs.* H_1, at least one parameter $\neq 0$. The likelihood function under the null hypothesis is given by

$$L(a = d = 0, \mathbf{b}, \sigma^2) = \prod_{j=1}^{n} f(y_j), \quad (10)$$

where $f(y_j) = (1/\sqrt{2\pi}\sigma)\exp[-(y_j - \mathbf{X}_j\mathbf{b})^2/2\sigma^2]$. The ML estimates for $\mathbf{b}$ and σ^2 under the null hypothesis are given by

$$\tilde{\mathbf{b}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

$$\tilde{\sigma}^2 = \tfrac{1}{n}(\mathbf{y} - \mathbf{X}\tilde{\mathbf{b}})'(\mathbf{y} - \mathbf{X}\tilde{\mathbf{b}}).$$

The test statistic is estimated as the log-likelihood ratio (LR) of Equations 10 and 2,

$$\mathrm{LR} = -2\ln\left[\frac{L(a=0, d=0, \hat{\mathbf{b}}, \hat{\sigma}^2)}{L(\hat{a}, \hat{d}, \hat{\mathbf{b}}, \hat{\sigma}^2, \hat{w})}\right], \quad (11)$$

which follows asymptotically a chi-square distribution.

SIMULATION

Simulation studies are carried out to illustrate the properties of CIM modified to map QTL using haploid tissues from a heterozygous individual. For a detailed discussion about the behavior of the test statistic of the ML method and advantages and disadvantages of CIM based on a controlled cross, see Zeng (1994).

Test statistic under the null hypothesis: The statistical behavior of the method proposed under a series of realistic conditions is examined by simulation experiments. Consider a species in which a haploid tissue derived from the same meiotic event as a gamete can be currently genotyped. Conifers, with 12 pairs of chromosomes, represent a significant group of such species. Assume that a genomic size of 2400 cM is composed of 12 chromosomes with identical lengths. On each chromosome, 11 markers are situated 20-cM apart from their immediate neighbors and cover the entire chromosome of 200-cM length. Based on different objectives of simulation experiments (see below), 300–1000 progeny individuals are simulated each with maternal gametic genotype M_t or m_t at the tth marker. By simulating a normally distributed quantitative trait on these individuals, the maximum LR test statistic (*i.e.*, θ is treated as a parameter) is calculated by (11) throughout a single chromosome for each of 1000 replicated simulations. The 95th percentiles for the simulated test statistics over the 1000 replicates are used as the critical values to declare the existence of a QTL in the chromosomes. Because outcrossing rate and QTL-allelic frequency may affect the accuracy of parameter estimation, I first assume that they are fixed in the simulations by setting $\rho = 0.90$ and $w = 0.50$. The choice of $\rho = 0.90$ is based on empirical observations on outcrossing rate in coniferous populations: for example, $\rho = 0.80$–0.90 for *P. caribaea* (Zheng and Ennos 1997) and $\rho = 0.89$–0.97 for *P. attenuata* (Burczyk *et al.* 1997).

Experimental designs: Three types of simulation experiments are performed to explore how differences in genetic architecture, sample size, heritability level, and parental-population composition affect the accuracy and power of parameter estimation. Experiment 1 is based on a genetic model in which some underlying QTL have larger effects on the phenotype than others (nonpolygenic model). Variable effects of QTL have been experimentally found in many species that are subject to QTL mapping (reviewed by Wu *et al.* 1999). Experiment 2 is based on a polygenic model that assumes a large number of loci with small effects. The polygenic model forms the basis of quantitative genetics theories (Bulmer 1980; Falconer and Mackay 1996) that have been applied successfully to genetic improvement of many species, such as forest trees (Wu *et al.* 1999). The strategy of QTL mapping based on a heterozygous individual is powerful because one can estimate two important population genetic parameters: outcrossing rate and QTL-allelic frequency. However, an issue arises about how the estimates of these additional parameters influence the statistical behavior of the method. Thus, I perform experiment 3 to explore the influences of different outcrossing rates and QTL-allelic frequencies on parameter estimation.

In experiments 1 and 2, I assume that a quantitative trait is controlled by 14 QTL with uneven distributions on the chromosomes. For example, chromosome 9 has three QTL, whereas chromosomes 4, 6, 8, 10, and 12 have no QTL at all. In experiment 1, a nonpolygenic trait is assumed in which the effects of the simulated QTL vary over loci (Tables 2 and 3). The statistical behavior of the method is examined under two different broad-sense heritability levels ($H^2 = 0.60$ *vs.* 0.20) as well as two different sample sizes ($n = 800$ *vs.* 300). Experiment 2 assumes the same QTL locations but in which each QTL has a similar, small effect (polygenic trait, Table 4). In this experiment, assume broad-sense heritability $H^2 = 0.20$ and sample size $n = 1000$. Experiment 3 assumes a QTL located at 50 cM in a 200-cM-long chromosome uniformly covered by 11 markers. The additive and dominant effects of the QTL on a simulated trait are set at 1.2 and 1.0, respectively. The broad-sense heritability of the trait is 0.60. Simulations are repeated 100 times on 400 individuals.

In all three experiments, the trait value of an individual is determined by the sum of additive and dominant effects of the simulated QTL plus a random variable that is normally distributed with mean zero and variance scaled to give the expected heritabilities. Simulations were repeated 100 times to estimate the average values and sampling errors of QTL parameters. The statistical power of a test is the probability of detecting the effect of the QTL when it exists. The empirical power was estimated from the 100 repeated simulations. It has been shown that the estimation of parameters using CIM is affected by the number, type, and space of markers used as cofactors in the multiple linear regression analysis (Equation 1; Zeng 1994). In this study, instead of markers throughout the entire genome, markers are chosen as cofactors only throughout a single chromosome on which the stimulated QTL are located. Such a scheme to choose markers is based on the fact that, when including too many markers (regardless of ones linked or unlinked to the QTL), the power of CIM would be largely reduced (Broman 1997).

Results: In experiment 1, QTL of larger effects can be detected more easily than those of small effects. However, the precision and power of parameter estimates are strongly affected by heritability levels and sample sizes (Tables 2 and 3). If a large sample size ($n = 800$) is used, the method can detect 86% (12/14) of the simulated QTL for a trait of high heritability ($H^2 = 0.60$; Table 2). Also, as indicated by low sampling errors, the method can precisely estimate the positions and additive and dominant effects of these QTL, even those

TABLE 2

Experiment 1: Average values (±SE) and empirical power of the ML estimates of QTL affecting a nonpolygenic trait of different heritability level

Chromosome	Simulated value			Estimated value											
				$H^2 = 0.60$						$H^2 = 0.20$					
	Position	a	d	Position	$\hat{a}$	$\hat{d}$	Power	$\hat{\rho}$	$\hat{w}$	Position	$\hat{a}$	$\hat{d}$	Power	$\hat{\rho}$	$\hat{w}$
1	32	0.5	0.4	35 ± 4	0.6 ± 0.03		0.41	0.89 ± 0.10	0.47 ± 0.07						
	175	1.2	0.8	174 ± 8	1.3 ± 0.08	0.9 ± 0.05	0.88	0.89 ± 0.09	0.49 ± 0.05	174 ± 25	1.3 ± 0.33		0.61	0.88 ± 0.09	0.48 ± 0.14
2	19	1.0	0.2	18 ± 1	1.2 ± 0.09		0.80	0.90 ± 0.10	0.48 ± 0.05	22 ± 7	1.1 ± 0.27		0.50	0.89 ± 0.10	0.47 ± 0.15
3	15	−0.9	−0.4	19 ± 2	−0.8 ± 0.04		0.62	0.89 ± 0.11	0.47 ± 0.06						
	134	−0.2	0.2												
	186	1.0	0.8	183 ± 5	1.1 ± 0.07	0.8 ± 0.04	0.83	0.89 ± 0.09	0.48 ± 0.06	189 ± 27	1.1 ± 0.32		0.55	0.89 ± 0.10	0.48 ± 0.16
4	—	—	—												
5	48	0.5	0.2	45 ± 5	0.6 ± 0.04		0.32	0.89 ± 0.10	0.46 ± 0.07						
	150	0.3	0.5	153 ± 9		0.5 ± 0.03	0.33	0.89 ± 0.09	0.46 ± 0.08						
	193	0.2	0.3												
6	—	—	—												
7	110	0.6	1.2	113 ± 4	0.7 ± 0.04	1.3 ± 0.11	0.80	0.89 ± 0.09	0.49 ± 0.06	114 ± 28		1.3 ± 0.37	0.52	0.89 ± 0.08	0.47 ± 0.13
8	—	—	—												
9	126	1.0	1.0	125 ± 6	0.9 ± 0.05	1.1 ± 0.08	0.98	0.89 ± 0.10	0.49 ± 0.05	128 ± 26	0.9 ± 0.24	1.1 ± 0.25	0.72	0.90 ± 0.11	0.48 ± 0.14
	155	1.2	0.8	156 ± 7	1.2 ± 0.10	0.8 ± 0.04	1.00	0.88 ± 0.08	0.49 ± 0.06	157 ± 27	1.3 ± 0.30		0.77	0.89 ± 0.10	0.47 ± 0.16
	191	1.0	1.4	188 ± 9	1.1 ± 0.10	1.3 ± 0.09	1.00	0.89 ± 0.09	0.50 ± 0.05	194 ± 34	1.0 ± 0.34	1.3 ± 0.30	0.80	0.87 ± 0.09	0.49 ± 0.16
10	—	—	—												
11	29	0.3	0.5	28 ± 4		0.7 ± 0.4	0.34	0.88 ± 0.08	0.47 ± 0.08						
12	—	—	—												

The sample size is $n = 800$. Results are obtained from 100 replicated simulations.

TABLE 3

Experiment 1: Average values (±SE) and empirical power of the ML estimates of QTL affecting a nonpolygenic trait of different heritability levels

Chromosome	Simulated value			Estimated value											
				$H^2 = 0.60$						$H^2 = 0.20$					
	Position	a	d	Position	$\hat{a}$	$\hat{d}$	Power	$\hat{\rho}$	$\hat{w}$	Position	$\hat{a}$	$\hat{d}$	Power	$\hat{\rho}$	$\hat{w}$
1	32	0.5	0.4												
	175	1.2	0.8	179 ± 24	1.4 ± 0.30		0.69	0.88 ± 0.11	0.46 ± 0.15	179 ± 31	1.5 ± 0.42		0.32	0.87 ± 0.13	0.46 ± 0.17
2	19	1.0	0.2	22 ± 6	1.5 ± 0.23		0.53	0.87 ± 0.13	0.47 ± 0.17						
3	15	−0.9	−0.4	11 ± 4	−1.2 ± 0.17		0.30	0.89 ± 0.11	0.47 ± 0.16						
	134	−0.2	0.2												
	186	1.0	0.8	189 ± 27	1.2 ± 0.29		0.60	0.88 ± 0.12	0.48 ± 0.16						
4		—	—												
5	48	0.5	0.2												
	150	0.3	0.5												
	193	0.2	0.3												
6	—	—	—												
7	110	0.6	1.2	107 ± 25		1.5 ± 0.34	0.54	0.87 ± 0.11	0.45 ± 0.14	115 ± 37		1.6 ± 0.42	0.42	0.89 ± 0.13	0.46 ± 0.18
8	—	—	—												
9	126	1.0	1.0	128 ± 22	1.2 ± 0.21	1.3 ± 0.20	0.79	0.87 ± 0.13	0.47 ± 0.15						
	155	1.2	0.8	157 ± 23	1.3 ± 0.28		0.81	0.87 ± 0.12	0.49 ± 0.14						
	191	1.0	1.4	195 ± 32	1.3 ± 0.31	1.6 ± 0.23	0.82	0.89 ± 0.12	0.48 ± 0.14	195 ± 44	1.3 ± 0.38	1.7 ± 0.35	0.62	0.88 ± 0.14	0.47 ± 0.18
10	—	—	—												
11	29	0.3	0.5												
12	—	—	—												

The sample size is $n = 300$. Results are obtained from 100 replicated simulations.

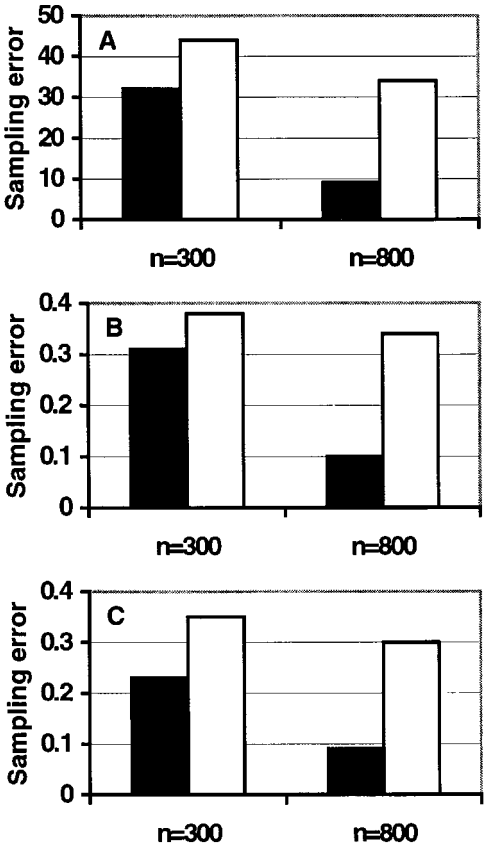

Figure 1.—Interaction effects of sample sizes and heritabilities (solid bars, $H^2 = 0.60$ and open bars, $H^2 = 0.20$) on sampling errors of parameter estimate for the QTL of large effect at 191 cM from the top of chromosome 9. A, QTL position; B, QTL-additive effect; and C, QTL-dominant effect.

with relatively small effects. When both sample size and heritability are large, the statistical power to detect the small QTL is ~0.30–0.40 but the power to detect those QTL of large effects can be >0.90. Both estimation precision and power are largely reduced when the trait has a low heritability (Table 2) or when the sample size used is small (Table 3). In the case where the heritability of a trait is low ($H^2 = 0.20$) but the sample size used is large ($n = 800$), the method can detect 50% (7/14) of the assumed QTL. If the trait's heritability is large ($H^2 = 0.60$) but the sample size used is low ($n = 300$), 57% (8/14) of the assumed QTL can be estimated. If both the heritability and sample size are low, the method can only estimate the three largest QTL (21%) with low power (0.32–0.62; Table 3).

It is found that estimates of QTL positions and effects are affected by interactions between heritabilities and sample sizes. Figure 1 describes the sampling errors of parameter estimation for a QTL of large effect at 191 cM from the top of chromosome 9 under different heritability and sample size combinations. A similar trend is detected for the estimation of QTL position (Figure 1A) and QTL effects (Figure 1, B and C). Parameter estimation displays greater precision under a sample

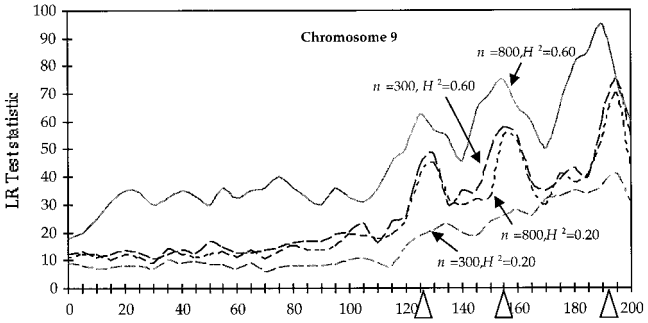

Figure 2.—The profile of likelihood-ratio test statistic calculated at every 1-cM position of chromosome 9. Three simulated QTL are indicated by triangles. Results are drawn from a single simulation experiment of each of the four different combinations between two sample sizes ($n = 800$ and 300) and broad-sense heritabilities ($H^2 = 0.60$ and 0.20). Only when both sample size and heritability are small, the three simulated QTL cannot be well separated.

size of 800 than 300. However, heritability of $H^2 = 0.60$ produces a much more significant increase in estimation precision than does heritability of $H^2 = 0.20$. If the sample size used is large or if the trait mapped is strongly inherited, the method displays high genetic resolution for linked QTL; for example, using this method, the three QTL can be mapped to the correct locations on chromosome 9 (Figure 2). However, when neither of the two conditions is met, the advantage of the interval test in discriminating adjacent QTL on a chromosome is lost. For QTL of small or medium effects, the influence of interactions between heritabilities and sample sizes is more remarkable.

The estimate of outcrossing rate appears to be unaffected by heritability levels, although it is slightly sensitive to the sample sizes used in the simulations (Tables 2 and 3). Also, the estimate of outcrossing rate is consistent based on all the simulated QTL. Although chromosomes 4, 6, 8, 10, and 12 have no QTL, outcrossing rate can still be estimated with very high accuracy when procedures are implemented to search for possible QTL on these chromosomes (data not shown). The estimate of QTL-allelic frequency is affected by both sample sizes and heritabilities. A smaller sample size or heritability results in more biased estimates for this population parameter than a larger sample size or heritability.

In experiment 1, QTL of effects ~0.5 cannot be detected with a sample size of 800. Results from experiment 2 show that QTL with such sizes of effects can be detected when a larger sample size ($n = 1000$) is used (Table 4). Of the 14 simulated QTL of relatively small effects, 8 (57%) can be detected with reasonable accuracy and power.

Experiment 3 includes two parts. In the first part, the frequency of allele Q for the QTL is fixed ($w = 0.5$) in the pollen pool, whereas outcrossing rate ρ is allowed to change from 0 to 1. Both QTL locations and additive

TABLE 4

Experiment 2: Average values (±SE) and empirical power of the ML estimates of a QTL affecting a polygenic trait

Chromosome	Simulated value			Estimated value					
	Position	a	d	Position	$\hat{a}$	$\hat{d}$	Power	$\hat{\rho}$	$\hat{w}$
1	32	0.5	0.4	33 ± 6	0.7 ± 0.08		0.54	0.88 ± 0.11	0.48 ± 0.07
	175	0.4	0.3						
2	19	0.3	0.3						
3	15	−0.4	−0.4						
	134	−0.3	−0.5	137 ± 18		−0.6 ± 0.06	0.56	0.88 ± 0.09	
	186	0.5	−0.4	187 ± 20	0.6 ± 0.07		0.62	0.90 ± 0.10	0.49 ± 0.08
4	—	—	—						
5	48	0.5	0.4	46 ± 8	0.7 ± 0.08		0.40	0.89 ± 0.11	0.45 ± 0.06
	150	0.3	0.3						
	193	0.3	0.3						
6	—	—	—						
7	110	0.5	0.3	114 ± 20	0.7 ± 0.07		0.61	0.88 ± 0.09	0.48 ± 0.06
8	—	—	—						
9	126	0.4	0.5	125 ± 26	0.5 ± 0.04	0.6 ± 0.08	0.82	0.90 ± 0.11	0.51 ± 0.07
	155	0.3	0.5	158 ± 27		0.5 ± 0.06	0.74	0.89 ± 0.09	0.46 ± 0.06
	191	0.5	0.4	187 ± 29	0.5 ± 0.05		0.81	0.88 ± 0.09	0.49 ± 0.06
10	—	—	—						
11	29	0.3	0.5	32 ± 7		0.6 ± 0.08	0.45	0.89 ± 0.10	0.52 ± 0.09
12	—	—	—						

n = 1000. Results are obtained from 100 replicated simulations.

effects are little influenced by the changes of outcrossing rate, although better estimation is obtained when ρ deviates from 0.5 (Table 5). Estimates of the dominant effect seem to be more sensitive to the change of ρ. The dominant effects can be better estimated when ρ is close to 0.5. In the second part, aimed at observing the influence of *Q*-allelic frequency, ρ is set to be fixed (ρ = 0.9). It is found that variability in allelic frequency for the QTL affects the estimation of QTL parameters (Table 5). When two QTL alleles are in equal frequency, the QTL can be mapped and estimated more accurately than when the QTL-allelic frequencies tend toward extremes. This is not unexpected because the frequency of informative families will decrease in the extreme case. The dominant effects are overestimated if the frequency of allele *Q* is low. In general, the frequency of the QTL allele can be well estimated, especially when the two QTL alleles have equal frequency.

DISCUSSION

Theoretically, the strategy based on a well-defined pedigree, such as F_2 or backcross, is not effective to map QTL in outcrossing species. As a result, the strategies based on a half- (HS) or full-sib (FS) family design have been developed for these species (Knott and Haley 1992; Mackinnon and Weller 1995; Hoeschele *et al.* 1997; Uimari and Hoeschele 1997; Liu and Dekkers 1998; Xu 1998). However, these new strategies are strongly dependent on parental selection and controlled crosses. If the parents used for crosses are not randomly selected from a population or if a particular cross does not produce adequate progeny, then type II error would occur with these strategies. In this article, a similar strategy based on an open-pollination (OP) design is proposed to overcome the limitation of controlled crosses.

The OP test is the easiest and least expensive means of creating a progeny population. Without requiring artificial crosses, this design collects open-pollinated seeds from parental plants that are to be tested. The design has been widely used to understand the overall genetic architecture of quantitative traits for outcrossing species (*e.g.*, Namkoong and Kang 1990). Field tests using the OP progeny have provided numerous estimates of additive genetic variance and heritability for the populations being tested. However, because only one parent is known, estimates of nonadditive genetic variance cannot be obtained. Through genomic mapping, I extend the OP design to estimate genetic parameters at the molecular level, such as the number of individual QTL and their positions, effects, gene action, and allelic frequencies in the population. These estimates may be obtained from the progeny population derived from any single plant that is only required to be heterozygous at the markers and QTL of interest.

A variety of statistical methodologies have been developed for mapping QTL in plants, animals, and humans (reviewed by Hoeschele *et al.* 1997). The advantage of simple regression analyses is that they are computation-

TABLE 5

Experiment 3: Average values (±SE) and empirical power of ML estimates of a simulated QTL obtained from 100 replicated simulations under different outcrossing rates (ρ) or QTL-allelic frequencies (*w*)

	Effect of different ρ's (*w* is fixed)					Effect of different *w*'s (ρ is fixed)				
ρ or *w*	Position	$a = 1.2$	$d = 1.0$	$w = 0.50$	Power	Position	$a = 1.2$	$d = 1.0$	$w = 0.50$	Power
0	50 ± 5	1.2 ± 0.09	1.0 ± 0.08	—	1.00	47 ± 11	1.3 ± 0.18	1.2 ± 0.19	0.05 ± 0.08	0.82
0.1	49 ± 6	1.2 ± 0.10	1.0 ± 0.13	0.49 ± 0.09	0.95	48 ± 10	1.3 ± 0.17	1.1 ± 0.16	0.14 ± 0.08	0.84
0.2	48 ± 7	1.2 ± 0.12	1.0 ± 0.14	0.50 ± 0.10	0.95	48 ± 9	1.3 ± 0.16	1.1 ± 0.16	0.23 ± 0.12	0.86
0.3	48 ± 8	1.3 ± 0.13	1.1 ± 0.15	0.49 ± 0.11	0.94	49 ± 9	1.2 ± 0.15	1.1 ± 0.13	0.32 ± 0.11	0.89
0.4	48 ± 10	1.3 ± 0.14	1.2 ± 0.17	0.50 ± 0.12	0.93	49 ± 8	1.2 ± 0.15	1.0 ± 0.12	0.41 ± 0.10	0.91
0.5	48 ± 12	1.4 ± 0.16	1.2 ± 0.20	0.49 ± 0.14	0.88	50 ± 7	1.2 ± 0.13	1.0 ± 0.10	0.50 ± 0.08	0.95
0.6	48 ± 11	1.3 ± 0.15	1.2 ± 0.19	0.48 ± 0.13	0.85	49 ± 8	1.2 ± 0.15	1.0 ± 0.12	0.59 ± 0.10	0.94
0.7	49 ± 10	1.3 ± 0.14	1.1 ± 0.16	0.51 ± 0.12	0.95	48 ± 8	1.2 ± 0.16	1.2 ± 0.14	0.69 ± 0.11	0.90
0.8	49 ± 8	1.2 ± 0.13	1.1 ± 0.14	0.49 ± 0.11	0.96	48 ± 9	1.2 ± 0.16	1.1 ± 0.16	0.78 ± 0.09	0.89
0.9	50 ± 7	1.2 ± 0.13	1.0 ± 0.10	0.50 ± 0.08	0.95	47 ± 10	1.3 ± 0.17	1.1 ± 0.17	0.85 ± 0.10	0.85
1.0	50 ± 6	1.2 ± 0.11	1.0 ± 0.08	0.50 ± 0.07	0.98	47 ± 11	1.3 ± 0.19	1.1 ± 0.18	0.94 ± 0.10	0.82

ally efficient. However, these methods cannot extract all the information in the data. Using these methods, one should perform data permutation to determine significant thresholds and Monte Carlo algorithms to estimate the sampling variances of parameters. For these reasons, simple methods have been recommended as initial data exploration from which more sophisticated methods, such as ML and Bayesian analysis, will be pursued (Hoeschele *et al.* 1997). The ML-based statistical methods have been extensively developed to map QTL segregating in a progeny population (Weller 1986; Lander and Botstein 1989; Zeng 1994; Mackinnon and Weller 1995; Xu and Atchley 1995; Jansen *et al.* 1998; Xu 1998). Mackinnon and Weller (1995) combined the ML method and a HS design to estimate QTL parameters in a segregating outbred population. Their method can simultaneously estimate several parameters related to a marker-linked QTL, *i.e.*, the additive and dominant effect, the recombination frequency between the marker and QTL, and the QTL-allelic frequency. However, their method needs information about marker-allelic frequency, which may cause large sampling errors for parameter estimates. When wrong marker-allelic frequencies are used, the QTL is estimated to be larger and more distant from the marker than it really is. In addition, their analysis was based on only a single linked marker and did not take full advantage of Zeng's CIM method.

In this article, I have combined CIM and an OP design to estimate QTL parameters through haploid tissues from single meiotic events. Methodologically, this combination has three favorable properties. First, the molecular characterization of individual alleles at markers is simple and accurate from the haploid tissues. By scoring the presence *vs.* absence of bands, the haploid tissues can be genotyped using PCR-dominant markers such as RAPDs and AFLPs (Plomion *et al.* 1995). Second, based on only a single heterozygous plant, the new method can provide information about the genetic architecture of a population by estimating outcrossing rate and QTL-allelic frequency. Simulation results show that these two parameters can be estimated with high accuracy and that their influences on estimates of other parameters can be ignored. Third, because the conditional probabilities of QTL genotypes upon marker genotypes of haploid tissues are independent of marker-allelic frequencies and linkage disequilibria, results from the new method are not affected by these two variables. When molecular markers are derived from diploid tissues, the accuracy for estimating QTL parameters is very sensitive to estimates of marker-allele frequency in the population (Mackinnon and Weller 1995) and linkage disequilibria between the markers and QTL (R. L. Wu, unpublished data).

Results from simulation experiments have demonstrated that the new method can be well used in practice. However, estimates from this method are asymptotically

unbiased; reduced sample sizes will result in reduced power to detect a QTL and increased biases in estimating this QTL's position and effect (see also Beavis 1994; Carson *et al.* 1996; Kaeppler 1997; Wilcox *et al.* 1997). For example, when a sample size of 800 is used, 87% of the simulated QTL can be detected for a trait of $H^2 = 0.60$, whereas the use of a sample size of 300 can only detect 57% for the same trait. For a quantitative trait of small heritability, improvements in the accuracy of parameter estimation with increased sample sizes are not as evident as those for a trait of large heritability (Figure 1). This is especially true for those traits that are not strongly inherited or for polygenic traits in which only QTL of small effects are involved. In addition, the mapping population suited for the current method includes mixed selfed and outcrossed progenies, with the percentage of selfed progeny depending on outcrossing rate. By affecting the phenotypic distribution of the mixed progeny population, inbreeding depression, frequently observed in the selfed progeny of conifers (Zobel and Talbert 1991), may have some impact on the reliability of parameter estimates. However, the extent to which inbreeding depression affects parameter estimation should be assessed via simulation experiments.

Two simplifying assumptions have been used to derive the statistical method proposed in this article. The first is that the QTL to be mapped are biallelic. This assumption can be relaxed by developing a normal-effects QTL model. Under a normal-effects or random QTL model, segregating variances instead of genetic effects for the QTL are estimated without prior knowledge about the number of QTL alleles (Xu and Atchley 1995; Xu 1996, 1998). The second assumption used in the present model is that no epistatic effects exist between QTL. Although multiple QTL have been modeled in recent years (Uimari and Hoeschele 1997; Jansen *et al.* 1998), a method for estimating epistasis between alleles of different QTL has not been well developed. The multiple-QTL model based on ML or Bayesian analysis can be further extended to map epistatic QTL (*e.g.*, Kao and Zeng 1997; Kao *et al.* 1999).

The mapping strategy proposed in this article is dependent on the availability of haploid nongametic tissues derived from single meiotic events. Such tissues that naturally exist for marker analysis include the megagametophyte of gymnosperms (Bierhorst 1971). However, for those species in which genotyping of haploid tissues is currently not available, a pseudo-testcross strategy based on a FS family is still an effective means of mapping QTL (Grattapaglia and Sederoff 1994). The FS family mapping is advantageous over the present mapping approach when there do not exist adequate heterozygous loci for the maternal parent. Also, as compared to an OP family, smaller residual variance in an FS family due to a single pollen parent can increase the power to detect a QTL. The haploid-based strategy cannot make use of the existing gymnosperm populations whose megagametophytes have not been stored. The megagametophyte is a temporary tissue with small amounts of DNA. Thus, it is difficult to use the same mapping population at a later time when new marker techniques become feasible. Despite these limitations, however, it is anticipated that the new method can be broadly useful for mapping quantitative traits in outcrossing species, because modern biotechnology can potentially develop to a point where it is possible to genotype haploid products based on a single cell.

Conclusions: The study shows that a number of genetic parameters regarding QTL positions and effects, and QTL-allelic frequencies and outcrossing rate in a parental population, can be estimated by ML methodology. However, the accuracy of parameter estimates and power to detect a QTL may be reduced when sample sizes and heritability levels are small. The sensitivity of parameter estimates to these two variables indicates that the prior knowledge of heritability is necessary for designing an appropriate experiment for QTL mapping. For those traits with lower heritability, for example, one should increase either sample size or environmental homogeneity, or both, to achieve acceptable precision for QTL detection. With an adequately large mapping population, the method proposed has a capacity to study the genetic basis underlying a polygenic trait.

Estimates of QTL-allelic frequencies and outcrossing rate in a parental population obtained from this method are of great importance to both breeders and population geneticists. If these two parameters are known for a breeding population, the probability is increased of selecting individuals carrying favorable QTL alleles through marker-assisted selection. From a population genetics perspective, these two parameters are essential for understanding the genetic architecture of natural populations.

I thank Prof. R. R. Sederoff and all other members of the Forest Biotechnology Group at North Carolina State University for encouragement and support on this and other studies. I am grateful to Dr. D. M. O'Malley, Dr. Z-B. Zeng, Mr. D. L. Remington, and Dr. B-H. Liu for much discussion regarding QTL mapping using megagametophytes. I especially appreciate Dr. Zhao-Bang Zeng, Dr. Shizhong Xu, Dr. Ruth Shaw, and three anonymous referees for thoughtful comments that led to a better presentation of this work. This work is partially supported by the North Carolina State University Forest Biotechnology Industrial Associates Consortium.

LITERATURE CITED

Beavis, W. D., 1994 The power and deceit of QTL experiments: lessons from comparative QTL studies. Proceedings of the 49th Annu. Corn and Sorghum Indus. Res. Conf., American Seed Trade Association, Washington, DC, pp. 250–266.

Benelli, C., and G. Bucci, 1994 A genetic linkage map of *Picea abies* Karst, based on RAPD markers as a tool in population genetics. Theor. Appl. Genet. **88:** 283–288.

Bierhorst, D. W., 1971 *Morphology of Vascular Plants.* Macmillan, New York.

Broman, K. W., 1997 *Identifying Quantitative Trait Loci in Experimental Crosses.* Ph.D. Thesis, University of California, Berkeley, CA.

Bulmer, M. G., 1980 *The Mathematical Theory of Quantitative Genetics*. Oxford University Press, London.

Burczyk, J., W. T. Adams and J. Y. Shimizu, 1997 Mating system and genetic diversity in natural populations of knobcone pine (*Pinus attenuata*). For. Genet. **4:** 223–226.

Carson, S. D., T. E. Richardson, G. E. Corbett, J. R. Lee and P. L. Wilcox, 1996 Validation of statistically significant linkages of RAPD markers and wood density in *Pinus radiata* D. Don. (Abstr. P245), Plant Genome, January 14–18, 1996, San Diego.

Dale, G. T., 1994 Genetic mapping in quantitative trait analysis in the *Pinus elliotti* Enghelm. × *Pinus carribaea* More. hybrid and in *Pinus radiata* Don. Ph.D. Thesis, University of Queensland, Brisbane, Australia.

Dragani, T. A., Z. B. Zeng, F. Canzian, M. Gariboldi, G. Manenti *et al.*, 1995 Molecular mapping of body weight loci on mouse chromosome X. Mamm. Genome **6:** 778–781.

Falconer, D. S., and T. F. C. MacKay, 1996 *Introduction to Quantitative Genetics*, Ed. 4. Longman Sci. and Tech., Harlow, UK.

Grattapaglia, D., and R. R. Sederoff, 1994 Genetic linkage maps of *Eucalyptus grandis* and *E. urophylla* using a pseudo-testcross mapping strategy and RAPD markers. Genetics **137:** 1121–1137.

Grattapaglia, D., P. Wilcox *et al.*, 1991 A RAPD map of loblolly pine in 60 days. Paper presented at the International Society for Plant Molecular Biology International Congress, Tucson, Arizona.

Hamrick, J. L., M. J. W. Godt and S. L. Sherman-Broyles, 1992 Factors influencing levels of genetic diversity in woody plant species. New For. **6:** 95–124.

Hoeschele, I., P. Uimari, F. E. Grignola, Q. Zhang and K. M. Gage, 1997 Advances in statistical methods to map quantitative trait loci in outbred populations. Genetics **147:** 1445–1457.

Huang, Q. Q., N. Tomaru, L. H. Wang and K. Ohba, 1994 Genetic control of isozyme variation in Masson pine, *Pinus massoniana* Lamb. Silvae Genet. **43:** 285–292.

Jansen, R. C., and P. Stam, 1994 High resolution of quantitative traits into multiple loci via interval mapping. Genetics **136:** 1447–1455.

Jansen, R. C., D. L. Johnson and J. A. M. van Arendonk, 1998 A mixture model approach to the mapping of quantitative trait loci in complex populations with an application to multiple cattle families. Genetics **148:** 391–399.

Jordan, A. P., 1997 *Fusiform Rust Disease Resistance and Genomic Mapping in Loblolly Pine*. Master's Thesis, North Carolina State University, Raleigh, NC.

Kaeppler, S. M., 1997 Quantitative trait loci mapping using sets of near-isogenic lines: relative power comparisons and technical considerations. Theor. Appl. Genet. **95:** 384–392.

Kao, C.-H., and Z.-B. Zeng, 1997 General formulas for obtaining the MLEs and the asymptotic variance-covariance matrix in mapping quantitative trait loci when using the EM algorithm. Biometrics **53:** 653–665.

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

Knott, S. A., and C. S. Haley, 1992 Maximum likelihood mapping of quantitative traits loci using full-sib families. Genetics **132:** 1211–1222.

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

Liu, Z., and J. C. M. Dekkers, 1998 Least squares interval mapping of quantitative trait loci under the infinitesimal genetic model in outbred populations. Genetics **148:** 495–505.

Liu, J., J. M. Mercer, L. F. Stem, G. C. Gibson, Z.-B. Zeng *et al.*, 1996 Genetic analysis of a morphological shape difference in the male genitalia of *Drosophila simulans* and *D. mauritiana*. Genetics **142:** 1129–1145.

Mackinnon, M. J., and J. I. Weller, 1995 Methodology and accuracy of estimation of quantitative trait loci parameters in a half-sib design using maximum likelihood. Genetics **141:** 755–770.

Meng, X.-L., and D. B. Rubin, 1993 Maximum likelihood estimation via the ECM algorithm: a general framework. Biometrika **80:** 267–278.

Millar, C. I., 1983 A steep cline in *Pinus muricata*. Evolution **37:** 311–319.

Namkoong, G., and H. H. Kang, 1990 Quantitative genetics of forest trees. Plant Breed. Rev. **8:** 139–188.

Nelson, C. D., W. L. Nance and R. L. Doudrick, 1993 A partial genetic linkage map of slash pine (*Pinus elliottii* Enghelm. var. *elliottii*) based on random amplified polymorphic DNAs. Theor. Appl. Genet. **87:** 145–151.

Nuzhdin, S. V., E. G. Pasyukova, C. L. Dilda, Z.-B. Zeng and T. F. C. MacKay, 1997 Sex-specific quantitative trait loci affecting longevity in Drosophila melanogaster. Proc. Natl. Acad. Sci. USA **94:** 9734–9739.

O'Malley, D. M., 1996 Complex trait dissection in forest trees using molecular markers, pp. 49–70 in *The Impact of Plant Molecular Genetics*, edited by B. W. S. Sobral. Birkhäuser, Boston.

Plomion, C., N. Bahrman, C.-E. Durel and D. M. O'Malley, 1995 Genomic mapping in *Pinus pinaster* (maritime pine) using RAPD and protein markers. Heredity **74:** 661–668.

Rogers, D. L., 1997 Inheritance of allozymes from seed tissues of the hexaploid gymnosperm, *sequoia sempervirens* (D. Don) Edl. (coast redwood). Heredity **78:** 166–175.

Tulserium, L. K., J. C. Glaubitz, G. Kiss and J. Carlson, 1992 Single tree genetic linkage mapping using haploid DNA from megagametophytes. BioTechnology **10:** 686–690.

Uimari, P., and I. Hoeschele, 1997 Mapping linked quantitative trait loci using Bayesian analysis and Markov chain Monte Carlo algorithms. Genetics **146:** 735–743.

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

Weller, J. I., 1986 Maximum likelihood techniques for the mapping and analysis of quantitative trait loci with the aid of genetic markers. Biometrics **42:** 627–640.

Wheeler, N. C., and R. P. Guries, 1982 Population structure, genetic diversity, and morphological variation in *Pinus contorta* Dougl. Can. J. For. Res. **12:** 595–606.

Wilcox, P. L., H. V. Amerson, E. G. Kuhlman, B.-H. Liu, D. O. O'Malley *et al.*, 1996 Detection of a major gene for resistance to fusiform rust disease in loblolly pine by genomic mapping. Proc. Natl. Acad. Sci. USA **93:** 3859–3864.

Wilcox, P. L., T. E. Richardson and S. D. Carson, 1997 Nature of quantitative trait variation in *Pinus radiata*: insights from QTL detection experiments, pp. 304–312 in *Proc. IUFRO '97: Genetics of Radiata Pine*, edited by R. D. Burdon and J. M. Moore. 1–5 December 1997, Rotorua, New Zealand. FRI Bulletin No. 203.

Wu, R. L., Z.-B. Zeng, D. M. O'Malley, S. E. McKeand and R. R. Sederoff, 1999 The case for molecular mapping in forest tree breeding. Plant Breed. Rev. (in press).

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

Xu, S., 1998 Mapping quantitative trait loci using multiple families of line crosses. Genetics **148:** 517–524.

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

Yan, T., 1997 *Construction of Genetic Linkage Maps in* Populus *and* Pinus massoniana. Ph.D. Thesis, Nanjing Forestry University, Nanjing, China.

Yazdani, R., F. C. Yeh and J. Rimsha, 1995 Genomic mapping of *Pinus sylvestris* (L.) using random amplified polymorphic DNA markers. For. Genet. **2:** 109–116.

Zeng, Z.-B., 1993 Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. Proc. Natl. Acad. Sci. USA **90:** 10972–10976.

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

Zheng, Y., and R. Ennos, 1997 Changes in the mating systems of populations of *Pinus caribaea* Morelet var. *caribaea* under domestication. For. Genet. **4:** 209–215.

Zobel, B., and J. Talbert, 1991 Applied forest tree improvement. Waveland Press, Prospect Heights, IL.

Communicating editor: R. G. Shaw