

Bayesian Model Choice and Search Strategies for Mapping Interacting Quantitative Trait Loci

Nengjun Yi,^{*,†,1} Shizhong Xu[‡] and David B. Allison^{*,†,§}

^{*}Department of Biostatistics, [†]Section on Statistical Genetics, [§]Clinical Nutrition Research Center, University of Alabama, Birmingham, Alabama 35294 and [‡]Department of Botany and Plant Sciences, University of California, Riverside, California 92521

Manuscript received March 10, 2003
Accepted for publication May 19, 2003

ABSTRACT

Most complex traits of animals, plants, and humans are influenced by multiple genetic and environmental factors. Interactions among multiple genes play fundamental roles in the genetic control and evolution of complex traits. Statistical modeling of interaction effects in quantitative trait loci (QTL) analysis must accommodate a very large number of potential genetic effects, which presents a major challenge to determining the genetic model with respect to the number of QTL, their positions, and their genetic effects. In this study, we use the methodology of Bayesian model and variable selection to develop strategies for identifying multiple QTL with complex epistatic patterns in experimental designs with two segregating genotypes. Specifically, we develop a reversible jump Markov chain Monte Carlo algorithm to determine the number of QTL and to select main and epistatic effects. With the proposed method, we can jointly infer the genetic model of a complex trait and the associated genetic parameters, including the number, positions, and main and epistatic effects of the identified QTL. Our method can map a large number of QTL with any combination of main and epistatic effects. Utility and flexibility of the method are demonstrated using both simulated data and a real data set. Sensitivity of posterior inference to prior specifications of the number and genetic effects of QTL is investigated.

MOST complex traits important to evolution, animal and plant breeding, and medical genetics are influenced by multiple genetic and environmental inputs. It has been speculated that epistases (interactions among two or more genetic loci) are ubiquitous in the genetic control and evolution of complex traits (CHEVERUND and ROUTMAN 1995; FRANKEL and SCHORK 1996; LYNCH and WALSH 1998; WADE 2001). However, the contribution of epistasis to genetic variance components is difficult to estimate by traditional quantitative genetic approaches, except when clones are available (WU 1996). With the availability of saturated linkage maps of molecular markers, it is possible to identify genes that significantly contribute to the variation of complex traits and to evaluate not only the marginal (main) effects of the identified quantitative trait loci (QTL) but also the epistatic effects (KAO *et al.* 1999; SEN and CHURCHILL 2001; YI and XU 2002). An in-depth understanding of the epistatic interactions is important in the genetic study of complex traits. It has been described from both empirical and theoretical aspects that the potential improvement in power in discovering and localizing genes can be gained by allowing epistatic interaction when it is present (GAUDERMAN and THOMAS 2000; KAO and ZENG 2002). Although most gene-map-

ping studies have not incorporated epistatic effects into the models, epistatic interactions have been described in recent studies in different species (*e.g.*, LARK *et al.* 1995; FIJNEMAN *et al.* 1996; ROUTMAN and CHEVERUD 1997; YU *et al.* 1997; FERNANDEZ *et al.* 2000; SUGIYAMA *et al.* 2001; DONG *et al.* 2003).

Statistical methodology for epistatic modeling in gene mapping is in its infancy. Early methods of QTL mapping were built mainly on single-QTL models in which one QTL is fitted to the model at a time and only marginal QTL effects are detected (*e.g.*, LANDER and BOTSTEIN 1989; JANSEN and STAM 1994; ZENG 1994). These approaches are not directly extendable to analyzing epistatic effects. To evaluate epistatic interactions, multiple-QTL models must be used and the marginal effects of multiple putative QTL and associated epistatic effects must be modeled simultaneously. To date, statistical methods for mapping multiple QTL with epistasis were developed mainly on the basis of non-Bayesian methods. Most of these methods either first conduct a search for single-gene main effects and then search for interactions among those loci with significant individual effects, which may make sense in some contexts (*e.g.*, CULVERHOUSE *et al.* 2002), or fit two QTL at a time and the two-dimensional searches along the genome are used to detect QTL and estimate two-locus interactions (*e.g.*, HALEY and KNOTT 1992; WANG *et al.* 1999). KAO *et al.* (1999) and ZENG *et al.* (1999) developed a multiple-interval mapping (MIM) approach to mapping multiple

¹Corresponding author: Department of Biostatistics, University of Alabama, Birmingham, AL 35294-0022. E-mail: nyi@ms.soph.uab.edu

QTL and estimating epistasis in backcross designs. They used a stepwise model selection approach to progressively add or delete a QTL as well as an epistatic effect in the model. Recently, CARLBORG *et al.* (2000) proposed a genetic algorithm to reduce the computational demand for simultaneous mapping of multiple interactive QTL. JANNINK and JANSEN (2001) and BOER *et al.* (2002) described a statistical method to map multiple epistatic QTL with one-dimensional genome searches.

There has been a surge of interest in Bayesian approaches to mapping multiple QTL. Earlier Bayesian approaches to QTL mapping were developed to estimate the locations and the effect parameters for multiple QTL with a prespecified number of QTL (STEPHENS and SMITH 1993; SATAGOPAN *et al.* 1996; UIMARI *et al.* 1996). Recently, SEN and CHURCHILL (2001) described a Bayesian method to map interacting QTL, but still assumed a known QTL number. However, the number of QTL is in fact unknown, and hence the dimensionality of the parameter space is also unknown. GREEN (1995) introduced the reversible jump Markov chain Monte Carlo (MCMC) algorithm, a variant of the Metropolis-Hastings algorithm for problems where the dimension of the parameter space is itself one of the unknowns. Using the reversible jump MCMC method, we can treat the number of QTL as a random variable and generate its posterior distribution. The Bayesian methods implemented via the reversible jump MCMC have been applied to map QTL in both line-crossing designs (SATAGOPAN and YANDELL 1996; SILLANPÄÄ and ARJAS 1998, 1999; STEPHENS and FISCH 1998; YI and XU 2000) and pedigree data (HEATH 1997; UIMARI and HOESCHELE 1997; XU and YI 2000; UIMARI and SILLANPÄÄ 2001; YI and XU 2001; CORANDER and SILLANPÄÄ 2002). Several of these approaches have been successfully applied to real data analyses (SATAGOPAN and YANDELL 1996; DAW *et al.* 1999; HURME *et al.* 2000; YUAN *et al.* 2000; BINK *et al.* 2002). YI and XU (2002) extended the reversible jump algorithm to map interacting QTL. Their method included all possible epistatic effects in the model and applied the reversible jump only to the number of QTL. With a large or moderate number of putative QTL, however, including all possible epistatic effects must accommodate a very large number of potential parameters, which causes problems in determination of the genetic model.

In this article, we treat the number of QTL as a random variable. Under the current number of QTL, we introduce indicator variables to represent the number of effects and the actual effects presented in the model. We use the methodology of the Bayesian model and variable selection to develop strategies for searching for multiple QTL with complex epistatic patterns and estimating the associated parameters simultaneously. Incorporation of epistasis not only necessitates determining the genetic model conditional on the number of QTL, but also necessitates a new algorithm for inferring

the number of QTL since adding or deleting QTL will depend on the presence of either main or epistatic effects. An efficient reversible jump MCMC algorithm is proposed to update the number of QTL under complicated epistatic models. We also develop a reversible jump MCMC method to infer main and epistatic effects conditional on the number of QTL.

METHODS

Epistatic model: We describe the method primarily for a mapping population with only two segregating genotypes, *e.g.*, a backcross, double haploid lines (DHLs), or recombinant inbred lines (RILs). Assume that a complex trait is affected by l QTL in a mapping population. Among the l QTL, some may exhibit only main effects, some may show both main effects and epistasis, and others may show only epistasis. For a continuously distributed trait, the observed phenotypic value of individual i , y_i , can be described by the following linear model,

$$y_i = \mu + \sum_{q=1}^l \gamma_q x_{iq} a_q + \sum_{q_1 < q_2}^l \gamma_{q_1 q_2} x_{iq_1} x_{iq_2} b_{q_1 q_2} + e_i, \quad i = 1, 2, \dots, n, \quad (1)$$

where n is the number of individuals in the mapping population, μ is the overall mean, l is the number of QTL on the genome, a_q is the main effect of putative QTL q , $b_{q_1 q_2}$ is the epistatic effect between putative QTL q_1 and q_2 , x_{iq} is the indicator variable denoting the genotype of putative QTL q for individual i and is defined by 0.5 or -0.5 for the two genotypes in the mapping population (KAO *et al.* 1999; ZENG *et al.* 1999), e_i is the residual error assumed to follow $N(0, \sigma_e^2)$, γ_q is a binary indicator variable for main effect of putative QTL q , taking value one if QTL q has main effect and zero otherwise, and $\gamma_{q_1 q_2}$ is a binary indicator variable for epistatic effect between QTL q_1 and q_2 , taking value one if QTL q_1 and q_2 interact and zero otherwise. Hereafter, we call these binary indicator variables effect indicators. Apparently, an effect indicator γ taking value zero is equivalent to setting zero to the corresponding effect or removing the corresponding effect from the model. As will be seen, however, the introduction of the effect indicator variables facilitates our description and analysis of the problem. Finally, it is worth noting that each QTL included in model (1) has at least one nonzero effect indicator.

In practice, we observe the phenotypic values $\mathbf{y} = \{y_i\}_{i=1}^n$ and a set of marker genotypes $\mathbf{M} = \{m_{ik}\}_{i=1, k=1}^{n, K}$, where K is the number of markers. For a mapping population with only two segregating genotypes, m_{ik} takes one of two values, *e.g.*, 0 and 1, denoting two different genotypes. Assume that marker linkage maps have been developed on the basis of observed marker data so that the locations of markers on each chromosome are known *a priori*. The observed marker data are used to

infer the probability distribution of the unobserved putative QTL. Since it may be difficult to distinguish multiple QTL on a marker interval (*e.g.*, LYNCH and WALSH 1998), we assume that there is at most one QTL on any marker interval although this is not a fundamental requirement for our method. For a dense linkage map, this assumption is reasonable.

Based on the observed data, our aim is to search all types of QTL and to infer their locations, main effects, and epistatic effects simultaneously. There are several special statistical problems, which complicate mapping multiple interacting QTL. First, one does not know *a priori* how many QTL to expect for a given trait. Therefore, the number of QTL has to be treated as an unknown parameter and inferred along with other parameters in the model. Under the epistatic model, the l putative QTL are chosen on the basis of either their significant main effects or their epistatic effects; this complicates inference about the number of QTL. Second, the locations of multiple QTL are unknown. Therefore, one needs to search for QTL on the entire genome. Third, mapping multiple interacting QTL requires accommodating a very large number of potential genetic effects, even when one assumes only a moderate number of QTL. For example, if there are 10 QTL, 10 parameters would be required to model the main effects, and 45 parameters to model digenic epistasis. In fact, many of these possible effects, particularly the epistatic effects, are not significantly different from zero and thus should be removed from the model. However, one does not know *a priori* which effects are significantly different from zero. Moreover, in the traditional frequentist framework, such a *backward selection* procedure model would increase the type I error rate if not accommodated in some way (POPE and WEBSTER 1972). In this article, we develop a Bayesian approach to search all types of QTL and infer the number, locations, and main and epistatic effects simultaneously.

Prior distributions and joint posterior distribution:

One can view the problem of mapping epistatic QTL as one of model determination and variable selection. Conditional on the observed data \mathbf{y} and \mathbf{M} , we seek to determine the number of genes and associated effect indicator variables and to estimate the locations of identified genes and the model parameters. Hereafter, we denote the number of QTL by l , locations by λ , vector of the effect indicators by γ , main effects by \mathbf{a} , epistatic effects by \mathbf{b} , genotypic indicator for the genotypes of putative QTL by \mathbf{x} , and all model parameters by $\theta = (\mathbf{a}, \mathbf{b}, \mu, \sigma_e^2)$. In some situations, the genotypes of some markers and the phenotypic values of some individuals are missing. For convenience of description, all marker genotypes and phenotypic values are assumed known, although they can be imputed (see the next section).

We employ a Bayesian framework in which statistical inference is based on the joint posterior distribution of

all unknowns. The joint posterior distribution of all unknowns, $(l, \gamma, \lambda, \theta, \mathbf{x})$, can be written as

$$p(l, \gamma, \lambda, \theta, \mathbf{x} | \mathbf{y}, \mathbf{M}) \propto p(\mathbf{y} | l, \gamma, \theta, \mathbf{x}) p(\mathbf{x} | l, \lambda, \mathbf{M}) p(l, \gamma, \lambda, \theta). \quad (2)$$

The first term in this equation is the conditional distribution of the phenotypic data given all unknowns and has the form

$$\begin{aligned} p(\mathbf{y} | l, \gamma, \theta, \mathbf{x}) &= \prod_{i=1}^n p(y_i | l, \gamma, \theta, \mathbf{x}_i) \\ &= \prod_{i=1}^n (2\pi\sigma_e^2)^{-n/2} \exp\left\{-\frac{(y_i - g_i)^2}{2\sigma_e^2}\right\}, \end{aligned}$$

where $g_i = \mu + \sum_{q=1}^l \gamma_q x_{iq} a_q + \sum_{q_1 < q_2} \gamma_{q_1 q_2} x_{iq_1} x_{iq_2} b_{q_1 q_2}$.

The second term $p(\mathbf{x} | l, \lambda, \mathbf{M})$ is the probability distribution of QTL genotypic indicator conditional on the number, locations of QTL and marker genotypes. Assuming that there is at most one QTL on any marker interval, $p(\mathbf{x} | l, \lambda, \mathbf{M})$ can be factorized into the products $p(\mathbf{x} | l, \lambda, \mathbf{M}) = \prod_{i=1}^n \prod_{q=1}^l p(x_{iq} | \lambda_{iq}, m_{iq}^L, m_{iq}^R)$, where m_{iq}^L and m_{iq}^R represent the left and the right flanking marker genotypes, respectively. Methods with which to calculate $p(x_{iq} | \lambda_{iq}, m_{iq}^L, m_{iq}^R)$ in different experimental designs are described in JIANG and ZENG (1997).

The third term $p(l, \gamma, \lambda, \theta)$ in Equation 2 is the joint prior distribution of the parameters $(l, \gamma, \lambda, \theta)$. Assuming prior conditional independence of the parameters, we can factorize the joint prior distribution into the following products:

$$p(l, \gamma, \lambda, \theta) = p(l) p(\gamma | l) p(\lambda | l) p(\mu) p(\sigma_e^2) \prod_{q: \gamma_q \neq 0}^l p(a_q) \prod_{q_1 < q_2: \gamma_{q_1 q_2} \neq 0}^l p(b_{q_1 q_2}).$$

The prior distribution of l , $p(l)$, is chosen to be uniform $(0, l_{\max})$, or Poisson (l_{mean}) , for a prespecified integer l_{\max} or Poisson mean l_{mean} . To specify the conditional prior for the vector of QTL positions λ , $p(\lambda | l)$, we assume that (1) a QTL is uniformly distributed across the entire genome and (2) there is at most one QTL on a marker interval. Under these two assumptions, the conditional prior distribution of QTL locations can be calculated as

$$p(\lambda | l) = p(\lambda_1) \prod_{q=2}^l p(\lambda_q | \lambda_1, \dots, \lambda_{q-1}),$$

where $p(\lambda_1)$ is uniform across the entire genome, and $p(\lambda_q | \lambda_1, \dots, \lambda_{q-1})$ is uniform across the regions of the unoccupied marker intervals (without QTL). The prior distribution $p(\gamma | l)$ reflects our subjective belief on the genetic effects conditional on the number of QTL. When no information regarding the genetic architecture of the trait is available, $p(\gamma | l)$ can be decomposed into the products of all elements, *i.e.*, $p(\gamma | l) = \prod_{q=1}^l p(\gamma_q) \prod_{q_2 > q_1}^l p(\gamma_{q_1 q_2})$, and a natural choice for $p(\gamma_q)$ and $p(\gamma_{q_1 q_2})$ is uniform at two states of 0 and 1. This prior specification gives the same prior weight to all models conditional on the number of QTL and is widely used as a noninformative prior in variable selection problems (NTZOUFRAS 1999). The priors for the overall mean μ

and the QTL effects a_q and $b_{q_1q_2}$ are assumed to be independently normal, *i.e.*, $\mu \sim N(\eta_0, \tau_0^2)$, $a_q \sim N(\eta, \tau^2)$, and $b_{q_1q_2} \sim N(\eta, \tau^2)$, with prespecified prior means η_0 and η and variances τ_0^2 and τ^2 . The prior for σ_e^2 is assumed to be uniform on a predefined interval. The lower and upper bounds for σ_e^2 are usually set to zero and the phenotypic variance present in the data, respectively.

Posterior computations: The calculation of the above joint posterior distributions is analytically intractable, and thus the MCMC approach is required to obtain observations from the joint posterior distribution. We use the reversible jump MCMC method of GREEN (1995) to generate samples from the joint posterior distribution. Specifically, our reversible jump MCMC algorithm proceeds as follows:

- a. Update the missing marker genotypes and phenotypic values.
- b. Update the model parameters (θ).
- c. Update the genotypic indicators (\mathbf{x}).
- d. Update the locations of QTL (λ).
- e. Update the effect indicator (γ): add or delete a main or epistatic effect.
- f. Add one new QTL with main effect or epistatic effects between existing QTL, or delete an existing QTL.
- g. Add two new QTL with only an epistatic effect between themselves, or delete two existing QTL.

It is noted that deleting an effect and deleting one or two QTL from the model may result in some QTL having all corresponding effect indicators equal to zero. If it occurs, we remove such QTL from the model. A complete pass through these steps is referred to as a sampling iteration. The algorithm starts from an initial point $(l^{(0)}, \gamma^{(0)}, \lambda^{(0)}, \theta^{(0)}, \mathbf{x}^{(0)})$ and then proceeds to update each of the groups of the unknowns in turn until a certain criterion of convergence is reached (BROOKS and GIUDICI 2000). Once the sampler converges, the posterior samples $\{(l^{(t)}, \gamma^{(t)}, \lambda^{(t)}, \theta^{(t)}, \mathbf{x}^{(t)}) : t = 1, 2, \dots\}$ are random drawings from the joint posterior distribution $p(l, \gamma, \lambda, \theta, \mathbf{x} | \mathbf{y}, \mathbf{M})$, and samples of individual parameters can be regarded as random drawings from the appropriate marginal posterior distribution. These posterior samples are used to draw inferences about parameters of interest.

Except for the effect indicator vector and the number of QTL, other unknowns do not result in a change in the dimension of the parameter space. Therefore, these unknowns can be updated using traditional Metropolis-Hastings (M-H) algorithms (GELMAN *et al.* 1995). Conditional on l, γ , and \mathbf{x} , model (1) is a conventional linear model, and thus the model parameters θ can be generated using the Gibbs sampler (GELMAN *et al.* 1995; YI and XU 2002). Given l, γ, \mathbf{x} , and θ , the missing phenotypic values can be sampled from normal distributions (GELMAN *et al.* 1995). The missing marker genotypes and the genotypic indicators of QTL can be updated

from discrete distributions using the Gibbs sampler (YI and XU 2000). We use our previous method developed in YI and XU (2001, 2002) to update the locations of QTL, but with the assumption of there being at most one QTL at any marker interval under consideration. To update λ_q , we propose a new location λ_q^* from uniform $(\lambda_q - d, \lambda_q + d)$, where d is a predetermined tuning parameter, and generate genotypic indicators at the new location for all individuals. The proposals for the new location and the genotypic indicators are then accepted or rejected using the M-H algorithm (YI and XU 2002). Our updating scheme updates the locations and the genotypic indicators of QTL jointly and results in a reasonable acceptance rate and mixing behavior (UIMARI and SILLANPÄÄ 2001; YI and XU 2001, 2002). Updating the effect indicator vector and the number of QTL results in a change in the dimension of the parameter space and thus needs reversible jump steps. We here develop novel reversible jump steps for updating these two parameters as follows.

The effect indicator vector γ represents which effects are present in the model. Conditional on other unknowns, updating γ is a variable selection problem and thus a variety of Bayesian variable selection methods can be applied to update γ (NTZOUFRAS 1999). We here develop a simple reversible jump MCMC algorithm. Assume that the current number of QTL is l , and a total of $l(l+1)/2$ elements are in the vector γ , corresponding to possible QTL effects for model (1), including l main effects and $l(l-1)/2$ digenic epistasis. To update γ , we randomly select one of all $l(l+1)/2$ elements and then propose a change for this element. If the element associated with the epistasis between QTL q_1' and q_2' is selected, for example, then we propose to change γ to γ' where $\gamma'_{q_1'q_2'} = 1 - \gamma_{q_1'q_2'}$ with all other components remaining the same. In other words, we propose to add the epistasis $b_{q_1'q_2'}$ into the model if the effect $b_{q_1'q_2'}$ is out of the current model, or delete the epistasis $b_{q_1'q_2'}$ from the model if the effect $b_{q_1'q_2'}$ is included in the current model. The proposal can be accepted or rejected according to the reversible jump MCMC method (GREEN 1995). A detailed description of the reversible jump algorithm for updating γ can be found in APPENDIX A.

To detect multiple QTL with any combination of main and epistatic effects, we can add four types of QTL into the current model. These four types of QTL include a new QTL with only main effect, a new QTL having only epistatic effects with the current QTL, a new QTL having both main effect and epistatic effects with the current QTL, and two new QTL with only epistatic effects between themselves. We develop two novel reversible jump steps to search for all four types of QTL. Specifically, we use step f to search the first three types of QTL and step g to search the fourth type of QTL. With these two steps, we are able to identify QTL with complex epistatic patterns.

In step f, we randomly decide whether to propose (a)

adding one new QTL with main effect or epistatic effects between existing QTL or (b) deleting an existing QTL with equal probability. Let $j(l+1|l)$ and $j(l-1|l) = 1 - j(l+1|l)$ be the probabilities of choosing either move type. Of course, $j(l-1|l) = 0$ if $l = 0$ and $j(l+1|l) = 0$ if $l = l_{\max}$, where l_{\max} is the maximum value allowed for l , and otherwise we choose $j(l-1|l) = j(l|l-1) = 0.5$, for $l = 2, 3, \dots, l_{\max} - 1$. Adding one QTL requires generating additional parameters for the new QTL, *i.e.*, new location, genotypic indicators, effect indicators, and selected new effects. A detailed description for generating these additional parameters and the acceptance probability are given in APPENDIX B. To delete one QTL, we randomly select one QTL among the existing QTL and then calculate the relevant acceptance probability. The acceptance probability for deleting QTL is also derived in APPENDIX B.

In step g, we make a random choice between attempting to add two new QTL with only epistatic effects or deleting two existing QTL, with probabilities $j(l+2|l)$ and $j(l-2|l) = 1 - j(l+2|l)$, respectively. Similar to step f, we choose $j(l-2|l) = j(l|l-2) = 0.5$, for $l = 2, 3, \dots, l_{\max} - 2$, and $j(l-2|l) = 0$ if $l \leq 1$ and $j(l+2|l) = 0$ if $l \geq l_{\max} - 1$. To add two QTL, we need to generate two new locations on the genome, genotypic indicators at these two locations for all individuals, and an epistatic effect for these two QTL. A detailed sampling scheme for generating these unknowns and the acceptance probability are given in APPENDIX C. The acceptance probability for deleting two QTL is also given in APPENDIX C.

SIMULATION STUDIES AND REAL DATA ANALYSIS

Simulation studies: The applicability of the proposed method was demonstrated by analyzing three simulation experiments. Each experimental population was a backcross containing 300 segregating individuals. In designs I and II, one chromosome of length 150 cM was simulated, and 16 codominant markers were placed on the simulated chromosome with positions shown at the numerical labels of the horizontal axis of Figure 1. Zero and three QTL were simulated in designs I and II, respectively. The overall mean and the residual variance were set to $\mu = 0$ and $\sigma_e^2 = 1$. In design II, three QTL were placed at 25, 63, and 103 cM, respectively, and the nonzero main and epistatic effects of the three loci were $a_1 = -0.70$, $a_2 = 0.75$, and $b_{23} = 1.0$, where a_q is the main effect of QTL q , and $b_{q_1q_2}$ is the epistatic effect between QTL q_1 and q_2 . Therefore, the third QTL had no main effect but interaction with the second QTL. The proportions of phenotypic variance explained by the nonzero effects were $\sim 9, 10, \text{ and } 8\%$, respectively. In design III, four chromosomes with length 100 cM each were simulated, and 11 codominant markers were placed on each chromosome with positions shown at the numerical labels of the horizontal axis of Figure 3.

We simulated seven QTL with complex epistatic patterns controlling the expression of a quantitative trait. The locations of the simulated QTL, the genetic effects (main and epistatic effects), and the heritability explained by each effect are shown in Table 1. The overall proportion of phenotypic variance explained by all QTL was $\sim 55\%$. The overall mean and the residual variance were $\mu = -1$ and $\sigma_e^2 = 1$, respectively. To evaluate the ability of our method for handling missing markers and missing phenotypic values, we randomly generated missing markers of 15% and missing phenotypic values of 10% in design III.

Both designs I and II were replicated five times and the averaged results were reported. For all analyses, the MCMC were started with no QTL in the model, and the same prior distributions for the overall mean and the effect indicator vector were used. We tried to run the MCMC sampler using other starting values for the number of QTL and found the results to be consistent. The prior for the overall mean was $N(0, 2)$. The residual variance took uniform(0, V), where V represents the phenotypic variance observed in each design. The prior for the indicator vector γ was taken as uniform on the space of γ as described earlier. The tuning parameter of proposals in the random-walk Metropolis-Hastings algorithm for updating QTL locations was chosen to be 2.0 cM. Different prior distributions for the number of QTL and the genetic effects were chosen for the analyses of three designs and are described later.

In each analysis, the MCMC sampler was run for 3×10^5 cycles after discarding the first 2000 cycles for the burn-in period. The chain was thinned (one iteration in every 15 cycles was saved) to reduce serial correlation in the stored samples so that the total number of samples kept in the post-Bayesian analysis was 2×10^4 (GELMAN *et al.* 1995). The stored sample was used to infer the parameters of interest. We now describe our results for each experimental design.

Zero-QTL model: In design I, we tested whether any spurious loci were detected when there is no true QTL. The priors for all main and epistatic effects were chosen to be $N(0, 0.5)$. The prior distribution of l was taken as uniform(0, 6) and Poisson distribution with three different prior means (see Table 2). For all analyses, we checked the plots of the changes in the number of QTL against the number of the iterations (not shown here). These plots showed that the MCMC sampler mixes well, with excursions into high values being short lived. The estimated posterior probability distributions for the number of QTL, averaged over five replicates, are given in Table 2. The posterior probability $p(l = x | \mathbf{y}, \mathbf{M})$ ($x = 0, 1, 2, \dots$) was obtained by counting the number of samples in which the number of QTL is l , divided by the total of number of samples. For all prior specifications, the posterior probabilities of $l = 0$ have the highest value. Table 2 also gives Bayes factors $B(0, 1)$ ($= p(l = 0 | \mathbf{y}, \mathbf{M}) / p(l = 1 | \mathbf{y}, \mathbf{M}) \times p(l = 1) / p(l = 0)$), comparing

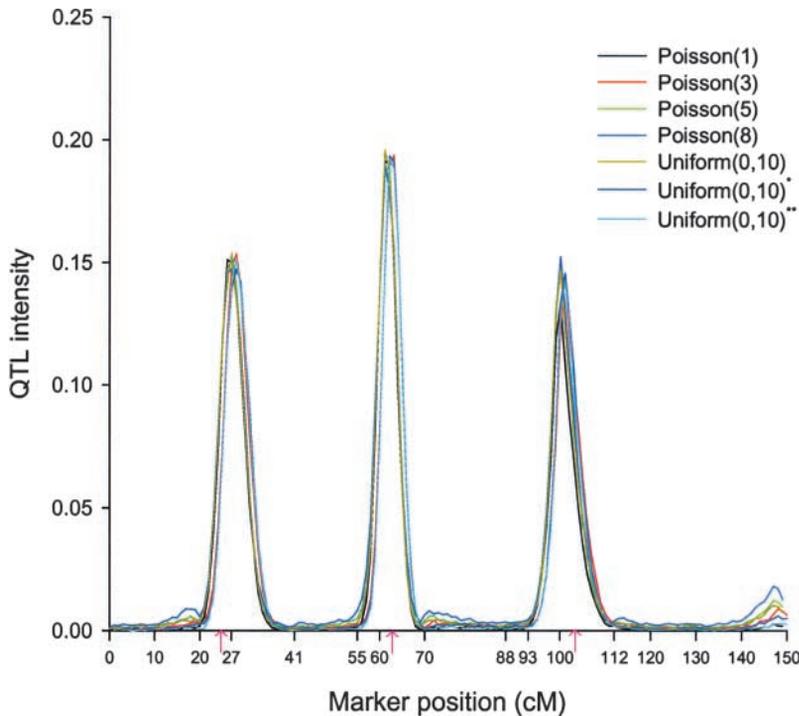


FIGURE 1.—Profiles of posterior QTL intensity for design II with respect to the different priors for the number and genetic effects of QTL. The true positions of the three simulated QTL are indicated by the arrows on the horizontal axis. *The prior variances for all QTL effects are 1.5. **The prior variances for all QTL effects are 2.5. For other cases, the prior variances for all QTL effects are 1.0.

a model with no QTL against a model with one QTL (SATAGOPAN and YANDELL 1996; GAFFNEY 2001). Evidently, the results supported a model with no QTL.

Three-QTL model: In the analyses of design II, we were primarily interested in checking for sensitivity to the choice of prior specifications for the number and genetic effects of QTL. The prior distribution for the number of QTL was taken as uniform(0, 10) and Poisson distribution with four different prior means (see Table 3). The priors for all main and epistatic effects were chosen to be $N(0, \tau^2)$ and three different values for τ^2 were considered (see Table 3). The prior distributions of the number and genetic effects of QTL appear in the acceptance ratios (see APPENDICES A–C) and thus may influence the posterior distributions of the parameters of interest and the mixing behavior of the chain.

As observed in the analyses of design I, the plots of the changes in the number of QTL against the number of the iterations showed that the MCMC sampler mixes well, changing frequently but always centralized around

the posterior mode of the QTL number (not shown here). The estimated posterior probability distributions for the number of QTL are given in Table 3. In all analyses, the posterior mode of the number of QTL coincided with the simulated value and the posterior mean approximately equaled the true value. The posterior mode of the number of QTL was not affected by the choice of prior mean for the number of QTL and prior variance for genetic effects. However, the posterior probabilities were affected by these prior specifications. When taking Poisson distribution as the prior for the number of QTL, increasing the prior mean favored a higher number of QTL. Similarly, reducing the prior variance for genetic effects also favored a higher number of QTL. Table 3 also gives Bayes factors $B(3, 2)$ and $B(3, 4)$ for all analyses. Apparently, Bayes factors did not seem to be affected by the prior of l , but were affected by the choice of prior variance of genetic effects. However, all analyses supported a model with three QTL.

QTL locations were estimated using the posterior

TABLE 1

Simulation of seven-QTL model: true locations, main effects, and epistatic effects of seven simulated QTL

Chromosome:	1	1	2	2	3	3	4
QTL:	1	2	3	4	5	6	7
True location (cM):	25	45	25	55	25	55	15
Main effect:	$a_1 = -0.70$ (0.054)	$a_2 = 0.90$ (0.089)	$a_3 = 0.00$	$a_4 = 0.00$	$a_5 = 0.00$	$a_6 = -0.90$ (0.089)	$a_7 = -0.90$ (0.089)
Nonzero epistasis:	$b_{12} = 1.70$ (0.079)	$b_{12} = 1.70$ (0.079)	$b_{34} = 1.70$ (0.079)	$b_{34} = 1.70$ (0.079)	$b_{56} = -1.7$ (0.079)	$b_{56} = -1.7$ (0.079)	

Heritability explained by each nonzero effect is given in parentheses.

TABLE 2
Analysis of zero-QTL model: estimates of the posterior distribution
of the number of QTL, its expectation, and Bayes factor

Prior	Estimated distribution for l					Bayes factor: $B(0, 1)$
	0	1	2	3	>4	
Poisson(1)	0.9125	0.0731	0.0141	0.0003	0.0000	12.4829
Poisson(3)	0.6741	0.2658	0.0469	0.0124	0.0007	7.6083
Poisson(5)	0.5723	0.2906	0.0918	0.0407	0.0040	9.8469
Uniform(0, 6)	0.8571	0.1121	0.0276	0.0028	0.0004	7.6459

QTL intensity function (SILLANPÄÄ and ARJAS 1998). For convenience, we divided each chromosome into many small intervals of equal length (bins), say 1 cM, and then calculated the frequency of hits by the QTL in each interval from all posterior samples (SILLANPÄÄ and ARJAS 1998). A more accurate method to calculate the posterior QTL intensity can be found in HOTI *et al.* (2002). For the different priors of the number and genetic effects of QTL, the QTL intensity profiles are depicted in Figure 1. These profiles were fairly similar and thus these priors did not seem to affect the posterior inference of QTL positions. Apparently, these profiles had three obvious peaks, each close to a true QTL. Therefore, three QTL were localized with high precision. The posterior inference about the effect indicators and the genetic effects were not sensitive to the priors of the number and genetic effects of QTL.

Seven-QTL model: In design III, we primarily evaluated the ability of the proposed method for mapping complex epistatic genes. The prior for the number of QTL was taken as uniform(0, 20). The priors for all main and epistatic effects were chosen to be $N(0, 1)$. We tried to run the MCMC sampler using other prior distributions for the number of QTL and the main and epistatic effects and found the results to be similar. When Poisson with a small prior mean was taken as prior distribution for the number of QTL, however, the chain did not mix

well and converged slowly. For uniform or Poisson distributions with suitable prior means, the chains mixed well.

The estimated posterior probability distribution for the number of QTL is given in Figure 2. It is immediately apparent from Figure 2 that the estimated posterior mode of the number of QTL coincides with the simulated value and the estimated posterior mean approximately equals the true value. The posterior probability of the true value of the QTL number was much greater than those of other values. Therefore, the number of QTL was estimated accurately using our method.

The profiles of posterior QTL intensity are shown in Figure 3 (left side). The regions of 95% highest posterior density (HPD; GELMAN *et al.* 1995) are given in Table 4. As shown in Figure 3 and Table 4, the profiles of posterior QTL intensity are concentrated around the true locations of the simulated QTL and the HPD regions include the true locations. The modes of the posterior distribution of QTL locations are given in Table 4. It can be seen that, for most of the identified QTL, the estimated positions were close to the true values. As expected, the third and fourth QTL, which have only an epistatic effect, were localized slightly inaccurately compared with other QTL.

Inference for the QTL effects and the proportion of phenotypic variance explained by each effect was obtained conditional on the estimated number of QTL

TABLE 3
Analysis of three-QTL model: estimates of the posterior distribution of the number of QTL, its expectation, and Bayes factors

Prior	Estimated distribution for l							Estimated expectation	Bayes factor	
	<2	2	3	4	5	6	>6		$B(3, 2)$	$B(3, 4)$
Poisson(1)	0.0002	0.1549	0.7847	0.0583	0.0021	0.0001	0.0000	2.9079	15.1975	3.3649
Poisson(3)	0.0001	0.0618	0.7524	0.1658	0.0189	0.0008	0.0000	3.1447	12.1747	3.4035
Poisson(5)	0.0000	0.0298	0.6608	0.2573	0.0465	0.0052	0.0004	3.3378	13.3046	3.2103
Poisson(8)	0.0000	0.0143	0.5474	0.3251	0.0949	0.0161	0.0019	3.5575	14.3549	3.3676
Uniform(0, 10)	0.0000	0.0499	0.7024	0.2096	0.0323	0.0052	0.0006	3.2423	14.0762	3.3511
Uniform(0, 10)*	0.0001	0.0620	0.7927	0.1312	0.0126	0.0014	0.0000	3.0988	12.7855	6.0419
Uniform(0, 10)**	0.0000	0.0975	0.8246	0.0731	0.0046	0.0003	0.0000	2.9855	8.4574	11.2804

*The prior variances for all QTL effects are 1.5; **the prior variances for all QTL effects are 2.5. For other cases, the prior variances for all QTL effects are 1.0.

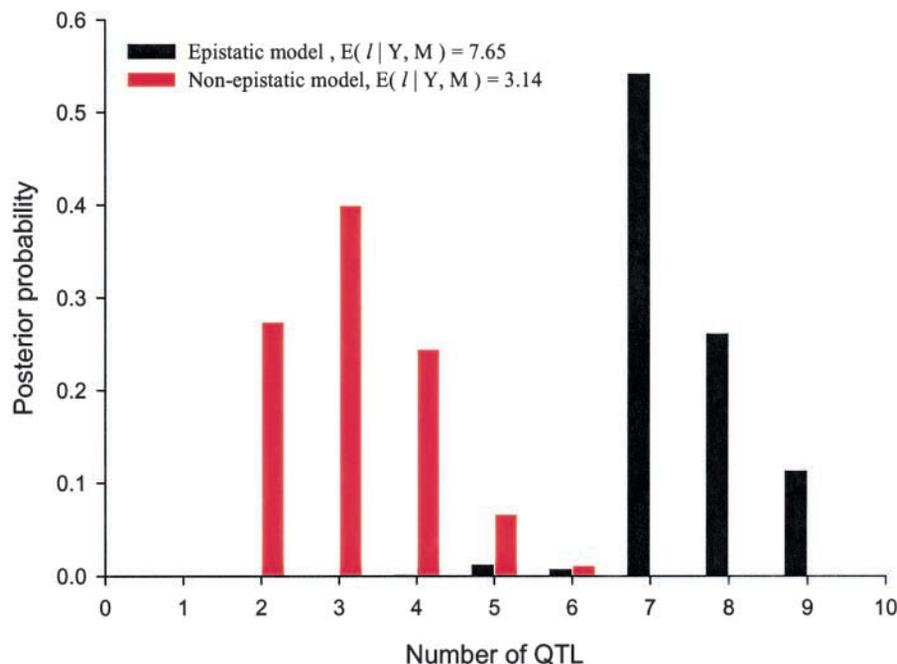


FIGURE 2.—Posterior distributions of the number of QTL for design III under the epistatic and nonepistatic models.

(estimated mode) and the estimated loci falling into the corresponding 95% HPD regions. Although the empirical marginal posterior distribution for each effect and heritability can be depicted, for simplicity we report only the posterior means and standard deviations. These estimates were calculated to average over all models conditional on $l = 7$ with the effect set to zero if the effect was not selected. Table 4 gives the estimated posterior means and posterior standard deviations for the main and epistatic effects of the identified QTL, as well as the heritability of each effect. Compared with Table 1, we can see that most of the effects and the heritabilities of the effects were estimated accurately. Table 4 also gives the posterior probability that each effect is included in the model (*i.e.*, that the corresponding indicator is non-zero). For all simulated main and epistatic effects, the corresponding posterior probabilities were estimated to be $>90\%$. However, the posterior probabilities for all nonexistent effects were small.

The estimates of the overall mean and the residual variance were -0.911 and 1.052 , respectively. The posterior standard deviations for these estimates were 0.137 and 0.102 , respectively. Therefore, the overall mean and the residual variance were estimated with high precision.

We reanalyzed the simulated data using the nonepistatic model, where only main effects of QTL are fit in the model. The priors for the number, locations, and main effects of QTL and the tuning parameter for updating the locations of QTL were set to be the same as those in the above analysis. The estimated posterior probabilities for the number of QTL are given in Figure 2. Obviously, the estimated posterior mode was much smaller than the true number of QTL. The posterior mean was estimated to be 3.14 , which was also smaller

than the true number of QTL. Therefore, it is immediately apparent that some QTL remained undetected when ignoring epistasis. The profiles of the posterior QTL intensity are depicted in Figure 3 (right side). Clearly, all QTL with only epistatic effects were not detected in this analysis. Surprisingly, the two QTL simulated on chromosome one were actually not detected, although these two QTL have main effects. It was found that the main effects were estimated with bias. Finally, the estimates of the overall mean and the residual variance were -0.653 and 1.672 , respectively. As expected, the residual variance was estimated to be greater than the true value because the epistatic variation was absorbed into the residual variance.

Real data analysis: Data from the North American Barley Genome Mapping Project (TINKER *et al.* 1996) were reanalyzed using the proposed Bayesian method. Seven traits were investigated in the project: heading, yield, maturity, height, lodging, kernel weight, and test weight. We represent only the result of heading here. The DH (double haploid) population contained 145 lines ($n = 145$) and each was grown in a range of environments. A total of 127 mapped markers ($K = 127$) covering 1500 cM of the genome along seven linkage groups were used in the analysis. The average phenotypic values across the environments were calculated for each line and these average values were treated as the original phenotypic values (y_i) for the analysis. These phenotypic values were further standardized using $y_i = (y_i - \bar{y})/s$, where \bar{y} is the mean and s is the standard deviation of y . The standardized records were subject to the Bayesian analysis.

The priors for the number, locations, and main and epistatic effects of QTL and the tuning parameter for updating the locations of QTL were set to be the same

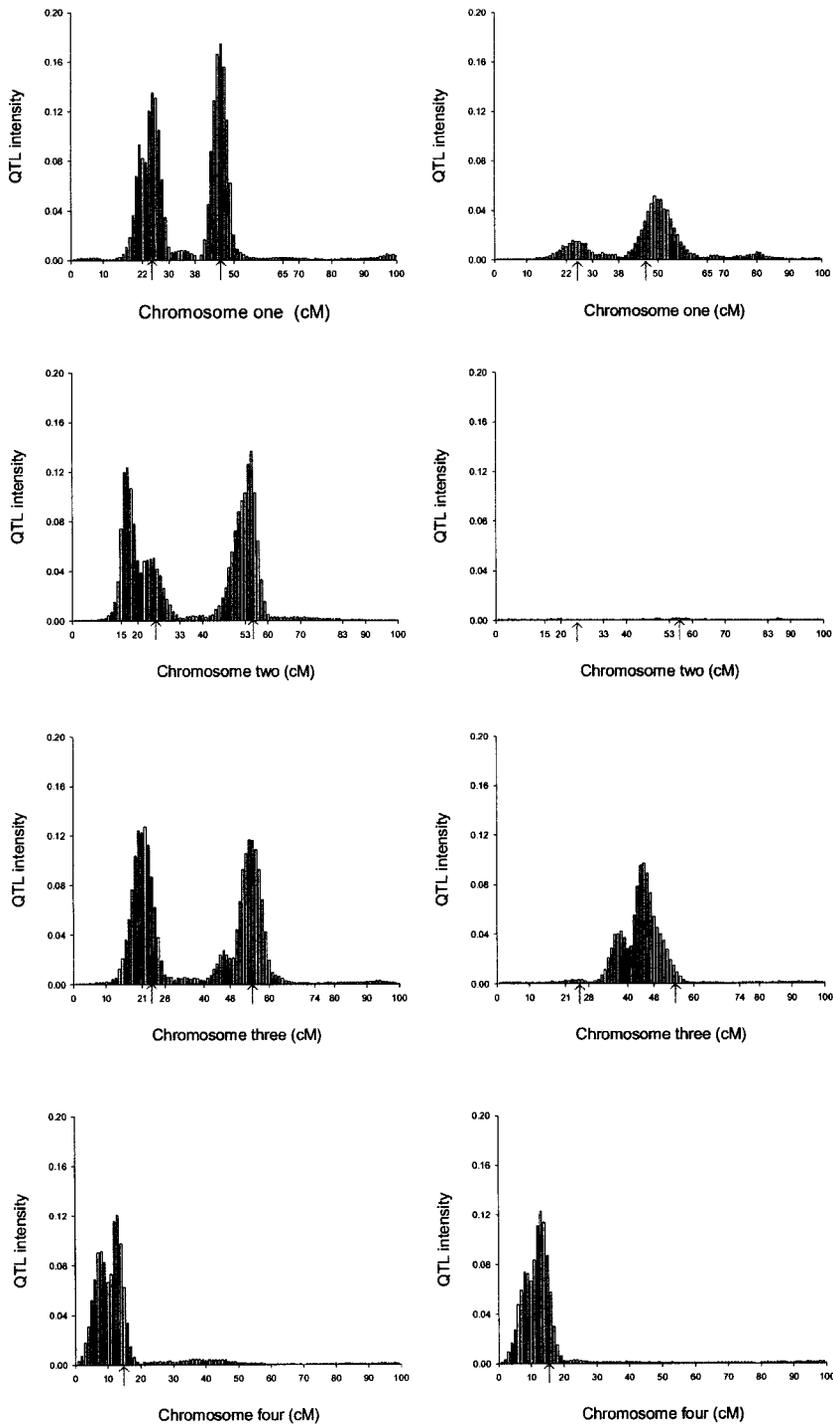


FIGURE 3.—Profiles of posterior QTL intensity for design III under the epistatic (left side) and nonepistatic models (right side). The true positions of the seven simulated QTL are indicated by the arrows on the horizontal axes.

as those in the analysis of design III described above. The length of the chain was also set to be the same as that in the simulated studies. The posterior probability distribution of the number of QTL is depicted in Figure 4, showing that the variation of the trait is most likely contributed by seven or eight loci with an equal chance ($\sim 30\%$), and a slight chance ($\sim 18\%$) contributed by nine loci. The estimated posterior expectation of the number of QTL was 7.98. The posterior probability that there are at least seven loci was $\sim 92\%$. Therefore, a conservative conclusion is that at least seven loci are

contributing to the genetic variation. In contrast, using the interval mapping method (LANDER and BOTSTEIN 1989) and the composite interval mapping method (ZENG 1994), TINKER *et al.* (1996) detected only three QTL.

The posterior QTL intensity is shown in Figure 5. The 95% HPD regions and the posterior modes of the locations of the seven QTL are given in Table 5. Seven QTL were localized at six chromosomes with width of 95% HPD from 20 to 38 cM. The three regions, on chromosomes one, four, and seven, respectively, were declared by TINKER *et al.* (1996) as significant. But TIN-

TABLE 4
Analysis of seven-QTL model: 95% high posterior density regions of QTL locations, posterior modes of QTL locations, main and epistatic effects, and heritability explained by each effect

Chromosome:	1	2	3	4	5	6	7
QTL:	1	2	3	4	5	6	7
95% HPD region:	~16-29	~40-49	~12-30	~44-58	~13-26	~42-59	~2-16
Mode of location:	24	45	16	54	21	54	12
Main effect:	$a_1 = -0.714$ (0.172, 0.952)	$a_2 = 0.965$ (0.171, 0.967)	$a_3 = 0.013$ (0.070, 0.091)	$a_4 = 0.046$ (0.099, 0.101)	$a_5 = -0.107$ (0.146, 0.146)	$a_6 = -0.892$ (0.153, 0.958)	$a_7 = -0.960$ (0.126, 0.966)
Heritability of main effect:	$h_1 = 0.066$ (0.030)	$h_2 = 0.117$ (0.041)	$h_3 = 0.001$ (0.002)	$h_4 = 0.001$ (0.003)	$h_5 = 0.004$ (0.007)	$h_6 = 0.099$ (0.032)	$h_7 = 0.113$ (0.029)
Epistatic effect:	$b_{12} = 1.491$ (0.317, 0.933)	$b_{13} = -0.047$ (0.179, 0.111)	$b_{14} = -0.077$ (0.199, 0.136)	$b_{15} = 0.053$ (0.177, 0.102)	$b_{16} = 0.003$ (0.159, 0.067)	$b_{17} = 0.032$ (0.164, 0.193)	$b_{23} = -0.083$ (0.201, 0.239)
	$b_{24} = 0.083$ (0.206, 0.136)	$b_{25} = 0.028$ (0.157, 0.176)	$b_{36} = 0.011$ (0.155, 0.062)	$b_{27} = -0.079$ (0.191, 0.125)	$b_{34} = 1.58$ (0.282, 0.937)	$b_{35} = -0.015$ (0.149, 0.165)	$b_{36} = 0.002$ (0.164, 0.083)
	$b_{37} = 0.022$ (0.153, 0.175)	$b_{45} = -0.007$ (0.157, 0.066)	$b_{46} = 0.211$ (0.279, 0.400)	$b_{47} = 0.165$ (0.246, 0.348)	$b_{56} = -1.707$ (0.286, 0.930)	$b_{57} = 0.157$ (0.242, 0.339)	$b_{67} = 0.046$ (0.168, 0.101)
Epistatic heritability:	$h_{12} = 0.046$ (0.003)	$h_{13} = 0.001$ (0.002)	$h_{14} = 0.001$ (0.003)	$h_{15} = 0.001$ (0.001)	$h_{16} = 0.000$ (0.001)	$h_{17} = 0.001$ (0.002)	$h_{23} = 0.001$ (0.004)
	$h_{24} = 0.002$ (0.004)	$h_{25} = 0.001$ (0.002)	$h_{26} = 0.001$ (0.002)	$h_{27} = 0.001$ (0.003)	$h_{34} = 0.059$ (0.018)	$h_{35} = 0.001$ (0.002)	$h_{36} = 0.001$ (0.002)
	$h_{37} = 0.001$ (0.002)	$h_{45} = 0.001$ (0.002)	$h_{46} = 0.004$ (0.006)	$h_{47} = 0.005$ (0.005)	$h_{56} = 0.065$ (0.021)	$h_{57} = 0.003$ (0.005)	$h_{67} = 0.001$ (0.003)

Posterior standard deviations of the estimates (the first value) and the posterior probability that each effect is included in the model (the second value) are given in parentheses.

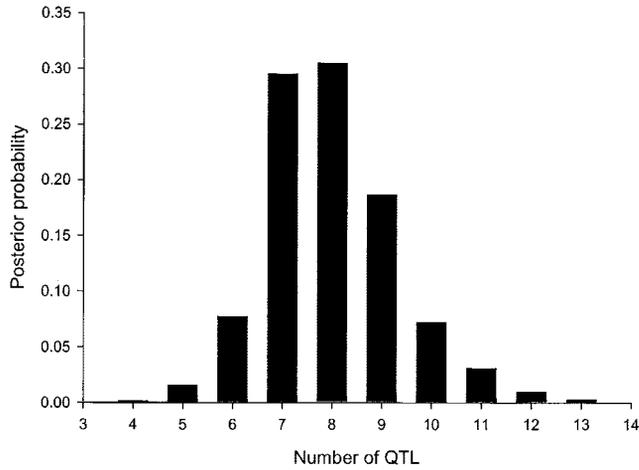


FIGURE 4.—Posterior distribution of the number of QTL for real data.

KER *et al.* (1996) failed to detect the other four regions. This illustration suggests (but does not unequivocally demonstrate) that our method is more powerful than certain existing approaches in discovering QTL.

Conditional on the estimated number of QTL ($l = 7$) and the estimated loci falling into the corresponding 95% HPD regions, we calculated the main and epistatic effects and the heritability explained by each effect (Table 5). Four QTL exhibited negative main effects, and the other three QTL showed positive main effects. The posterior standard deviations for the estimates of the main effects were small, indicating that the estimates were fairly precise. Table 5 also gives the posterior probability that each effect is included in the model. The estimated posterior probability that each main effect is included in the model was $>85\%$. It seems that there were four nonzero epistatic effects, *i.e.*, b_{14} , b_{35} , b_{45} , and b_{46} , whose heritabilities are 2.5, 1, 0.9, and 1.1%, respectively. The posterior probability including the strongest epistatic effect, b_{14} in the model, was $\sim 85\%$, and other posterior probabilities were $<55\%$. Therefore, we can conclude that there is at least one epistatic effect for heading in this DH population. The proportion of phenotypic variance explained by each main effect ranged from 5.3 to 11.3%. However, the overall proportion of phenotypic variance explained by the seven QTL was $\sim 60\%$.

DISCUSSION

Epistatic variance can be an important source of variation for complex traits. However, detecting complex epistatic genes is difficult primarily due to analytical methods, insufficient sample sizes, and study designs that exclude epistasis (FRANKEL and SCHORK 1996). In this article, we develop a Bayesian method implemented via the reversible jump MCMC algorithm to search for epistatic QTL across the entire genome and to jointly infer the number of QTL, their genomic positions, and

their significant main and epistatic effects. The proposed method was shown, via simulation, to be effective in detecting a large number of QTL with complex epistatic patterns. We investigated sensitivity of posterior inference to prior specifications of the number and genetic effects of QTL. These prior distributions do not affect the posterior mode of the number of QTL and the positions and genetic effects of QTL, but affect the posterior distribution of the number of QTL. The Bayes factor for the number of QTL is relatively insensitive to the choice of prior distribution of the number of QTL, but sensitive to the prior distributions of genetic effects. In this aspect, therefore, the proposed method retains the properties of the existing reversible jump MCMC algorithms for mapping nonepistatic QTL (SATAGOPAN and YANDELL 1996; THOMAS *et al.* 1997; GAFFNEY 2001). Recently, GAFFNEY (2001) proposed adjusting the prior variances of genetic effects as the number of QTL changes and showed that such a strategy can reduce sensitivity to the priors of genetic effects. We plan to incorporate the priors of GAFFNEY (2001) into our procedure in the future.

Our approach differs from the existing reversible jump algorithms for QTL mapping in several essential ways. First, most of the reversible jump algorithms have been developed on the basis of nonepistatic models. In nonepistatic models, only an effect parameter is associated with one QTL and thus it is relatively easy to implement the reversible jump for adding one QTL or deleting one QTL. In epistatic models, however, many epistatic effects are involved when adding or deleting a QTL. In our previous study (YI and XU 2002), we generated all possible effects corresponding to the proposed QTL when one new QTL is proposed to be added. Although such a method has been shown to work well when the number of QTL is small, *e.g.*, $l = 2$ or 3 , the acceptance rate for adding one QTL may be drastically reduced as the number of QTL increases. In the present study, the number of effects and the actual effects are first proposed when adding one QTL. The proposed effects are then generated from their fully conditional posterior distribution. We introduced an indicator vector γ to represent the number of effects and the actual effects presented in the model. The number of effects and the epistatic effects included in the model are proposed from the prior distribution of γ . With such an algorithm, we have a chance at each step to pick a “good” combination of effects for the new QTL in arbitrarily complicated situations. The acceptance rate for adding or deleting one QTL in our method is typically 3–5%. Importantly, the acceptance rates are rarely influenced by the true number of QTL and the pattern of epistasis. These properties make our algorithm generally useful for mapping QTL under complicated epistatic models. Second, our method can search QTL with only epistatic effects between themselves, which is achieved in step g. Since this type of QTL does not depend on other types

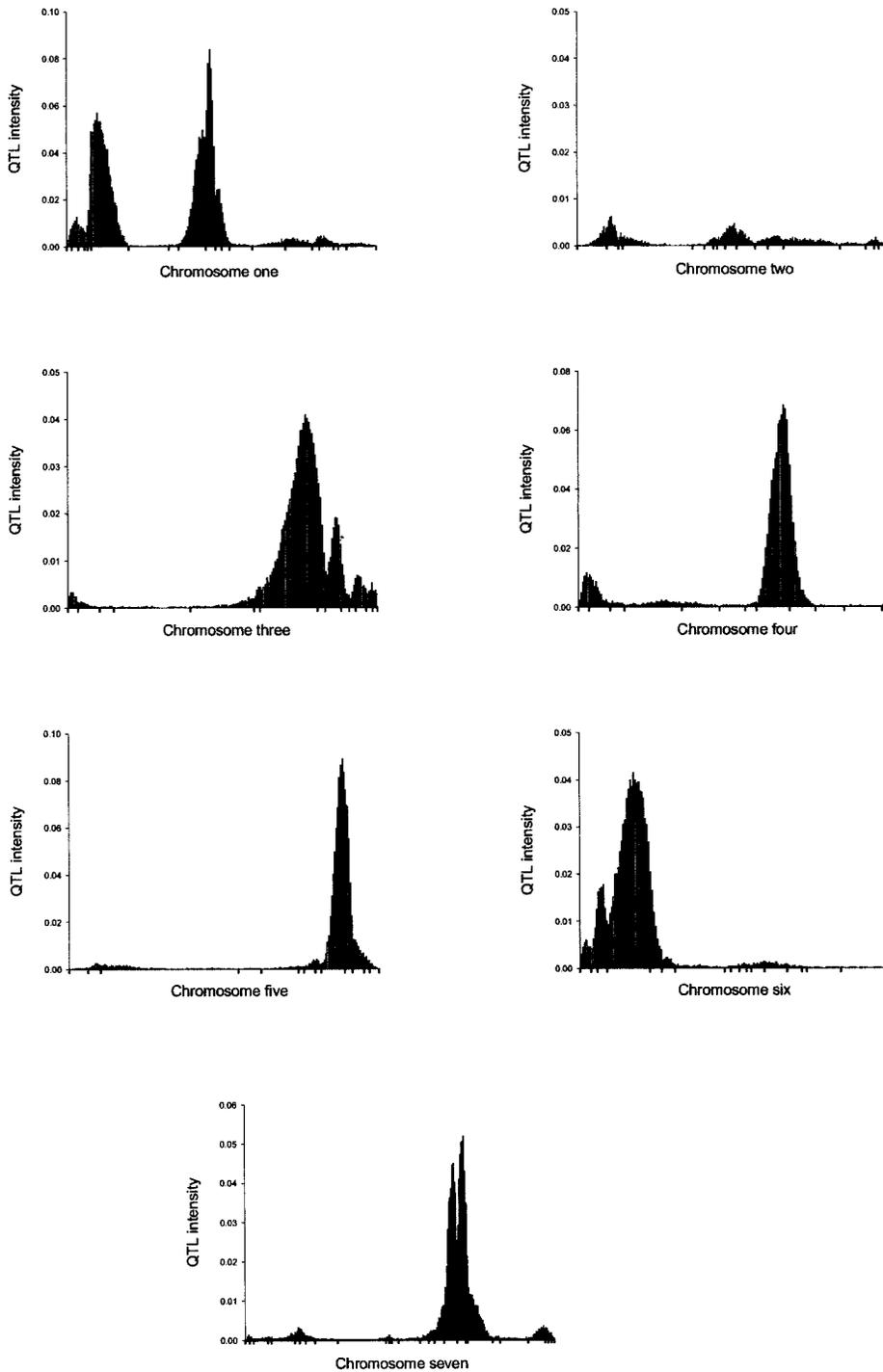


FIGURE 5.—Profiles of posterior QTL intensity for real data.

of QTL, we search them in a separate step, which simplifies the MCMC algorithm. In our simulation studies, the acceptance rate for step g is typically 0.5–1%. Third, our approach uses a new reversible jump step to determine the genetic models in terms of the number of effects. We have treated this problem as a variable selection problem. Model updating is performed by local moves where single terms are added or deleted from the model. When a variable is added or deleted from the model, all remaining parameters are unchanged. Such a

procedure has been successfully applied to model determination and variable selection in a log-linear model (DELLAPORTAS and FORSTER 1999) and a general linear model (DELLAPORTAS *et al.* 2002). When an effect is proposed to be added to the model, we generate a value for the parameter corresponding to the proposed effect variable from its fully conditional posterior distribution. The acceptance rates for these moves are typically ~ 25 – 30% . For our complex simulated data sets and the real data, these proportions are satisfactory and show that

our proposal based on the local moves and the fully conditional posterior distribution is sensible.

The aim of our approach is similar to the MIM method proposed by KAO *et al.* (1999) and reviewed by ZENG *et al.* (1999). The idea of both our approach and the MIM approach is to determine the genetic model of a complex trait and to estimate the associated genetic parameters. The essential difference between our approach and the MIM approach lies in the statistical methodologies that the two approaches use. Our method is Bayesian implemented via modern MCMC algorithms whereas the MIM uses an expectation-maximization algorithm to evaluate the likelihood and a stepwise search procedure to build the model. Advantages of the Bayesian approaches to QTL mapping over the non-Bayesian methods have been elucidated in a number of articles (*e.g.*, HOESCHELE *et al.* 1997; SILLANPÄÄ and ARJAS 1998; YI and XU 2000; HOESCHELE 2001). These advantages are more obvious in the context of mapping interacting QTL. In our Bayesian approach, we can add or delete two QTL at a time and thus uncover epistatic QTL without main effects. For maximum likelihood, however, it is difficult to perform such a search (ZENG *et al.* 1999). Our procedure can infer the posterior distributions for all parameters of interest as well as the Bayes factors for the number of QTL, which facilitates model selection and comparison (GAFFNEY 2001). In the non-Bayesian methods, many statistical tests are required to determine the genetic model, and hence a high statistical threshold must be adopted to avoid false positives among those tests. A high threshold would drastically reduce the power of detecting QTL and finding significant epistasis (JANNINK and JANSEN 2001). We aim to derive and present posterior distributions of the parameters of interest and not just approximate “best estimates.” Inferences about any particular parameter of interest can be obtained by either averaging over possible models or using a single selected model. Therefore, the Bayesian approach can provide more robust inferences than non-Bayesian methods (BALL 2001; SILLANPÄÄ and CORANDER 2002). The common aim of the non-Bayesian methods is to find a single “optimal” model regarded as the number, locations, and effects of QTL and then to make inferences as if the selected model were the true model. However, this ignores uncertainty about the model itself.

Our method appears to work well when tested using both the complex simulated data and real data set. However, we do not want to claim that the proposed algorithm is absolutely the optimal one. Improvements in several aspects can be made, *e.g.*, computational efficiency, acceptance rates for adding or deleting QTL, and mixing behavior. The proposed reversible jump algorithm is computationally intensive. For analysis of a single data set of average size, our current program takes several hours, which makes a full-scale simulation study with many replicates of each choice of parameters

TABLE 5
Real data analysis: 95% high posterior density regions of QTL locations, posterior modes of QTL locations, main and epistatic effects, and heritability explained by each effect

Chromosome:	1	2	3	4	5	6	7
QTL:	1	2	3	4	5	6	7
95% HPD region:	~6–35	~84–107	~125–163	~114–136	~167–187	~10–44	~175–199
Mode of location:	21	98	142	126	177	30	191
Main effect:	$a_1 = -0.445$ (0.129, 0.871)	$a_2 = -0.613$ (0.138, 0.901)	$a_3 = -0.612$ (0.140, 0.899)	$a_4 = 0.664$ (0.128, 0.912)	$a_5 = 0.608$ (0.132, 0.887)	$a_6 = -0.499$ (0.135, 0.886)	$a_7 = 0.513$ (0.128, 0.890)
Heritability of main effect:	$h_1 = 0.053$ (0.028)	$h_2 = 0.099$ (0.042)	$h_3 = 0.098$ (0.043)	$h_4 = 0.113$ (0.042)	$h_5 = 0.097$ (0.039)	$h_6 = 0.067$ (0.035)	$h_7 = 0.070$ (0.033)
Epistatic effect:	$h_{12} = -0.044$ (0.155, 0.129)	$h_{13} = 0.026$ (0.153, 0.113)	$h_{14} = -0.593$ (0.260, 0.845)	$h_{15} = -0.009$ (0.125, 0.103)	$h_{16} = -0.040$ (0.149, 0.116)	$h_{17} = 0.054$ (0.158, 0.101)	$h_{23} = -0.016$ (0.144, 0.063)
	$b_{24} = -0.028$ (0.149, 0.104)	$b_{25} = 0.001$ (0.133, 0.028)	$b_{26} = -0.025$ (0.138, 0.128)	$b_{27} = -0.082$ (0.189, 0.142)	$b_{34} = 0.141$ (0.228, 0.273)	$b_{35} = 0.346$ (0.263, 0.529)	$b_{36} = -0.258$ (0.256, 0.274)
	$b_{37} = 0.008$ (0.142, 0.095)	$b_{45} = 0.315$ (0.260, 0.512)	$b_{46} = -0.361$ (0.279, 0.526)	$b_{47} = -0.048$ (0.149, 0.106)	$b_{56} = -0.047$ (0.152, 0.144)	$b_{57} = 0.085$ (0.188, 0.156)	$b_{67} = -0.024$ (0.168, 0.084)
Epistatic heritability:	$h_{12} = 0.001$ (0.005)	$h_{13} = 0.002$ (0.005)	$h_{14} = 0.025$ (0.020)	$h_{15} = 0.001$ (0.003)	$h_{16} = 0.002$ (0.004)	$h_{17} = 0.002$ (0.005)	$h_{23} = 0.001$ (0.004)
	$h_{24} = 0.001$ (0.004)	$h_{25} = 0.001$ (0.003)	$h_{26} = 0.001$ (0.002)	$h_{27} = 0.003$ (0.006)	$h_{34} = 0.004$ (0.009)	$h_{35} = 0.010$ (0.009)	$h_{36} = 0.006$ (0.011)
	$h_{37} = 0.001$ (0.001)	$h_{45} = 0.009$ (0.008)	$h_{46} = 0.011$ (0.010)	$h_{47} = 0.002$ (0.004)	$h_{56} = 0.002$ (0.005)	$h_{57} = 0.003$ (0.006)	$h_{67} = 0.001$ (0.004)

Posterior standard deviations of the estimates (the first value) and the posterior probability that each effect is included in the model (the second value) are given in parentheses.

a difficult task. However, the time required could be greatly reduced by optimizing the program codes and using a more efficient sampler (GAFFNEY 2001). One possible way to improve the acceptance rate for adding QTL is to choose the location(s) of the new QTL to target the more "important" regions of the genome, rather than randomly selecting location(s) from the unoccupied regions. In the context of human nuclear families, LEE and THOMAS (2000) developed a method to propose a location by scanning the unoccupied regions of the entire genome for evidence of linkage of the trait residuals, although they did not discuss epistasis. They showed that their method has greatly improved the acceptance rate. On the other hand, we can apply the multiple-try MCMC to first select randomly several locations from the unoccupied regions and then choose the most important one among these locations (LIU *et al.* 2000). The multiple-try MCMC may compromise Lee-Thomas's genome-scanning method and the randomly choosing method as used here in terms of statistical efficiency and computational load. Finally, it is also important to extend our procedure to handle experimental designs with three segregating genotypes (*e.g.*, F_2) and include higher-order epistatic effects. For such a mapping population, we need to include dominance, additive-by-dominance, dominance-by-additive, and dominance-by-dominance effects into the model. In principle, our algorithm could be easily extended to accommodate these additional parameters. Under these extensions, however, the number of parameters is largely increased. Further work will be devoted to improving computational efficiency and investigating the mixing behavior of our reversible jump algorithm.

We are very grateful to Dr. Alfred Bartolucci, the associate editor, and two anonymous reviewers for their helpful comments and criticisms. This work was supported in part by the National Institutes of Health (NIH) no. GM55321; the U.S. Department of Agriculture National Research Initiative Competitive Grants Program 00-35300-9245 to S.X.; and NIH grants R01ES009912, P41RR006009, R01DK054298, and P30DK56336 to D.B.A.

LITERATURE CITED

- BALL, R. D., 2001 Bayesian methods for quantitative trait loci mapping based on model selection: approximate analysis using the Bayesian information criterion. *Genetics* **159**: 1351–1364.
- BINK, M. C. A. M., P. UIMARI, M. J. SILLANPÄÄ, L. L. G. JANSSEN and R. C. JANSEN, 2002 Multiple QTL mapping in related plant populations via a pedigree-analysis approach. *Theor. Appl. Genet.* **104**: 751–762.
- BOER, M. P., C. J. F. TER BRAAK and R. C. JANSEN, 2002 A penalized likelihood method for mapping epistatic quantitative trait loci with one-dimensional genome searches. *Genetics* **162**: 951–960.
- BROOKS, S. P., and P. GIUDICI, 2000 Markov chain Monte Carlo convergence assessment via two-way analysis of variance. *J. Comput. Graph. Stat.* **9**: 266–276.
- CARLBORG, O., L. ANDERSSON and B. KINGHORN, 2000 The use of a genetic algorithm for simultaneous mapping of multiple interacting quantitative trait loci. *Genetics* **155**: 2003–2010.
- CHEVERUND, J. M., and E. J. ROUTMAN, 1995 Epistasis and its contribution to genetic variance components. *Genetics* **139**: 1455–1461.
- CORANDER, J., and M. J. SILLANPÄÄ, 2002 A unified approach to joint modeling of multiple quantitative and qualitative traits in gene mapping. *J. Theor. Biol.* **218**: 435–446.
- CULVERHOUSE, R., B. K. SUAREZ, J. LIN and T. REICH, 2002 A perspective on epistasis: limits of models displaying no main effect. *Am. J. Hum. Genet.* **70** (2): 461–471.
- DAW, E. W., S. C. HEATH and E. M. WIJSMAN, 1999 Multiple oligogenic analysis of age-at-onset data with applications to Alzheimer disease pedigrees. *Am. J. Hum. Genet.* **64**: 839–851.
- DELLAPORTAS, P., and J. J. FORSTER, 1999 Markov chain Monte Carlo model determination for hierarchical and graphical log-linear models. *Biometrika* **86**: 615–633.
- DELLAPORTAS, P., J. J. FORSTER and I. NTZOUFRAS, 2002 On Bayesian model and variable selection using MCMC. *Stat. Comput.* **12**: 27–36.
- DONG, C., S. WANG, W.-D. LI, D. LI, H. ZHAO *et al.*, 2003 Interacting genetic loci on chromosomes 20 and 10 influence extreme human obesity. *Am. J. Hum. Genet.* **72**: 115–124.
- FERNANDEZ, J. R., L. M. TARANTINO, S. M. HOFER, G. P. VOGLER and G. E. MCCLEARN, 2000 Epistatic quantitative trait loci for alcohol preference in mice. *Behav. Genet.* **30** (6): 431–437.
- FIJNEMAN, R. J. A., S. S. DE VRIES, R. C. JANSEN and P. DEMANT, 1996 Complex interactions of new quantitative trait loci, *Sluc1*, *Sluc2*, *Sluc3*, and *Sluc4* that influence the susceptibility to lung cancer in mouse. *Nat. Genet.* **14**: 465–467.
- FRANKEL, W. N., and N. J. SCHORK, 1996 Who's afraid of epistasis. *Nat. Genet.* **14**: 371–373.
- GAFFNEY, P. J., 2001 An efficient reversible jump Markov chain Monte Carlo approach to detect multiple loci and their effects in inbred crosses. Ph.D. Dissertation, Department of Statistics, University of Wisconsin, Madison, WI.
- GAUDERMAN, W. J., and D. C. THOMAS, 2000 The role of interacting determinants in the localization of genes. *Adv. Genet.* **42**: 393–412.
- GELMAN, A., J. CARLIN, H. STERN and D. RUBIN, 1995 Bayesian data analysis. Chapman & Hall, London.
- GREEN, P. J., 1995 Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82**: 711–732.
- 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.
- HEATH, S. C., 1997 Markov chain Monte Carlo segregation and linkage analysis for oligogenic models. *Am. J. Hum. Genet.* **61**: 748–760.
- HOESCHELE, I., 2001 Mapping quantitative trait loci in outbred pedigrees, pp. 599–644 in *Handbook of Statistical Genetics*, edited by D. J. BALDING, M. BISHOP and C. CANNINGS. John Wiley & Sons, New York.
- 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.
- HOTI, F. J., M. J. SILLANPÄÄ and L. HOLMSTROM, 2002 A note on estimating the posterior density of a quantitative trait locus from a Markov chain Monte Carlo sample. *Genet. Epidemiol.* **22**: 369–376.
- HURME, P., M. J. SILLANPÄÄ, E. ARJAS, T. REPO and O. SAVOLAINEN, 2000 Genetic basis of climatic adaptation in Scots pine by Bayesian quantitative trait locus analysis. *Genetics* **156**: 1309–1322.
- JANNINK, J., and R. JANSEN, 2001 Mapping epistatic quantitative trait loci with one-dimensional genome searches. *Genetics* **157**: 445–454.
- JANSEN, R. C., and P. STAM, 1994 High resolution of quantitative traits into multiple loci via interval mapping. *Genetics* **136**: 1447–1455.
- JIANG, C., and Z.-B. ZENG, 1997 Mapping quantitative trait loci with dominant and missing markers in various crosses from two inbred lines. *Genetica* **101**: 47–58.
- KAO, C. H., and Z.-B. ZENG, 2002 Modeling epistasis of quantitative trait loci using Cockerham's model. *Genetics* **160**: 1243–1261.
- KAO, C. H., Z.-B. ZENG and R. D. TEASDALE, 1999 Multiple interval mapping for quantitative trait loci. *Genetics* **152**: 1203–1216.
- LANDER, E. S., and D. BOTSTEIN, 1989 Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**: 185–199.
- LARK, K. G., K. CHASE, F. R. ADLER, L. M. MANSUR and J. J. ORF, 1995 Interactions between quantitative trait loci in soybean in

- which trait variation at one locus is conditional upon a specific allele at another. *Proc. Natl. Acad. Sci. USA* **92**: 4656–4660.
- LEE, J. K., and D. C. THOMAS, 2000 Performance of Markov chain Monte Carlo approaches for mapping genes in oligogenic models with an unknown number of loci. *Am. J. Hum. Genet.* **67**: 1232–1250.
- LIU, J. S., F. LIANG and W. H. WONG, 2000 The multiple-try method and local optimization in Metropolis sampling. *J. Am. Stat. Assoc.* **95**: 121–134.
- LYNCH, M., and B. WALSH, 1998 *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, Sunderland, MA.
- NTZOUFRAS, I., 1999 Aspects of Bayesian model and variable selection using MCMC. Ph.D. Dissertation, Department of Statistics, Athens University of Economics and Business, Athens, Greece.
- POPE, P. T., and J. T. WEBSTER, 1972 The use of an F-statistic in stepwise regression procedures. *Technometrics* **14**: 327–340.
- ROUTMAN, E. J., and J. M. CHEVERUD, 1997 Gene effects on a quantitative trait: two-locus epistatic effects measured at microsatellite markers and at estimated QTL. *Evolution* **5**: 1654–1662.
- SATAGOPAN, J. M., and B. S. YANDELL, 1996 Estimating the number of quantitative trait loci via Bayesian model determination. Joint Statistical Meeting, October 1996, Chicago.
- SATAGOPAN, J. M., B. S. YANDELL, M. A. NEWTON and T. C. OSBORN, 1996 A Bayesian approach to detect quantitative trait loci using Markov chain Monte Carlo. *Genetics* **144**: 805–816.
- SEN, S., and G. CHURCHILL, 2001 A statistical framework for quantitative trait mapping. *Genetics* **159**: 371–387.
- 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.
- SILLANPÄÄ, M. J., and J. CORANDER, 2002 Model choice in gene mapping: what and why. *Trends Genet.* **18**: 301–307.
- 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.
- STEPHENS, D., and A. SMITH, 1993 Bayesian inference in multipoint gene mapping. *Ann. Hum. Genet.* **57**: 65–82.
- SUGIYAMA, F., G. A. CHURCHILL, D. C. HIGGINS, C. JOHNS, K. P. MAKARITSIS *et al.*, 2001 Concordance of murine quantitative trait loci for salt-induced hypertension with rat and human loci. *Genomics* **71**: 70–77.
- THOMAS, D. C., S. RICHARDSON, J. GAUDERMAN and J. PITKANEMI, 1997 A Bayesian approach to multipoint mapping in nuclear families. *Genet. Epidemiol.* **14**: 903–908.
- TINKER, N. A., D. E. MATHER, B. G. ROSSNAGEL, K. J. KASHA and A. KLEINHOF, 1996 Regions of the genome that affect agronomic performance in two-row barley. *Crop Sci.* **36**: 1053–1062.
- UIMARI, P., and I. HOESCHELE, 1997 Mapping linked quantitative trait loci using Bayesian method analysis and Markov chain Monte Carlo algorithms. *Genetics* **146**: 735–743.
- UIMARI, P., and M. J. SILLANPÄÄ, 2001 Bayesian oligogenic analysis of quantitative and qualitative traits in general pedigrees. *Genet. Epidemiol.* **21**: 224–242.
- UIMARI, P., G. THALLER and I. HOESCHELE, 1996 The use of multiple markers in a Bayesian method for mapping quantitative trait loci. *Genetics* **143**: 1831–1842.
- WADE, M. J., 2001 Epistasis, complex traits, and mapping genes. *Genetica* **112/113**: 59–69.
- WANG, D. L., J. ZHU, Z. K. LI and A. H. PATERSON, 1999 Mapping QTLs with epistatic effects and QTL×environment interactions by mixed linear model approaches. *Theor. Appl. Genet.* **99**: 1255–1264.
- WU, R. L., 1996 Detecting epistatic genetic variance with a clonally replicated design: model for low- vs. high-order nonallelic interactions. *Theor. Appl. Genet.* **93**: 109–119.
- XU, S., and N. YI, 2000 Mixed model analysis of quantitative trait loci. *Proc. Natl. Acad. Sci. USA* **97**: 14542–14547.
- YI, N., and S. XU, 2000 Bayesian mapping of quantitative trait loci for complex binary traits. *Genetics* **155**: 1391–1403.
- YI, N., and S. XU, 2001 Bayesian mapping of quantitative trait loci under complicated mating designs. *Genetics* **157**: 1759–1771.
- YI, N., and S. XU, 2002 Mapping quantitative trait loci with epistatic effects. *Genet. Res.* **79**: 185–198.
- YU, S. B., J. X. LI, C. G. XU, Y. F. TAN, Y. J. GAO *et al.*, 1997 Importance of epistasis as the genetic basis of heterosis in an elite rice hybrid. *Proc. Natl. Acad. Sci. USA* **94**: 9226–9231.
- YUAN, B., R. NEUMAN, S. H. DUAN, J. L. WEBER, P. Y. KWOK *et al.*, 2000 Linkage of a gene for familial hypobetalipoproteinemia to chromosome 3p21.1-22. *Am. J. Hum. Genet.* **66**: 1699–1704.
- ZENG, Z.-B., 1994 Precision mapping of quantitative trait loci. *Genetics* **136**: 1457–1468.
- ZENG, Z.-B., C.-H. KAO and C. J. BASTEN, 1999 Estimating the genetic architecture of quantitative traits. *Genet. Res.* **74**: 279–289.

Communicating editor: G. CHURCHILL

APPENDIX A: UPDATING THE INDICATOR VECTOR OF QTL EFFECTS γ AND CORRESPONDING EFFECTS

Assume that the current number of QTL is l and that there are a total of $l(l+1)/2$ possible QTL effects for model (1), including l main effects and $l(l-1)/2$ digenic epistasis. We randomly select one of the $l(l+1)/2$ possible effects. If the epistasis between QTL q'_1 and q'_2 is picked, then we propose to change γ to γ' , where $\gamma'_{q'_1q'_2} = 1 - \gamma_{q'_1q'_2}$ with all other components remaining the same.

If $\gamma_{q'_1q'_2} = 0$, *i.e.*, the epistasis $b_{q'_1q'_2}$ is currently out of the model, we propose to add this epistasis into the model; this needs to generate an additional effect parameter from a proposal density $q(b_{q'_1q'_2})$. We take the following fully conditional posterior distribution as the proposal density

$$N\left(\frac{\eta/\tau^2 + \sum_{i=1}^n x_{iq'_1} x_{iq'_2} (y_i - g_i)/\sigma_c^2}{1/\tau^2 + \sum_{j=1}^n (x_{iq'_1} x_{iq'_2})^2/\sigma_c^2}, \frac{1}{1/\tau^2 + \sum_{j=1}^n (x_{iq'_1} x_{iq'_2})^2/\sigma_c^2}\right),$$

where $g_i = \mu + \sum_{q=1}^l \gamma_q x_{iq} \alpha_q + \sum_{q_1 < q_2}^l \gamma_{q_1 q_2} x_{iq_1} x_{iq_2} b_{q_1 q_2}$, and η and τ^2 are prior mean and variance of QTL effects, respectively. The proposal move is accepted with probability $\min(1, r)$, where

$$r = \frac{p(\mathbf{y}|l, \gamma', \theta', \mathbf{x})}{p(\mathbf{y}|l, \gamma, \theta, \mathbf{x})} \cdot \frac{p(\gamma'|l) p(b_{q'_1q'_2})}{p(\gamma|l) q(b_{q'_1q'_2})},$$

and $\theta' = (\theta, b_{q'_1q'_2})$. If the proposed move is accepted, update γ by γ' and add the effect $b_{q'_1q'_2}$ into the model; otherwise leave the state unchanged.

If $\gamma_{q'_1q'_2} = 1$, *i.e.*, the epistasis $b_{q'_1q'_2}$ is in the current model, we propose deleting this epistasis from the model. The proposal move is accepted with probability $\min(1, r)$, where

$$r = \frac{p(\mathbf{y}|l, \gamma', \theta', \mathbf{x})}{p(\mathbf{y}|l, \gamma, \theta, \mathbf{x})} \cdot \frac{p(\gamma'|l)q(b_{q_1'q_2'})}{p(\gamma|l)p(b_{q_1q_2})},$$

where θ' is θ with $b_{q_1'q_2'}$ removed, and $q(b_{q_1'q_2'})$ takes the form

$$N\left(\frac{\eta/\tau^2 + \sum_{i=1}^n x_{iq_1'}x_{iq_2'}(y_i - g_i + x_{iq_1'}x_{iq_2'}b_{q_1'q_2'})/\sigma_c^2}{1/\tau^2 + \sum_{i=1}^n (x_{iq_1'}x_{iq_2'})^2/\sigma_c^2}, \frac{1}{1/\tau^2 + \sum_{i=1}^n (x_{iq_1'}x_{iq_2'})^2/\sigma_c^2}\right),$$

where $g_i = \mu + \sum_{q=1}^l \gamma_q x_{iq} a_q + \sum_{q_1 < q_2} \gamma_{q_1 q_2} x_{iq_1} x_{iq_2} b_{q_1 q_2}$. If the proposed move is accepted, update γ by γ' and delete the effect $b_{q_1'q_2'}$ from the model; otherwise the state remains unchanged.

APPENDIX B: ADDING OR DELETING ONE QTL

We first describe the sampling scheme for generating unknowns for the added QTL, *i.e.*, new location, new genotype indicators for all individuals, effect indicators, and selected new effects, and then derive the acceptance probabilities for adding and deleting one QTL.

The unknowns for new QTL are generated as follows:

1. Sample a location λ^* uniformly from all unoccupied intervals on the whole genome.
2. Sample the genotype indicator $\mathbf{x}^* = \{x_i^*\}_{i=1}^n$ for all individuals from the conditional distribution

$$p(\mathbf{x}^*) = \prod_{i=1}^n p(x_i^*|\lambda^*, \mathbf{M}).$$

3. A total of $l + 1$ possible new effects are associated with this proposed QTL, including one main effect and l epistatic effects between the proposed QTL and all existing QTL. In most situations, however, not all these effects are significant. Therefore, we first generate a random integral k between 1 and $l + 1$ and then pick k effects of $l + 1$ possible new effects at random. The proposed new effects, denoted by $\beta^* = (\beta_1^*, \dots, \beta_k^*)$, can be sampled sequentially from the conditional distributions,

$$q(\beta_1^*|\mathbf{y}, l, \gamma, \theta, \mathbf{x}, \mathbf{x}^*), \dots, q(\beta_k^*|\mathbf{y}, l, \gamma, \theta, \mathbf{x}, \mathbf{x}^*, \beta_1^*, \dots, \beta_{k-1}^*).$$

These conditional distributions are univariate normal and thus can be easily sampled from.

The change in the number of QTL from l to $l + 1$, together with the proposed parameters, is accepted with probability $\min(1, r)$, where the acceptance ratio is

$$r = \frac{p(\mathbf{y}|l + 1, \gamma', \theta', \mathbf{x}')}{p(\mathbf{y}|l, \gamma, \theta, \mathbf{x})} \cdot \frac{p(l + 1)p(\beta^*)j(l, l + 1)(1/l + 1)}{p(l)j(l + 1; l)q(\beta^*)},$$

where γ' is the indicator vector of QTL effects after adding the proposed QTL, $\theta' = (\theta, \beta^*)$, $\mathbf{x}' = (\mathbf{x}, \mathbf{x}^*)$, and

$$q(\beta^*) = q(\beta_1^*|\mathbf{y}, l, \gamma, \theta, \mathbf{x}, \mathbf{x}^*) \prod_{j=2}^k q(\beta_j^*|\mathbf{y}, l, \gamma, \theta, \mathbf{x}, \mathbf{x}^*, \beta_1^*, \dots, \beta_{j-1}^*).$$

Deleting a QTL is simply the reverse process. A QTL is randomly chosen among the existing QTL. The chosen QTL, together with all corresponding parameters, is then proposed to delete from the model with probability $\min(1, r)$, where

$$r = \frac{p(\mathbf{y}|l - 1, \gamma', \theta', \mathbf{x}')}{p(\mathbf{y}|l, \gamma, \theta, \mathbf{x})} \cdot \frac{p(l - 1)j(l, l - 1)q(\beta^*)}{p(l)p(\beta^*)j(l - 1; l)(1/l)},$$

where β^* is the main and epistatic effects of the deleted QTL, and θ' , γ' , and \mathbf{x}' are θ , γ , and \mathbf{x} , respectively, with the items corresponding to the deleted QTL removed.

APPENDIX C: ADDING OR DELETING TWO EPISTATIC QTL

In step g, we need to propose two QTL with only an epistatic effect between themselves. The associated parameters can be generated as follows:

1. Sample two locations, λ_1^* and λ_2^* , uniformly from all unoccupied intervals on the whole genome.
2. Sample the genotype indicators $\mathbf{x}_1^* = \{x_{i1}^*\}_{i=1}^n$ and $\mathbf{x}_2^* = \{x_{i2}^*\}_{i=1}^n$, for the two proposed locations, respectively, from the conditional distributions $p(\mathbf{x}_1^*) = \prod_{i=1}^n p(x_{i1}^*|\lambda_1^*, \mathbf{M})$ and $p(\mathbf{x}_2^*) = \prod_{i=1}^n p(x_{i2}^*|\lambda_2^*, \mathbf{M})$.

3. The new epistatic effect, denoted by β^* , is sampled from the full conditional posterior distribution, $q(\beta^*)$. This full conditional distribution is normal, *i.e.*,

$$N\left(\frac{\eta/\tau^2 + \sum_{i=1}^n x_{i1}^* x_{i2}^* (y_i - g_i)/\sigma_c^2}{1/\tau^2 + \sum_{j=1}^n (x_{j1}^* x_{j2}^*) \sigma_c^2}, \frac{1}{1/\tau^2 + \sum_{j=1}^n (x_{j1}^* x_{j2}^*)^2 / \sigma_c^2}\right),$$

where $g_i = \mu + \sum_{q=1}^l x_{iq} a_q + \sum_{q_1 < q_2}^l \gamma_{q_1 q_2} x_{iq_1} x_{iq_2} b_{q_1 q_2}$.

The change in the number of QTL from l to $l + 2$, together with the proposed parameters, is accepted with probability $\min(1, r)$, where the acceptance ratio is

$$r = \frac{p(\mathbf{y}|l+2, \gamma', \theta', \mathbf{x}')}{p(\mathbf{y}|l, \gamma, \theta, \mathbf{x})} \cdot \frac{p(l+2)p(\beta^*)j(l+2)(1/(l+1)(l+2))}{p(l)j(l+2; l)q(\beta^*)},$$

where γ' is the indicator vector of QTL effects after adding the two proposed QTL, $\theta' = (\theta, \beta^*)$, and $\mathbf{x}' = (\mathbf{x}, \mathbf{x}_1^*, \mathbf{x}_2^*)$.

To delete two QTL from the model, we randomly choose two QTL among the existing QTL. The chosen QTL, together with all corresponding parameters, is then proposed to be deleted from the model with probability $\min(1, r)$, where

$$r = \frac{p(\mathbf{y}|l-2, \gamma', \theta', \mathbf{x}')}{p(\mathbf{y}|l, \gamma, \theta, \mathbf{x})} \cdot \frac{p(l-2)j(l-2)q(\beta^*)}{p(l)p(\beta^*)j(l-2; l)(1/(l-1)l)},$$

where β^* is the main and epistatic effects of the two deleted QTL, and θ' , γ' , and \mathbf{x}' are θ , γ , and \mathbf{x} , respectively, with the items corresponding to the deleted QTL removed.

