

Bayesian Mapping of Quantitative Trait Loci Under the Identity-by-Descent-Based Variance Component Model

Nengjun Yi and Shizhong Xu

Department of Botany and Plant Sciences, University of California, Riverside, California, 92521-0124

Manuscript received December 6, 1999

Accepted for publication May 19, 2000

ABSTRACT

Variance component analysis of quantitative trait loci (QTL) is an important strategy of genetic mapping for complex traits in humans. The method is robust because it can handle an arbitrary number of alleles with arbitrary modes of gene actions. The variance component method is usually implemented using the proportion of alleles with identity-by-descent (IBD) shared by relatives. As a result, information about marker linkage phases in the parents is not required. The method has been studied extensively under either the maximum-likelihood framework or the sib-pair regression paradigm. However, virtually all investigations are limited to normally distributed traits under a single QTL model. In this study, we develop a Bayes method to map multiple QTL. We also extend the Bayesian mapping procedure to identify QTL responsible for the variation of complex binary diseases in humans under a threshold model. The method can also treat the number of QTL as a parameter and infer its posterior distribution. We use the reversible jump Markov chain Monte Carlo method to infer the posterior distributions of parameters of interest. The Bayesian mapping procedure ends with an estimation of the joint posterior distribution of the number of QTL and the locations and variances of the identified QTL. Utilities of the method are demonstrated using a simulated population consisting of multiple full-sib families.

THE identity-by-descent (IBD)-based variance component analysis is a powerful statistical method for quantitative trait loci (QTL) mapping in outbred populations, such as humans. This method requires fewer assumptions than other methods with regard to the genetic model underlying the expression of the trait in question. For instance, knowledge of the actual genetic mechanism of the trait, such as the number of loci, the number of alleles per locus, the allelic frequencies, or the marker linkage phases, is not absolutely required (GOLDGAR 1990; SCHORK 1993; AMOS 1994; FULKER and CARDON 1994; XU and ATCHLEY 1995; ALMASY and BLANGERO 1998). Conventionally, the method decomposes the overall genetic variance into several variance components, one being due to the segregation of a putative QTL at the position being tested and the other due to the effect of a polygenic term (the collective effects of all other quantitative loci affecting the trait). The key to separating the contribution of a putative QTL from that of the polygene is the differentiated proportion of IBD alleles shared by relatives at the QTL and the polygene. The IBD proportion varies from one locus to another, which provides the capability of locating QTL on the chromosome.

Existing methods of QTL mapping under the IBD-based variance component model are developed under

a single QTL model. When multiple QTL exist in the same chromosome, a proportion of effects of QTL not included in the model is confounded with the effect of the QTL fitted in the model and the remaining proportion is absorbed into the polygenic component. As a consequence, the single-QTL model can lead to biased estimates of QTL positions and effects because of the interference between QTL located on the same chromosome (*e.g.*, HALEY and KNOTT 1992; GRIGNOLA *et al.* 1996). In theory, effects of multiple QTL can be simultaneously fitted in the same model, but this can be difficult to implement in practice because even the number of QTL is unknown. For line-crossing experiments, JANSEN (1993) and ZENG (1994) developed the idea of composite interval mapping in which selected markers in untested regions are fitted in the model as cofactors to absorb effects of background QTL. Recently, KAO *et al.* (1999) developed a multiple interval mapping (MIM) approach, designed particularly for multiple QTL in line crosses. Yet, extension of the MIM to the IBD-based variance component approach under the maximum-likelihood framework is not straightforward.

Almost all methods of QTL mapping under the IBD-based variance component model are developed for normally distributed traits. However, many complex human diseases, such as breast cancer and type I diabetes, are dichotomous or binary. Although these traits have a simple qualitatively expressed phenotype, their genetic architectures are generally complex, involving multiple genetic factors. Furthermore, the expression of the phenotype is often sensitive to environmental variation. As

Corresponding author: Nengjun Yi, Department of Botany and Plant Sciences, University of California, Riverside, CA 92521.
E-mail: yi@genetics.ucr.edu

a consequence, these traits are usually called complex binary diseases and are commonly formulated under a threshold model. This model assumes a latent continuous variable (called the liability) controlling the expression of the binary trait (LYNCH and WALSH 1998). Methods of QTL mapping under the threshold model have been developed in line crosses (HACKETT and WELLER 1995; VISSCHER *et al.* 1996; XU and ATCHLEY 1996; REBAI 1997; RAO and XU 1998). YI and XU (1999a,b) recently developed a random model approach to directly estimate and test the QTL variances in outbred populations. Under a single-QTL model, DUGGIRALA *et al.* (1997) investigated the IBD-based variance component method using the Mendell-Elston algorithm (MENDELL and ELSTON 1974) to approximate the likelihood function.

Bayes methods of QTL mapping have been developed, in particular, for detection of multiple QTL (SATAGOPAN and YANDELL 1996; SATAGOPAN *et al.* 1996; UIMARI and HOESCHELE 1997; HEATH 1997; SILLANPÄÄ and ARJAS 1998, 1999; STEPHENS and FISCH 1998). In Bayesian analysis, Markov chain Monte Carlo (MCMC) methods are commonly used to evaluate complex integrals to summarize posterior distributions of all unknowns. A recent development in MCMC methodology is the reversible jump algorithm, an extension of the Metropolis-Hastings sampler, which permits posterior samples to be collected from posterior distributions with varying dimensions (GREEN 1995). Bayes methods, implemented via reversible jump MCMC, can yield posterior densities for not only the QTL locations and the corresponding effects of a specified number of QTL but also the QTL number itself. The reversible jump MCMC has been used to map QTL for normally distributed traits in both line-crossing experiments (SATAGOPAN and YANDELL 1996; SILLANPÄÄ and ARJAS 1998, 1999; STEPHENS and FISCH 1998) and complex pedigrees (HEATH 1997; UIMARI and HOESCHELE 1997).

In this article, we develop a Bayes method to map QTL for both normally distributed and binary traits under the IBD-based variance component model. We treat the additive and dominance effects of QTL as random variables so that their variances are directly estimated. The proposed method is implemented via the reversible jump MCMC, where we allow simultaneous estimation of the number of QTL and the locations and variances of the identified QTL.

THE GENETIC MODEL

Linear model for normally distributed traits: Consider n individuals in a population of interest. The population is assumed to consist of many independent families. For convenience of presentation, only full-sib families are considered with phenotypic values of the parents excluded from the data. Parents of these full-sib families are randomly sampled from a large outbred population in Hardy-Weinberg and linkage equilibrium. Extended pedigree structures can be handled fol-

lowing the same procedure as described for full-sib families.

Let \mathbf{y} represent an $n \times 1$ vector for the observed phenotypic values. When the observed phenotype is controlled by multiple genes acting independently, \mathbf{y} can be described by the linear model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \sum_{j=1}^l \mathbf{a}_j + \sum_{j=1}^l \mathbf{d}_j + \mathbf{g} + \mathbf{e}, \quad (1)$$

with $E(\mathbf{a}_j) = E(\mathbf{d}_j) = E(\mathbf{g}) = E(\mathbf{e}) = \mathbf{0}$, $\text{Var}(\mathbf{a}_j) = \Pi_j \sigma_{a_j}^2$, $\text{Var}(\mathbf{d}_j) = \Delta_j \sigma_{d_j}^2$, $\text{Var}(\mathbf{g}) = \mathbf{A} \sigma_{\Lambda}^2$, and $\text{Var}(\mathbf{e}) = \mathbf{I} \sigma_e^2$, where $\boldsymbol{\beta}$ is a $p \times 1$ vector of covariate effects (fixed effects, including the overall mean), \mathbf{X} is an $n \times p$ design matrix relating $\boldsymbol{\beta}$ to \mathbf{y} , l is the number of QTL on the chromosome(s) (or chromosomal segments) of interest, \mathbf{g} is an $n \times 1$ vector of the additive effects of the polygene (collective additive effects of all QTL residing on other chromosomes), \mathbf{e} is an $n \times 1$ vector of residual errors, \mathbf{a}_j and \mathbf{d}_j are $n \times 1$ vectors for the additive and dominance effects of the genotypes at the j th QTL, respectively, $\Pi_j = (\pi_{jii'})_{n \times n}$ is an IBD matrix with element $\pi_{jii'}$ being the proportion of genes IBD shared by individuals i and i' at the j th QTL, $\Delta_j = (\delta_{jii'})_{n \times n}$ is a matrix with element $\delta_{jii'}$ indicating whether individuals i and i' share two IBD alleles at the j th QTL, $\sigma_{a_j}^2$ and $\sigma_{d_j}^2$ are the additive and dominance variances of the j th QTL, respectively, $\mathbf{A} = (A_{ii'})_{n \times n}$ is the additive genetic relationship matrix with element $A_{ii'}$ being twice the coancestry coefficient between offspring i and i' (not conditional on marker information), σ_{Λ}^2 is the additive variance of the polygene (see Table 1), \mathbf{I} is an identity matrix, and σ_e^2 is the residual variance. Note that the dominance effect of the polygene is assumed to be absent in model (1).

The expectation and variance-covariance matrix of \mathbf{y} are

$$E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta} \quad (2)$$

and

$$\text{Var}(\mathbf{y}) = \mathbf{V} = \sum_{j=1}^l \Pi_j \sigma_{a_j}^2 + \sum_{j=1}^l \Delta_j \sigma_{d_j}^2 + \mathbf{A} \sigma_{\Lambda}^2 + \mathbf{I} \sigma_e^2, \quad (3)$$

respectively. When families are independent, there are no covariances between effects of members from different families. Therefore, the matrices Π_j , Δ_j , and \mathbf{A} are all blockdiagonal. If individuals i and i' are full sibs, $\pi_{jii'}$ takes one of the four states $\{(0+0)/2, (0+1)/2, (1+0)/2, (1+1)/2\}$. If i and i' are from different families, then $\pi_{jii'} = 0$. The four states show that individuals i and i' share no alleles, share the paternal but not the maternal alleles, share the maternal but not the paternal alleles, and share both alleles identity-by-descent, respectively. Similarly, $\delta_{jii'}$ takes one of the two values $\{1, 0\}$. If individuals i and i' share both alleles IBD, $\delta_{jii'} = 1$; otherwise, $\delta_{jii'} = 0$. Note that $\pi_{jii'}$ and $\delta_{jii'}$ vary from sib pair to sib pair but element $A_{ii'}$ of matrix \mathbf{A} is a constant across all sib pairs. Because of this, $\sigma_{a_j}^2$, $\sigma_{d_j}^2$, and

TABLE 1
Description of symbols used in the text

| Symbol | Description |
|---|---|
| β | Vector of covariate effects |
| $\mathbf{a}_j, \mathbf{d}_j$ | Vectors of the additive and dominance effects of the genotypes at j th QTL |
| $\sigma_{a_j}^2, \sigma_{d_j}^2$ | Additive and dominance variances of j th QTL |
| \mathbf{g}, σ_A^2 | Vector of additive effects of polygene, polygenic additive variance |
| $\Pi_j = (\pi_{jii'})_{n \times n}$ | $\pi_{jii'}$ being the proportion of genes IBD shared by individuals i and i' at the j th QTL |
| $\Delta_j = (\delta_{jii'})_{n \times n}$ | $\delta_{jii'}$ indicating whether individuals i and i' share two IBD alleles at the j th QTL |
| \mathbf{A} | Additive genetic relationship matrix |

σ_A^2 can be separated, which is the theoretical basis of the variance component model of QTL mapping.

Multipoint inference of the IBD matrices Π_j and Δ_j :

In linkage analysis, the IBD matrix of a QTL is not observable because the QTL genotype cannot be seen. In full-sib families without inbreeding, the diagonal elements of the matrices Π_j and Δ_j are unity, but the off-diagonal elements vary depending on how many IBD alleles are shared by the two siblings. Thus the rationale is to infer the distributions of IBD variables $\pi_{jii'}$ and $\delta_{jii'}$ using markers in the same linkage group. In outbred populations, markers may be partially informative, and thus it is important to use a multipoint method to extract the maximum amount of marker information. A number of multipoint methods have been proposed to calculate the distributions of IBD (FULKER *et al.* 1995; KRUGLYAK and LANDER 1995; ALMASY and BLANGERO 1998; XU and GESSLER 1998). The method of KRUGLYAK and LANDER (1995) is the most efficient with regard to extracting the maximum amount of information from markers. However, their method is also the most intensive in terms of computational time and complexity. The method of FULKER *et al.* (1995), on the other hand, is an approximate method, but it is a much simpler and faster algorithm. The method of XU and GESSLER (1998) is a compromise between the two. The multipoint method used in this study is a modified version of XU and GESSLER (1998) and is described below.

Consider M ordered markers on the chromosome of interest. If a marker is fully informative, the IBD states of the marker shared by sibs are observed. Otherwise, the probabilities of IBD states of a marker can be inferred based on the observed genotypes of this marker (XU and GESSLER 1998). Denote the probabilities of IBD states of marker locus k by $p_{00}^k, p_{01}^k, p_{10}^k, p_{11}^k$, respectively, for the four states. Assume that the QTL is located between marker k and $k + 1$ for $1 \leq k \leq M - 1$. What we want is to calculate the probabilities of IBD states of the QTL conditional on the probabilities of IBD states of all markers, *i.e.*, $\Pr(\pi_{i'ii'}^j | \pi_{i'ii'}^1, \dots, \pi_{i'ii'}^M)$. When the marker IBD states are not directly observed, $\Pr(\pi_{i'ii'}^j | \pi_{i'ii'}^1, \dots, \pi_{i'ii'}^M)$ should be denoted by $\Pr(\pi_{i'ii'}^j | I_M)$, where I_M means the marker information. From Bayes' theorem, the probability of IBD states of QTL given the marker information is

$$\Pr(\pi_{i'ii'}^j | I_M) = \frac{\Pr(\pi_{i'ii'}^j) \Pr(I_M | \pi_{i'ii'}^j)}{\sum_{\pi_{i'ii'}^j} \Pr(\pi_{i'ii'}^j) \Pr(I_M | \pi_{i'ii'}^j)},$$

where $\Pr(\pi_{i'ii'}^j)$ is the prior distribution of the IBD state, equal to $1/4$ for each of the four possible IBD states. After some algebraic manipulations, we have

$$\Pr\left(I_M | \pi_{i'ii'}^j = \frac{0+0}{2}\right) = \mathbf{1}^T \mathbf{D}_1 \mathbf{T}_1 \mathbf{D}_2 \cdots \mathbf{T}_{kq} \mathbf{D}_{(00)} \mathbf{T}_{qk+1} \cdots \mathbf{D}_{M-1} \mathbf{T}_{M-1} \mathbf{1},$$

$$\Pr\left(I_M | \pi_{i'ii'}^j = \frac{0+1}{2}\right) = \mathbf{1}^T \mathbf{D}_1 \mathbf{T}_1 \mathbf{D}_2 \cdots \mathbf{T}_{kq} \mathbf{D}_{(01)} \mathbf{T}_{qk+1} \cdots \mathbf{D}_{M-1} \mathbf{T}_{M-1} \mathbf{1},$$

$$\Pr\left(I_M | \pi_{i'ii'}^j = \frac{1+0}{2}\right) = \mathbf{1}^T \mathbf{D}_1 \mathbf{T}_1 \mathbf{D}_2 \cdots \mathbf{T}_{kq} \mathbf{D}_{(10)} \mathbf{T}_{qk+1} \cdots \mathbf{D}_{M-1} \mathbf{T}_{M-1} \mathbf{1},$$

and

$$\Pr\left(I_M | \pi_{i'ii'}^j = \frac{1+1}{2}\right) = \mathbf{1}^T \mathbf{D}_1 \mathbf{T}_1 \mathbf{D}_2 \cdots \mathbf{T}_{kq} \mathbf{D}_{(11)} \mathbf{T}_{qk+1} \cdots \mathbf{D}_{M-1} \mathbf{T}_{M-1} \mathbf{1},$$

where $\mathbf{1} = (1 \ 1 \ 1 \ 1)^T$, $\mathbf{D}_k = \text{diag}(p_{00}^k \ p_{01}^k \ p_{10}^k \ p_{11}^k)$, $\mathbf{D}_{(00)} = \text{diag}(0 \ 0 \ 0 \ 1)$, $\mathbf{D}_{(01)} = \text{diag}(0 \ 0 \ 1 \ 0)$, $\mathbf{D}_{(10)} = \text{diag}(0 \ 1 \ 0 \ 0)$, $\mathbf{D}_{(11)} = \text{diag}(1 \ 0 \ 0 \ 0)$, and \mathbf{T}_{kl} is the transition matrix between $\pi_{i'ii'}^k$ and $\pi_{i'ii'}^l$,

$$\mathbf{T}_{kl} = \begin{pmatrix} \Psi_{kl}^2 & \Psi_{kl}(1 - \Psi_{kl}) & (1 - \Psi_{kl})\Psi_{kl} & (1 - \Psi_{kl})^2 \\ \Psi_{kl}(1 - \Psi_{kl}) & \Psi_{kl}^2 & (1 - \Psi_{kl})^2 & (1 - \Psi_{kl})\Psi_{kl} \\ (1 - \Psi_{kl})\Psi_{kl} & (1 - \Psi_{kl})^2 & \Psi_{kl}^2 & \Psi_{kl}(1 - \Psi_{kl}) \\ (1 - \Psi_{kl})^2 & (1 - \Psi_{kl})\Psi_{kl} & \Psi_{kl}(1 - \Psi_{kl}) & \Psi_{kl}^2 \end{pmatrix},$$

where $\Psi_{kl} = r_{kl}^2 + (1 - r_{kl})^2$ and r_{kl} is the recombination fraction between loci k and l .

Since $\delta_{i'ii'}^j = 1$ if $\pi_{i'ii'}^j = 1$, and $\delta_{i'ii'}^j = 0$ otherwise, at the time when the probability distribution of $\pi_{i'ii'}^j$ is calculated, that of $\delta_{i'ii'}^j$ is also generated as a by-product. The conditional expectations of $\pi_{i'ii'}^j$ and $\delta_{i'ii'}^j$ are calculated by

$$E(\pi_{i'ii'}^j | I_M) = \Pr\left(\pi_{i'ii'}^j = \frac{1+1}{2} \middle| I_M\right) + \frac{1}{2} \left[\Pr\left(\pi_{i'ii'}^j = \frac{0+1}{2} \middle| I_M\right) + \Pr\left(\pi_{i'ii'}^j = \frac{1+0}{2} \middle| I_M\right) \right]$$

and

$$E(\delta_{i'ii'}^j | I_M) = \Pr\left(\pi_{i'ii'}^j = \frac{1+1}{2} \middle| I_M\right).$$

Both the conditional distributions and conditional

expectations of the IBD matrices $\{\Pi_j\}$ and $\{\Delta_j\}$ can be used in the Bayesian analysis of QTL mapping. With the conditional distribution method, the states of $\{\Pi_j\}$ and $\{\Delta_j\}$ have to be sampled from their conditional distributions in the Bayesian procedure. In a maximum-likelihood (ML) analysis of IBD-based variance components, GESSLER and XU (1996) discovered that the conditional distribution and expectation methods have virtually no difference, but the conditional expectation method is computationally much more efficient than the conditional distribution method. Because of this, we use only the conditional expectations in the proposed method, *i.e.*, replacing $\{\Pi_j\}$ and $\{\Delta_j\}$ by their conditional expectations.

BAYESIAN MAPPING FOR NORMALLY DISTRIBUTED TRAITS

The observables are the phenotypic values $\mathbf{y} = \{y_j\}_{j=1}^n$, the covariate data \mathbf{X} , and the marker data I_M . The locations of markers on chromosomes are known *a priori*. The list of unobservables contains the number of QTL l , the locations of QTL $\lambda = \{\lambda_j\}_{j=1}^l$, and the model effects $\theta = (\beta, \sigma_{a_1}^2, \dots, \sigma_{a_b}^2, \sigma_{d_1}^2, \dots, \sigma_{d_b}^2, \sigma_A^2, \sigma_c^2)$, where λ_j denotes the distance of the j th QTL from one end of the chromosome in which the QTL resides. With the IBD-based variance component approach, the polygenic effects and the additive and dominance effects have been integrated out in the likelihood, and, therefore, we do not have to generate them in the MCMC process.

From Bayes' theorem, the joint posterior distribution of the parameters $\{l, \lambda, \theta\}$ given the observables and prior information is

$$p(l, \lambda, \theta | \mathbf{y}) \propto p(\mathbf{y} | l, \lambda, \theta) p(l, \lambda, \theta). \quad (4)$$

Here, we have suppressed the notation for conditional on the observed markers and covariate data. The first term in Equation 4 is the conditional distribution of the phenotypic data given all unknowns, which is usually called the likelihood, and has the following form:

$$f(\mathbf{y} | l, \lambda, \theta) = (2\pi)^{-n/2} |\mathbf{V}|^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{y} - \mathbf{X}\beta)^T \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\beta)\right\}. \quad (5)$$

The second term in Equation 4 is the joint prior distribution of l, λ , and θ , the parameters of our interest. Assuming prior independence of the parameters, we can factorize the joint prior distribution $p(l, \lambda, \theta)$ into the following products:

$$p(l, \lambda, \theta) = p(l) p(\beta) p(\sigma_A^2) p(\sigma_c^2) \prod_{j=1}^l [p(\lambda_j) p(\sigma_{a_j}^2) p(\sigma_{d_j}^2)]. \quad (6)$$

The prior distribution of the number of QTL, l is assumed to be truncated Poisson with mean μ_l and a predefined maximum number L . When no information regarding the QTL locations is available, the prior of

λ_j takes the uniform distribution on the chromosome(s) into consideration. The prior distributions of $\beta, \sigma_A^2, \sigma_c^2, \sigma_{a_j}^2$ and $\sigma_{d_j}^2$ are assumed to be uniforms on predefined intervals, although other priors can be used. The lower and upper bounds for all variance components are usually set to zero and the phenotypic variance present in the data, respectively.

A Markov chain Monte Carlo method is used to generate the joint posterior distribution of all unknowns given in Equation 4. The idea of MCMC is to simulate a random walk in the space of the unknowns. The random walk eventually converges to a stationary distribution. The stationary distribution represents the posterior distribution of the unknowns. Various approaches have been suggested to conduct the MCMC. The Metropolis-Hastings algorithm is a general term for a family of Markov chain simulation methods that are useful for drawing samples from Bayesian posterior distributions (HASTINGS 1970). The Metropolis algorithm and the Gibbs sampler are two commonly used special cases of the Metropolis-Hastings algorithm (METROPOLIS *et al.* 1953; GEMAN and GEMAN 1984). The reversible jump MCMC is an extension of the Metropolis-Hastings sampler, permitting posterior samples to be collected from posterior distributions with varying dimensions (GREEN 1995). With our Bayesian analysis, the number of QTL is treated as an unobservable, which naturally leads us to consider the problem within the general framework of variable dimensional parameter estimation.

The proposed MCMC algorithm starts from an initial point $(l^0, \lambda^0, \theta^0)$ and proceeds to update each of the unknowns in turn. For an initial QTL position, we can calculate the corresponding expectations of the IBD matrices using the multipoint method. Updating θ and λ is implemented using the Metropolis algorithm. Updating the QTL number l requires a change in the dimension of the model and thus needs a reversible jump step. More specifically, given the current state of (l, λ, θ) , we proceed with the MCMC as follows.

Updating θ : All elements of θ are updated simultaneously. New proposals for the elements of θ are sampled from the symmetric uniform densities around their previous values. The proposals are accepted simultaneously with probability

$$\min\left\{1, \frac{f(\mathbf{y} | l, \lambda, \theta^*)}{f(\mathbf{y} | l, \lambda, \theta)}\right\}, \quad (7)$$

where θ^* represents the proposed θ . If the proposal is not accepted, then the state remains unchanged. Updating λ and l : we adopt the method of SILLANPÄÄ and ARJAS (1998, 1999) to implement the reversible jump MCMC. They used three different types of moves: (1) modify the QTL locations when the number of QTL is unchanged; (2) add one new QTL to the model; and (3) delete an existing QTL from the model. If $l = 0$, we cannot delete a QTL. On the

other hand, if $l = L$, the maximum number of QTL assigned *a priori*, we cannot add a QTL. When $0 < l < L$, the different types of moves have proposal probabilities of $p_m = 1/3$, $p_a = 1/3$, or $p_d = 1/3$ to remain unchanged, to add, or to delete a QTL, respectively. Note that other values of p_m , p_a , and p_d can be used.

When the number of QTL remains unchanged, we modify the locations of the existing QTL. Following SILLANPÄÄ and ARJAS (1998, 1999), we do not fix the order of QTL when updating the QTL locations. Elements of λ are modified one at a time using the Metropolis algorithm. For the j th QTL, a proposal λ_j^* is sampled from a uniform distribution on the interval $[\lambda_j - d, \lambda_j + d]$, where d is the tuning parameter. The expectations of the IBD matrices for new location λ_j^* , denoted by Π_j^* and Δ_j^* , are then formed according to the new position (λ_j^*) and the marker information. The proposal is accepted with probability

$$\min\left\{1, \frac{|\mathbf{V}^*|^{-1/2} \exp\{-1/2(\mathbf{y} - \mathbf{X}\beta)^T \mathbf{V}^{*-1}(\mathbf{y} - \mathbf{X}\beta)\}}{|\mathbf{V}|^{-1/2} \exp\{-1/2(\mathbf{y} - \mathbf{X}\beta)^T \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\beta)\}}\right\}, \quad (8)$$

where λ_{-j} means all elements of λ excluding the j th element,

$$\mathbf{V} = \sum_{j=1}^l \Pi_j \sigma_{a_j}^2 + \sum_{j=1}^l \Delta_j \sigma_{d_j}^2 + \mathbf{A}\sigma_{\Lambda}^2 + \mathbf{I}\sigma_{\epsilon}^2$$

and

$$\mathbf{V}^* = \sum_{j \neq l}^l \Pi_j \sigma_{a_j}^2 + \Pi_l^* \sigma_{a_l}^2 + \sum_{j \neq l}^l \Delta_j \sigma_{d_j}^2 + \Delta_l^* \sigma_{d_l}^2 + \mathbf{A}\sigma_{\Lambda}^2 + \mathbf{I}\sigma_{\epsilon}^2.$$

If the proposal is not accepted, the state remains unchanged, and the algorithm proceeds to update the location of the next QTL.

When p_a happens, we add a QTL to the model. The location of the new QTL, λ_{l+1} , is proposed from the uniform density on the chromosome(s) under consideration. The expectations of the IBD matrices for the new QTL, denoted by Π_{l+1} and Δ_{l+1} , are then formed according to the new position (λ_{l+1}) and the marker information. The additive and dominance variances of the new QTL, $\sigma_{a_{l+1}}^2$ and $\sigma_{d_{l+1}}^2$, are drawn from their priors. The proposal is accepted with probability

$$\min\left\{1, \frac{|\mathbf{V}^*|^{-1/2} \exp\{-1/2(\mathbf{y} - \mathbf{X}\beta)^T \mathbf{V}^{*-1}(\mathbf{y} - \mathbf{X}\beta)\}}{|\mathbf{V}|^{-1/2} \exp\{-1/2(\mathbf{y} - \mathbf{X}\beta)^T \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\beta)\}} \times \frac{\mu p_d}{(l+1)^2 p_a}\right\}, \quad (9)$$

where $\mathbf{V}^* = \sum_{j=1}^l \Pi_j \sigma_{a_j}^2 + \Pi_{l+1} \sigma_{a_{l+1}}^2 + \sum_{j=1}^l \Delta_j \sigma_{d_j}^2 + \Delta_{l+1} \sigma_{d_{l+1}}^2 + \mathbf{A}\sigma_{\Lambda}^2 + \mathbf{I}\sigma_{\epsilon}^2$. If the proposal is accepted, we add a QTL to the model, and the new QTL location and

the variances are all accepted simultaneously; otherwise, the QTL number remains unchanged.

When p_d happens, we delete a QTL from the model, each existing QTL being equally likely to be deleted. If the j 'th existing QTL is proposed to be deleted from the model, the acceptance probability is

$$\min\left\{1, \frac{|\mathbf{V}^*|^{-1/2} \exp\{-1/2(\mathbf{y} - \mathbf{X}\beta)^T \mathbf{V}^{*-1}(\mathbf{y} - \mathbf{X}\beta)\}}{|\mathbf{V}|^{-1/2} \exp\{-1/2(\mathbf{y} - \mathbf{X}\beta)^T \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\beta)\}} \times \frac{l^2 p_a}{\mu p_d}\right\}, \quad (10)$$

where $\mathbf{V}^* = \sum_{j=1}^l \Pi_j \sigma_{a_j}^2 - \Pi_j \sigma_{a_j}^2 + \sum_{j=1}^l \Delta_j \sigma_{d_j}^2 - \Delta_j \sigma_{d_j}^2 + \mathbf{A}\sigma_{\Lambda}^2 + \mathbf{I}\sigma_{\epsilon}^2$. If the proposal is accepted, we delete a QTL from the model. Otherwise, the number of QTL remains unchanged.

In Equations 9 and 10, the first term of the acceptance ratio is the likelihood ratio, and the second term contains the prior ratio and the proposal ratio. Note that here the Jacobian of the transformation is one because adding one QTL or deleting one QTL does not influence the parameters of the other QTL.

BAYESIAN MAPPING FOR BINARY TRAITS

For a complex disease trait, the observed phenotype is defined in a binary fashion, *i.e.*, $w_j = 1$ if an individual is affected, and $w_j = 0$ otherwise. We propose an underlying normal variable y_j that determines the binary observation. The link between y_j and w_j is through a threshold t , *i.e.*,

$$w_j = \begin{cases} 1 & \text{if } y_j > t \\ 0 & \text{if } y_j \leq t \end{cases} \quad (11)$$

The liability y is nothing new but an unobservable quantitative trait that can be described by Equation 1. The threshold model is overparameterized so that some constraints must be superimposed. As usual, we take $t = 0$ and $\sigma_{\epsilon}^2 = 1$ (ALBERT and CHIB 1993). The expectation of the liability y is the same as that of Equation 2, but the variance-covariance matrix is

$$\text{Var}(\mathbf{y}) = \mathbf{V} = \sum_{j=1}^l \Pi_j \sigma_{a_j}^2 + \sum_{j=1}^l \Delta_j \sigma_{d_j}^2 + \mathbf{A}\sigma_{\Lambda}^2 + \mathbf{I}. \quad (12)$$

In the threshold model, the observables are the binary phenotypic values $\mathbf{w} = \{w_i\}_{i=1}^n$, the covariate data, and the marker data. The unobservables include the vector of liability $\mathbf{y} = \{y_i\}_{i=1}^n$, the number of QTL l , the locations of QTL $\lambda = \{\lambda_j\}_{j=1}^l$, and the model effects $\theta = (\beta, \sigma_{a_1}^2, \dots, \sigma_{a_l}^2, \sigma_{d_1}^2, \dots, \sigma_{d_l}^2, \sigma_{\Lambda}^2)$.

The joint posterior distribution of the unobservables, $\{\mathbf{y}, l, \lambda, \theta\}$, after suppressing the notation for conditioning on the marker data and the covariate data, can be expressed as

$$p(\mathbf{y}, l, \lambda, \theta | \mathbf{w}) \propto p(\mathbf{w} | \mathbf{y}, l, \lambda, \theta) p(\mathbf{y} | l, \lambda, \theta) p(l, \lambda, \theta), \quad (13)$$

where the likelihood has the form of

$$p(\mathbf{w}|\mathbf{y}, l, \lambda, \theta) = \prod_{i=1}^n p(w_i|y_i) = \prod_{i=1}^n \{1(y_i > 0)1(w_i = 1) + 1(y_i \leq 0)1(w_i = 0)\}, \quad (14)$$

where $1(X \in A)$ is the indicator function, taking the value of 1 if X is contained in A , and 0 otherwise. Note that $p(\mathbf{w}|\mathbf{y}, l, \lambda, \theta) = p(\mathbf{w}|\mathbf{y})$ because \mathbf{w} is solely determined by \mathbf{y} . The second term in Equation 13 is the conditional distribution of the liability given all other unknowns and has the same form as Equation 5. Finally, the last term in Equation 13 is the joint prior distribution of l, λ , and θ .

From this joint posterior density, it can be seen that the posterior distribution of the parameters l, λ , and θ can be computed using the method for normally distributed data if the liability \mathbf{y} is known. To implement the MCMC algorithm for binary traits, therefore, we need to generate the liability \mathbf{y} . Since it is difficult to simulate directly from the conditional distribution of \mathbf{y} given \mathbf{w} and other parameters, we use the method of the Gibbs sampler (CHAN and KUK 1997). The details are as follows.

Starting with some initial values $\mathbf{y}^{(0)} = (y_1^{(0)}, \dots, y_n^{(0)})^T$ consistent in signs with the observed data \mathbf{w} , we proceed to generate $\mathbf{y}^{(t+1)} = (y_1^{(t+1)}, \dots, y_n^{(t+1)})^T$, $t = 0, 1, \dots$, sequentially in the following manner. Given the current state $(\mathbf{y}^{(t)}, \ell^{(t)}, \lambda^{(t)}, \theta^{(t)})$ of all unknowns, we simulate

$$\begin{aligned} & y_1^{(t+1)} \text{ from } f(y_1|y_2^{(t)}, y_3^{(t)}, \dots, y_n^{(t)}, \mathbf{w}; \ell^{(t)}, \lambda^{(t)}, \theta^{(t)}), \\ & \vdots \\ & y_i^{(t+1)} \text{ from } f(y_i|y_1^{(t+1)}, y_2^{(t+1)}, \dots, y_{i-1}^{(t+1)}, y_{i+1}^{(t)}, \dots, y_n^{(t)}, \mathbf{w}; \ell^{(t)}, \lambda^{(t)}, \theta^{(t)}), \\ & \vdots \\ & y_n^{(t+1)} \text{ from } f(y_n|y_1^{(t+1)}, y_2^{(t+1)}, \dots, y_{n-1}^{(t+1)}, \mathbf{w}; \ell^{(t)}, \lambda^{(t)}, \theta^{(t)}). \end{aligned}$$

To carry out the above simulations, we use the fact

$$f(y_i|y_j, j \neq i; w_1, \dots, w_n; l, \lambda, \theta) = f(y_i|y_j, j \neq i; w_i; l, \lambda, \theta).$$

Since $\mathbf{y} = (y_1, \dots, y_n)^T$ has a $N(\mathbf{X}\beta, \mathbf{V})$ distribution, it follows that $f(y_i|y_j, j \neq i; l, \lambda, \theta)$ is also normal with mean

$$E(y_i) + \text{Cov}(y_i, \mathbf{y}_{-i}) (\text{Var}(\mathbf{y}_{-i}))^{-1} (\mathbf{y}_{-i} - E(\mathbf{y}_{-i}))$$

and variance

$$\text{Var}(y_i) - \text{Cov}(y_i, \mathbf{y}_{-i}) (\text{Var}(\mathbf{y}_{-i}))^{-1} \text{Cov}(\mathbf{y}_{-i}, y_i),$$

where $\mathbf{y}_{-i} = (y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_n)^T$. Thus, $f(y_i|y_j, j \neq i; w_i, l, \lambda, \theta)$ is a truncated (above 0 if $w_i = 1$; below 0 if $w_i = 0$) normal distribution. Devroye's algorithm is then used to simulate a truncated normal variable (DEVROYE 1986).

A SIMULATION STUDY

Design of the simulation experiment: The proposed method is illustrated in the analysis of a simulated example. We simulated a single chromosome of length 100 cM. Eleven codominant markers were placed on the

chromosome with a marker distance of 10 cM. Four equally frequent alleles were simulated at each marker locus. Two QTL residing at positions 25 and 75 cM, respectively, and 12 additional independent loci of equal effect (called polygene) were simulated for a continuous phenotype. The dominance effect of the polygene was assumed to be absent. The variances of the first QTL were $\sigma_{a_1}^2 = \sigma_{d_1}^2 = 0.6$ and those of the second QTL were $\sigma_{a_2}^2 = \sigma_{d_2}^2 = 0.4$. The polygenic variance was set at $\sigma_\lambda^2 = 1.0$. The normally distributed phenotype of each individual took the sum of an overall mean, values of QTL additive and dominance effects, polygenic effect, and a residual error sampled from a standardized normal distribution. In addition, we generated a binary phenotype for each individual using the normally distributed phenotypic value as the underlying liability. The binary phenotype took 1 if the liability exceeded 0, and 0 otherwise. The overall mean was set to 0, which led to 50% population incidence. We simulated 500 full-sib families, each with 6 siblings (a total of 3000 individuals). For comparison, we analyzed both the normal data and the binary data. No phenotypic values and marker linkage phases were recorded for all parents.

In each of the MCMC analyses, we ran a single long chain with 5×10^5 cycles of simulations. The first 200 samples were discarded and the chain was then thinned (we saved one iteration in every 50 cycles) to reduce serial correlation in the stored samples so that the total number of samples kept in each analysis was 10^4 .

The initial value for the QTL number was set at 1 and the corresponding initial location was given at the 50-cM position of the chromosome. The prior Poisson mean of the number of QTL was $\mu_l = 2$ and the maximum number of QTL was $L = 4$. The starting values were 0.0 for the overall mean, 1.0 for both the polygenic and residual variances, and 0.5 for both the additive and dominance variances of the QTL. The prior for the overall mean was uniform over the range $[-2.0, 2.0]$. The priors for all variance components were chosen as uniform on $[0.0, 4.0]$, the right endpoint being equal to the true phenotypic variance of the liability. The prior for the QTL locations was uniform over the whole chromosome. The tuning parameter of the proposal distribution was set at 2 cM for QTL locations and 0.1 for the overall mean and all variance components.

We used the QTL intensity function of SILLANPÄÄ and ARJAS (1998) to detect the number and locations of QTL. In practice, we divided the chromosome into intervals of equal length (according to the Haldane distance) and then calculated the proportion of QTL in each interval from the MCMC samples. The interval length was chosen to be 1 cM long. As done in STEPHENS and FISCH (1998) and SILLANPÄÄ and ARJAS (1999), the QTL variances were estimated from samples in which QTL locations fall into the regions with sufficiently high estimated QTL intensities.

Results: The Bayesian posterior QTL intensities for

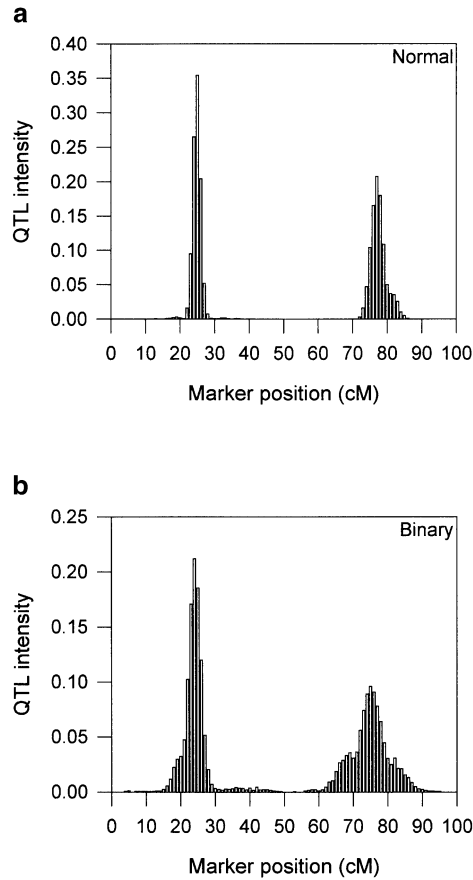


FIGURE 1.—Histograms of the posterior QTL intensity for (a) normally distributed data and (b) binary data. The true number of QTL is two, located at positions 25 and 75 cM, respectively.

both the normal and the binary data are presented in Figure 1. The QTL intensity graphs are concentrated around the true locations of the simulated QTL. One peak of the posterior QTL intensity for the normal data is on [24, 25] and the other on [76, 77]. The corresponding peaks for the binary data are on [23, 24] and [74, 75], respectively. These results support quite strongly a model having two QTL in the chromosome. Comparing the shapes of the posterior QTL intensities for the binary and normal data analyses, we can see that binary data analysis does lose some information, but the information retained is still sufficient to detect both QTL.

The approximate posterior distributions on the QTL number, obtained from the two types of data, are presented in Table 2. The posterior expectations are essentially the same for the binary data and the normal data

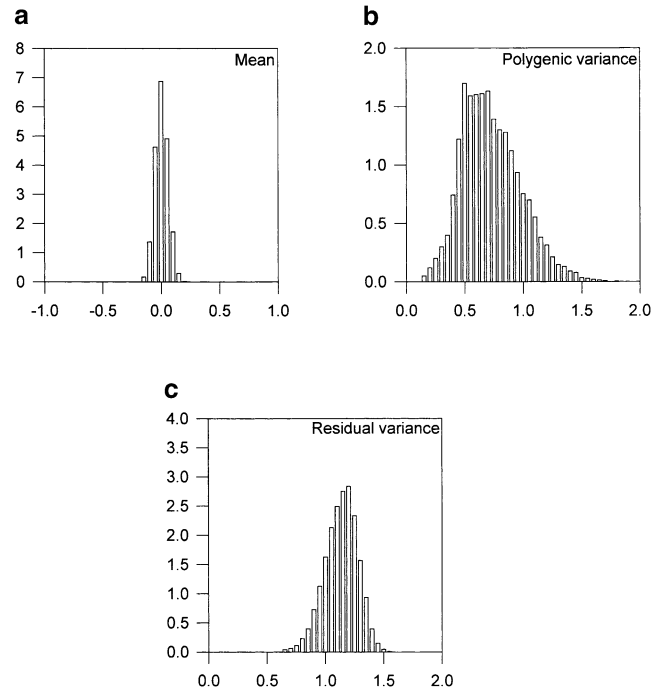


FIGURE 2.—Approximate posterior distributions of the (a) overall mean, (b) polygenic variance, and (c) residual variance for normally distributed data.

and coincide with the simulated number of QTL. As expected, the posterior variance of QTL number for the normal data is smaller than that for the binary data. Finally, the posterior modes of the QTL numbers overlap with the true number of QTL in both data analyses.

The posterior samples for the overall mean, the polygenic variance, and the residual variance from the normal data analysis, and those for the overall mean and the polygenic variance from the binary data analysis, are depicted in Figures 2 and 3, respectively. The posterior means and standard errors of these parameters are given in Table 3. From the figures and the table, it appears that estimates of these parameters are close to the simulated values with small standard errors from the normal data. However, the polygenic variance is greatly overestimated and the estimated standard error is large from the binary data analysis although the mean is still estimated accurately.

The chromosome regions with sufficiently high posterior QTL intensity are given in Table 3. The posterior samples, in which QTL locations fall into these regions, are used to estimate the QTL variances. Figures 4 and 5 depict such posterior samples for the QTL variances

TABLE 2
Inferred posterior distribution and posterior mean of the number of QTL

| No. of QTL | 0 | 1 | 2 | 3 | 4 | Mean |
|---------------------------|--------|--------|--------|--------|--------|--------|
| Frequency for normal data | 0.0000 | 0.0011 | 0.9705 | 0.0282 | 0.0002 | 2.0297 |
| Frequency for binary data | 0.0000 | 0.0022 | 0.9039 | 0.0904 | 0.0035 | 2.0952 |

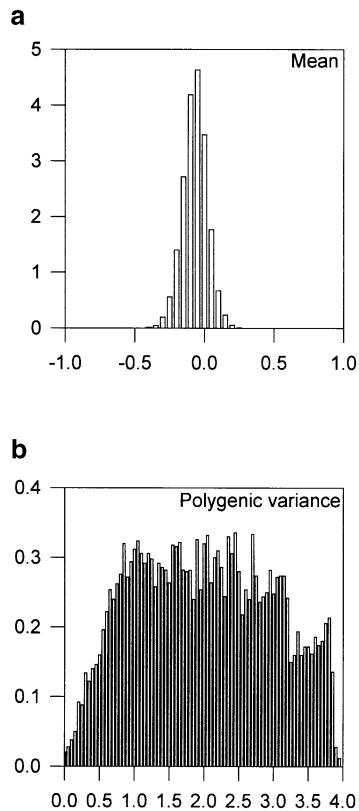


FIGURE 3.—Approximate posterior distributions of the (a) overall mean and (b) polygenic variance for binary data.

from the normal data and the binary data analyses, respectively. The posterior densities of QTL variances are concentrated around the corresponding true values for both types of analyses, although the binary data analysis shows a wider distribution. More explicitly, we calculated the means and the standard errors of the posterior samples for the QTL variances (see Table 3). For the normal data, the estimated posterior means of the additive and dominance variances of the QTL are close to the simulated true values, and the standard

errors are small. In the binary data analysis, however, the additive and dominance variances of the QTL are overestimated, and the standard errors are larger than those for the normal data.

The point estimates and the estimation errors of the QTL locations are also given in Table 3. For both types of data, the estimated QTL locations are very close to the corresponding true values. However, the standard errors are much larger for the binary data than those for the normal data.

For comparison, the simulated normal data set was analyzed using the ML method (XU and ATCHLEY 1995). Figure 6 shows the likelihood profile along the chromosome when a single QTL was fitted to the model. We can see that the likelihood-ratio profile shows two peaks, the higher peak being at position 25 cM and the lower peak at position 76 cM, overlapping with the true locations of the two simulated QTL (at 25 and 75 cM, respectively). This result is consistent with results of the reversible jump MCMC analysis. The ML estimates of variances of the first QTL are $\hat{\sigma}_{a_1}^2 = 0.8253$ and $\hat{\sigma}_{d_1}^2 = 0.8900$. The estimated variances of the second QTL are $\hat{\sigma}_{a_2}^2 = 0.7839$ and $\hat{\sigma}_{d_2}^2 = 0.5667$. The ML estimates of QTL variances are not so close to the corresponding true values as the estimated posterior means in the reversible jump MCMC analysis.

DISCUSSION

We have presented here a Bayesian procedure that allows detection of multiple QTL for both normally distributed and binary traits under IBD-based variance component analysis. The proposed Bayesian procedure is implemented via a reversible jump MCMC algorithm, which enables moves to be made between models with different numbers of QTL. We model complex binary traits under the classical threshold model of quantitative genetics (LYNCH and WALSH 1998). The Bayes method of mapping QTL for binary traits is developed based

TABLE 3

The highest posterior QTL intensity intervals, Bayesian estimates of QTL locations, the additive and dominance variances of the QTL, and polygenic and residual variances

| Data type | Interval (cM) | Sum of the QTL intensity | QTL location (cM) | μ | $\sigma_{a_j}^2$ | $\sigma_{d_j}^2$ | σ_A^2 | σ_c^2 |
|-----------|---------------|--------------------------|---------------------|---------------------|--------------------|--------------------|--------------------|--------------------|
| Normal | ~18–29 | 0.9967 | 24.3013 (1.1678) | 0.028 (0.0559) | 0.7101 (0.2480) | 0.6886 (0.1994) | 0.8696 (0.2415) | 1.1587 (0.1419) |
| | ~72–84 | 0.9938 | 77.0253 (2.2326) | | 0.5701 (0.1954) | 0.3362 (0.1534) | | |
| Binary | ~17–30 | 0.9405 | 23.2892 (2.1665) | -0.0396 (0.0816) | 0.8443 (0.4618) | 1.0681 (0.5966) | 1.5801 (0.7561) | |
| | ~64–86 | 0.9101 | 74.4903 (4.7986) | | 0.4493 (0.3237) | 0.7652 (0.5289) | | |

Posterior standard errors of the estimates are given in parentheses.

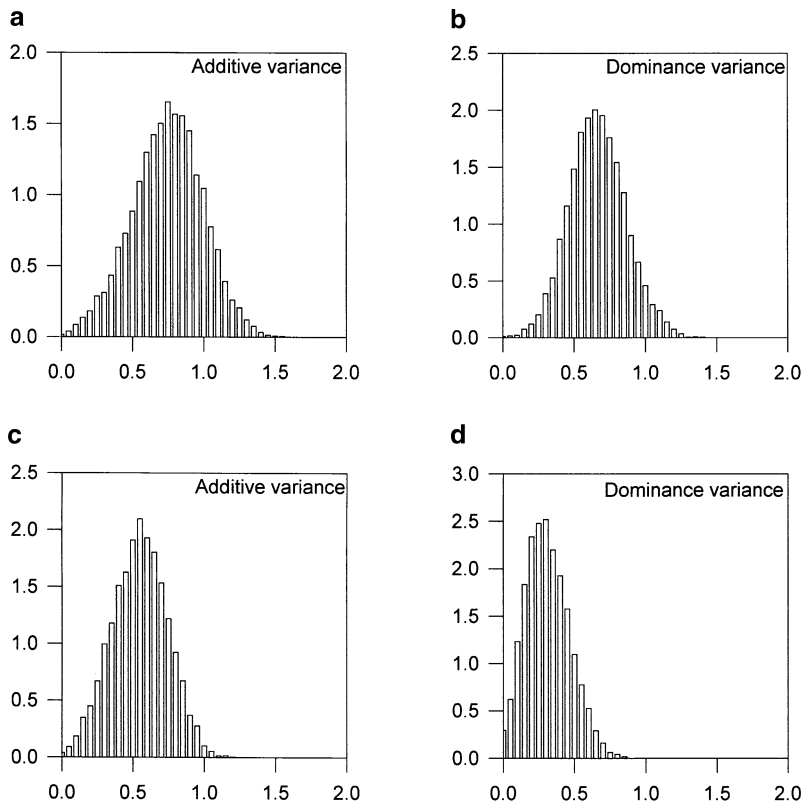


FIGURE 4.—Approximate posterior distributions of the QTL additive and dominance variances for normally distributed data. (a) The first QTL additive variance determined from interval 18 cM to ~ 29 cM. (b) The first QTL dominance variance determined from interval 18 cM to ~ 29 cM. (c) The second QTL additive variance determined from interval 72 cM to ~ 84 cM. (d) The second QTL dominance variance determined from interval 72 cM to ~ 84 cM.

on the idea of data augmentation. This treatment provides an easy way to generate the values of the normally distributed latent variable, which in turn allows the use of Bayesian mapping for normally distributed traits. The methodology can be easily generalized to multiple ordered categorical traits (ALBERT and CHIB 1993).

Bayesian mapping statistics have been well studied in the context of line-crossing experiments (SATAGOPAN and YANDELL 1996; SATAGOPAN *et al.* 1996; SILLANPÄÄ and ARJAS 1998, 1999; STEPHENS and FISCH 1998), in which the parameter of primary interest is the average effect of allelic substitution (the first moment). The variance of the allelic substitution is given as a prior and no further inference is made to the prior variance. When applied to complex pedigrees, the Bayes method can be formulated to fit different models of QTL variation (*e.g.*, biallelic or normal QTL effects; HEATH 1997; UIMARI and HOESCHELE 1997; BINK *et al.* 1998). In contrast to line-crossing experiments in which the number of alleles at the QTL is known exactly, in outbred populations, the number of alleles per locus is rarely known. In most cases, we can assume two alleles in a relatively homogeneous population, as made by HEATH (1997) and UIMARI and HOESCHELE (1997). As a consequence, one must also estimate the allelic frequency to infer the genetic variance via $\hat{\sigma}_a^2 = 2\hat{p}(1 - \hat{p})\hat{a}^2$. Although the biallelic model may be proper in most situations, it can potentially fail if the actual number of polymorphic alleles is not two. The IBD-based variance component Bayesian mapping goes up another level in the hierar-

chical model by treating the variance as the parameter of interest. As a result, information about the number of alleles is not required, which has significantly improved the robustness of the Bayes method.

The IBD-based variance component approach of QTL mapping is based on the proportion of alleles with identity-by-descent shared by relatives. Inference on IBD matrices does not need the information of marker linkage phases. The IBD states of markers are calculated in advance and remain as *a priori* in the Bayesian procedure. With the proposed method, we can completely avoid the complicated problem concerning sampling of additive and dominance effects of QTL and marker and QTL genotypes. The IBD-based variance component Bayesian mapping is not limited to nuclear families. In the simulation study, we used full-sib families to demonstrate the utility of the method solely for the purpose of simplicity. In theory, the Bayes method described in this article can be applied easily to extended or complex pedigrees by using a general algorithm to compute the IBD proportion shared by an arbitrary relative pair. Such a general algorithm has been given by ALMASY and BLANGERO (1998). For large and complex pedigrees, however, the computational burden may be prohibitive, due to the need for computation of the variance-covariance matrix \mathbf{V} and its inverse.

In the simulation study, we put two QTL on a single chromosome and 12 independent loci on other unspecified chromosomes as a "polygene." Our model always includes a polygenic term to absorb background QTL.

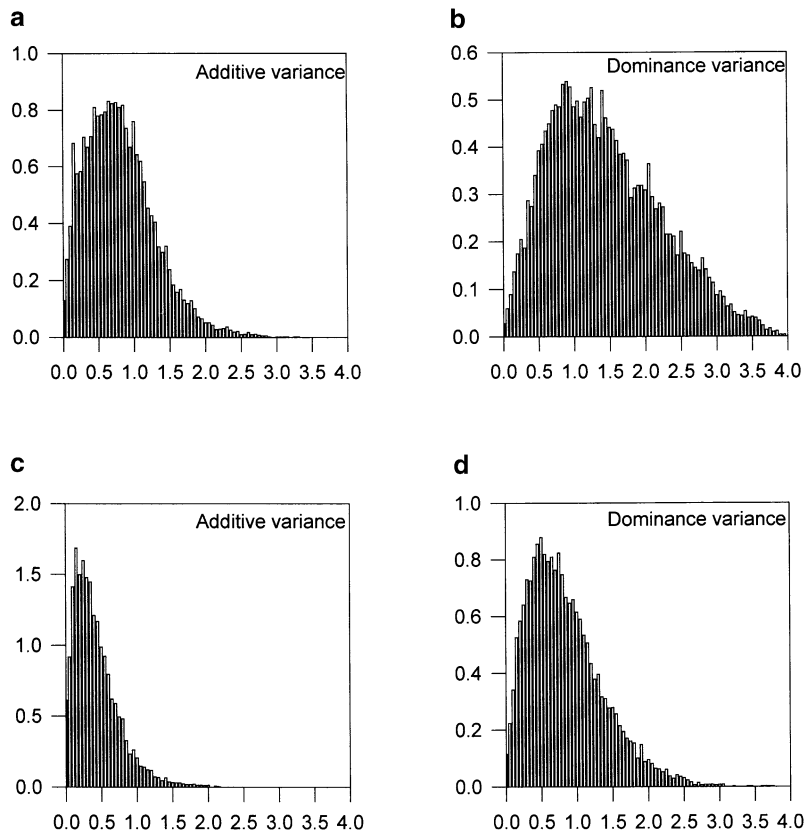


FIGURE 5.—Approximate posterior distributions of the QTL additive and dominance variances for binary data. (a) The first QTL additive variance determined from interval 17 cM to ~ 30 cM. (b) The first QTL dominance variance determined from interval 17 cM to ~ 30 cM. (c) The second QTL additive variance determined from interval 64 cM to ~ 86 cM. (d) The second QTL dominance variance determined from interval 64 cM to ~ 86 cM.

The polygenic variance varies depending on the model effects fitted. Without this polygenic term, variances due to QTL on unmarked chromosomes or chromosome regions will join the residual variance. A large residual variance will reduce the power of QTL detection. In practice, we can take one of three strategies for QTL mapping. One strategy is to analyze one chromosome at a time. Marker information on other chromosomes can be completely ignored. In this case, variances of QTL on other chromosomes will be collected by the polygene (not the residual). The second strategy is to include the previously identified QTL (at other chromo-

somes) in the model instead of being absorbed into the polygenic term. The other strategy is to map QTL simultaneously for all chromosomes. In the latter situation, if the marker map is sufficiently dense, the polygenic term can be ignored because eventually all QTL will be identified and their effects will be explicitly estimated. These three mapping strategies may produce similar results.

The reversible jump MCMC sampler performed well for the simulated normally distributed and binary data. A plot of the changes in the number of QTL against the number of iterations for the binary data used in the simulation study is presented in Figure 7. It shows that the reversible jump MCMC algorithm mixes well over the number of QTL. A similar plot was obtained for the normally distributed data. We detected no influence of initial values of unknowns on the mixing of the MCMC sampler. For example, with the simulated data, starting with $l_0 = 4$, the QTL number l quickly dropped to 1 after several hundred iterations and subsequently behaved the same as starting with $l_0 = 1$. The mixing behavior of the number of QTL is greatly affected by the prior distributions for QTL additive and dominance variances (STEPHENS and FISCH 1998). As expected, increasing the upper bounds of the uniform prior distributions has the general effect of decreasing the acceptance rates of adding a new QTL to the model and deleting a QTL from the model. With the upper bound $c = 4.0$ used in our simulation studies, the observed acceptance

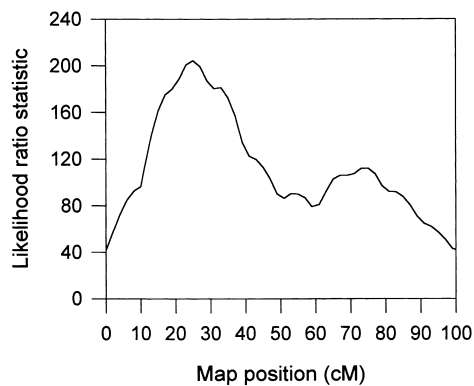


FIGURE 6.—Likelihood-ratio profile of ML mapping from normally distributed data. There are two QTL, located at positions 25 and 75 cM, respectively.

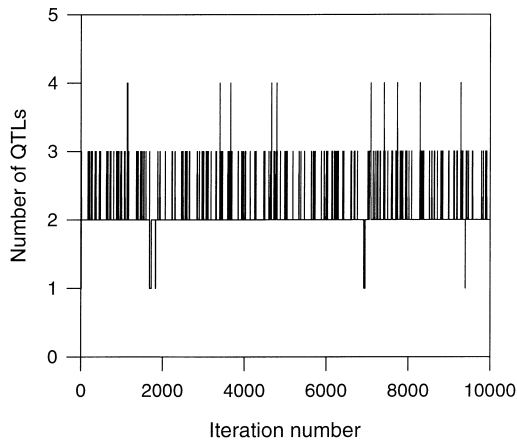


FIGURE 7.—The trace of the number of QTL for the binary data, for 10^4 saved samples after burn-in.

rates for both adding and deleting steps were $\sim 0.4\%$ for both the simulated normally distributed and binary data. These acceptance rates were similar to those in SILLANPÄÄ and ARJAS (1998). The acceptance rates for both adding and deleting steps increased to 3% when the upper bound was set at $c = 1.0$. By comparison, the acceptance proportions for updating QTL locations and model effects θ were found to be rather high in general and were $\sim 80\%$ for the simulated data.

A major implementation issue in MCMC is to determine the effective sample sizes. This issue is related to the assessment of convergence of the chain, the serial correlation between the samples, and the burn-in period. For a single long chain, one can examine time series graphs of a simulated sequence to judge whether the chain is sufficiently long (GEYER 1992). Subsampling the chain at regular intervals is a common practice to reduce the serial correlation between the samples. With the reversible jump MCMC algorithm, however, it is difficult to calculate autocorrelation of the MCMC samples for all parameters because the dimension keeps changing from one cycle to another. When the dimension changes, the identities of the QTL also change. The parameters in one cycle of the iteration may be different from those in the next cycle of iteration. In our simulation studies, therefore, we empirically determined the lengths of the chain, subsampling intervals and the burn-in period. We used the plots of the changes in the number of QTL against the number of iterations to determine an approximate burn-in period. The length of subsampling intervals was chosen to eliminate obvious changing trends for all parameters.

We only use marker information to infer the IBD matrices of QTL and allow multiple QTL located in the same marker interval in the proposed method. Alternatively, we can infer the IBD matrices of a QTL considered by utilizing the IBD information of all markers and other existing QTL in the same chromosome and assume at most one QTL in any marker interval is accept-

able. However, our treatment simplifies the MCMC algorithm and does not result in a significant reduction in efficiency particularly when the marker map is dense. It has been shown that, conditional on the IBD states of a marker locus, the IBD matrices of a QTL on one side of the marker are not correlated to those of a QTL on the other side (XU and ATCHLEY 1995). In case of two tightly linked QTL in the same marker interval, the IBD matrices of these two QTL are always highly correlated no matter what methods are used to infer them. The correlation between two QTL can be completely eliminated by the assumption that any marker interval includes at most one QTL. Under this assumption, the acceptance probabilities for adding one QTL and deleting one QTL should be modified (STEPHENS and FISCH 1998).

Finally, Bayesian statistics have the inherent flexibility introduced by their incorporation of multiple levels of randomness and the resultant ability to combine information from different sources. Therefore, the Bayesian approach could be extended to allow more complicated models under more complicated data structures, *e.g.*, the epistatic model. When a new QTL is proposed, its interactive effects with all existing QTL should be proposed and, if accepted, the epistatic effects should be included in the model. A QTL is finally added to the model if at least one effect caused by the QTL is accepted. This will involve additional reversible jumps on the dimension of the model even if the number of QTL remains the same.

We thank Dr. Damian Gessler for his helpful comments on the manuscript. We also thank two anonymous reviewers for their critical comments on an earlier version of the manuscript. This research was supported by the National Institutes of Health Grant GM55321 and the United States Department of Agriculture National Research Initiative Competitive Grants Program 97-35205-5075 to S.X.

LITERATURE CITED

- ALBERT, J. H., and S. CHIB, 1993 Bayesian analysis of binary and polychotomous response data. *J. Am. Stat. Assoc.* **88**: 669–679.
- ALMASY, L., and J. BLANGERO, 1998 Multipoint quantitative-trait linkage analysis in general pedigrees. *Am. J. Hum. Genet.* **62**: 1198–1211.
- AMOS, C. I., 1994 Robust variance-components approach for assessing genetic linkage in pedigrees. *Am. J. Hum. Genet.* **54**: 535–543.
- BINK, M. G. A. M., L. L. G. JANSSE and R. L. QUAAAS, 1998 Mapping a polyallelic quantitative trait locus using simulated tempering. *Proceedings of the 6th World Congress on Genetics Applied to Livestock Production Science, Armidale, Australia, Vol. 26*, pp. 277–280.
- CHAN, J. S. K., and A. C. KUK, 1997 Maximum likelihood estimation for probit-linear mixed models with correlated random effects. *Biometrics* **88**: 86–97.
- DEVROYE, T., 1986 *Non-uniform Random Variable Generation*. Springer-Verlag, New York.
- DUGGIRALA, R., J. T. WILLIAMS, S. WILLIAMS-BLANGERO and J. BLANGERO, 1997 A variance component approach to dichotomous trait linkage analysis using a threshold model. *Genet. Epidemiol.* **14**: 987–992.
- FULKER, D. W., and L. R. CARDON, 1994 A sib-pair approach to

- interval mapping of quantitative trait loci. *Am. J. Hum. Genet.* **54**: 1092–1103.
- FULKER, D. W., S. S. CHERNY and L. R. CARDON, 1995 Multipoint interval mapping of quantitative trait loci, using sib pairs. *Am. J. Hum. Genet.* **56**: 1224–1233.
- GEMAN, S., and D. GEMAN, 1984 Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Machine Intell.* **6**: 721–741.
- GESSLER, D. D. G., and S. XU, 1996 Using the expectation or the distribution of identity-by-descent for mapping quantitative trait loci under the random model. *Am. J. Hum. Genet.* **59**: 1382–1390.
- GEYER, C. J., 1992 Practical Markov chain Monte Carlo. *Stat. Sci.* **7**: 473–511.
- GOLDGAR, D. E., 1990 Multipoint analysis of human quantitative genetic variation. *Am. J. Hum. Genet.* **47**: 957–967.
- GREEN, P. J., 1995 Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82**: 711–732.
- GRIGNOLA, F. E., I. HOESCHELE and B. TIER, 1996 Mapping quantitative trait loci via residual maximum likelihood: I. Methodology. *Genet. Sel. Evol.* **28**: 479–490.
- HACKETT, C. A., and J. I. WELLER, 1995 Genetic mapping of quantitative trait loci for traits with ordinal distributions. *Biometrics* **51**: 1252–1263.
- HALEY, C. S., and S. A. KNOTT, 1992 A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* **69**: 315–324.
- HASTINGS, W. K., 1970 Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**: 97–109.
- HEATH, S. C., 1997 Markov chain Monte Carlo segregation and linkage analysis of oligogenic models. *Am. J. Hum. Genet.* **61**: 748–760.
- JANSEN, R. C., 1993 Interval mapping of multiple quantitative trait loci. *Genetics* **135**: 205–211.
- KAO, C. H., Z-B. ZENG and R. D. TEASDALE, 1999 Multiple interval mapping for quantitative trait loci. *Genetics* **152**: 1203–1216.
- KRUGLYAK, L., and E. S. LANDER, 1995 Complete multipoint sib-pair analysis of qualitative and quantitative traits. *Am. J. Hum. Genet.* **57**: 439–454.
- LYNCH, M., and B. WALSH, 1998 *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, Sunderland, MA.
- MENDELL, N. R., and R. C. ELSTON, 1974 Multifactorial qualitative traits: genetic analysis and prediction of recurrence risks. *Biometrics* **30**: 41–57.
- METROPOLIS, N., A. W. ROSENBLUTH, M. N. ROSENBLUTH, A. H. TELLER and E. TELLER, 1953 Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21**: 1087–1091.
- RAO, S., and S. XU, 1998 Mapping quantitative trait loci for ordered categorical traits in four-way crosses. *Heredity* **81**: 214–224.
- REBAI, A., 1997 Comparison of methods for regression interval mapping in QTL analysis with non-normal traits. *Genet. Res.* **69**: 69–74.
- SATAGOPAN, J. M., and B. S. YANDELL, 1996 Estimating the number of quantitative trait loci via Bayesian model determination. Special Contributed Paper Session on Genetic Analysis of Quantitative Traits and Complex Diseases, Biometric Section, Statistical Meeting, Chicago, IL.
- SATAGOPAN, J. M., B. S. YANDELL, M. A. NEWTON and T. G. OSBORN, 1996 A Bayesian approach to detect quantitative trait loci using Markov chain Monte Carlo. *Genetics* **144**: 805–816.
- SCHORK, N. J., 1993 Extended multipoint identity-by-descent analysis of human quantitative traits: efficiency, power, and modeling considerations. *Am. J. Hum. Genet.* **53**: 1306–1319.
- SILLANPÄÄ, M. J., and E. ARJAS, 1998 Bayesian mapping of multiple quantitative trait loci from incomplete inbred line cross data. *Genetics* **148**: 1373–1388.
- SILLANPÄÄ, M. J., and E. ARJAS, 1999 Bayesian mapping of multiple quantitative trait loci from incomplete outbred offspring data. *Genetics* **151**: 1605–1619.
- 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.
- UIMARI, P., and I. HOESCHELE, 1997 Mapping linked quantitative trait loci using Bayesian analysis and Markov chain Monte Carlo algorithms. *Genetics* **146**: 735–743.
- VISSCHER, P. M., C. S. HALEY and S. A. KNOTT, 1996 Mapping QTL for binary traits in backcross and F_2 populations. *Genet. Res.* **68**: 55–63.
- XU, S., and W. R. ATCHLEY, 1995 A random model approach to interval mapping of quantitative trait loci. *Genetics* **141**: 1189–1197.
- XU, S., and W. R. ATCHLEY, 1996 Mapping quantitative trait loci for complex binary diseases using line crosses. *Genetics* **143**: 1417–1424.
- XU, S., and D. D. G. GESSLER, 1998 Multipoint genetic mapping of quantitative trait loci using a variable number of sibs per family. *Genet. Res.* **71**: 73–83.
- YI, N., and S. XU, 1999a Mapping quantitative trait loci for complex binary traits in outbred populations. *Heredity* **82**: 668–676.
- YI, N., and S. XU, 1999b A random model approach to mapping quantitative trait loci for complex binary traits in outbred populations. *Genetics* **153**: 1029–1040.
- ZENG, Z-B., 1994 Precision mapping of quantitative trait loci. *Genetics* **136**: 1457–1468.

Communicating editor: J. B. WALSH