Genet. Res., Camb. (2002), **79**, *pp.* 185–198. *With 6 figures.* © 2002 Cambridge University Press
DOI: 10.1017/S0016672301005511 Printed in the United Kingdom

Mapping quantitative trait loci with epistatic effects

NENGJUN YI AND SHIZHONG XU*
Department of Botany and Plant Sciences, University of California, Riverside, CA 92521-0124, USA

(*Received 12 December 2000 and in revised form 1 July and 1 October 2001*)

Summary

Epistatic variance can be an important source of variation for complex traits. However, detecting epistatic effects is difficult primarily due to insufficient sample sizes and lack of robust statistical methods. In this paper, we develop a Bayesian method to map multiple quantitative trait loci (QTLs) with epistatic effects. The method can map QTLs in complicated mating designs derived from the cross of two inbred lines. In addition to mapping QTLs for quantitative traits, the proposed method can even map genes underlying binary traits such as disease susceptibility using the threshold model. The parameters of interest are various QTL effects, including additive, dominance and epistatic effects of QTLs, the locations of identified QTLs and even the number of QTLs. When the number of QTLs is treated as an unknown parameter, the dimension of the model becomes a variable. This requires the reversible jump Markov chain Monte Carlo algorithm. The utility of the proposed method is demonstrated through analysis of simulation data.

1. Introduction

Variation of a quantitative trait is determined by the segregation of multiple quantitative trait loci (QTLs). Classical quantitative genetics cannot separate the effects of individual QTLs; instead, only the collective or average effect of all QTLs is explored. Under some special mating designs, e.g. North Carolina Design III, epistatic variances can be estimated (Bulmer, 1980; Lynch & Walsh, 1998). However, a large number of crosses that inherit different proportions of the founder genetic material are required to estimate individual components of epistatic variance, e.g. additive-by-additive, additive-by-dominance, dominance-by-dominance. Even if there are a sufficient number of different crosses, some epistatic variance components may be still inseparable. For example, the additive-by-dominance variance for two loci is actually composed of the additive-by-dominance component (interaction between the additive effect of the first locus and the dominance effect of the second locus) and the dominance-by-additive component (interaction between the dominance effect of the first locus and the additive effect of the second locus). With the classical analysis, these two components are simply lumped together and termed additive-by-dominance variance. With the advent of modern molecular technology, various components of epistatic genetic variance can be separated and jointly estimated with saturated molecular markers. An in-depth understanding of the interlocus interactions is important in the genetic study of complex traits (e.g. Lark *et al.*, 1995; Yu *et al.*, 1997). The ultimate goal of any QTLs linkage study is not only to estimate the number, locations and marginal effects of QTLs, but also to explore the epistatic effects (e.g. Routman & Cheverud, 1997; Zeng *et al.*, 2000).

A prerequisite for detecting epistasis is to simultaneously model all QTLs. Previous methods of QTLs mapping, including the widely used interval mapping (IM) and composite interval mapping (CIM), were developed under single-QTL models (e.g. Lander & Botstein, 1989; Zeng, 1994; Jansen & Stam, 1994). The single-QTL models fit one QTL at a time, and detect only marginal QTL effects. Statistical methods for mapping multiple QTLs with epistasis were previously developed on the basis of the fixed number

* Corresponding author. Tel: +1 (909) 787 5898. Fax: +1 (909) 787 4437. e-mail: xu@genetics.ucr.edu

of QTLs and multiple-dimensional search approaches (e.g. Haley & Knott, 1992; Wang *et al.*, 1999). These methods have not been used in practice for cases where the number of QTLs is unknown and for genome-wide searches for QTLs due to high computational demand as the number of QTLs increases. Recently, Carlborg *et al.* (2000) proposed a genetic algorithm to reduce the computational demand for simultaneous mapping of multiple interactive QTLs. Jannink & Jansen (2001) described a statistical method to map multiple QTLs with epistasis with one-dimensional genome searches. Their method identifies loci of high QTLs by genetic background interaction and requires large populations derived from multiple related inbred lines. Kao *et al.* (1999) and Zeng *et al.* (2000) extended the idea of the CIM method and developed a multiple interval mapping (MIM) approach to mapping multiple QTLs and estimating epistasis in backcross designs. It has been shown that epistasis mapping can improve the chance of QTL recovery and the accuracy of parameter estimation. However, all these methods provide only point estimates for the number, locations and effects of QTLs. The critical values for significance tests and interval estimates of the parameters have to be established using a repeated sampling technique, e.g. a permutation test (Churchill & Doerge, 1994) or bootstrapping analysis (Visscher *et al.*, 1996).

Bayesian methods have been used to map multiple QTLs (Satagopan *et al.*, 1996; Uimari & Hoeschele, 1997; Satagopan & Yandell, 1998; Heath, 1997; Stephens & Fisch, 1998; Sillanpää & Arjas, 1998, 1999). QTLs linkage analysis is complicated considerably by the fact that the number of QTLs and thus the dimension of the parameter space are essentially unknown. Green (1995) introduced a reversible jump Markov chain Monte Carlo (MCMC) algorithm to sample variables from a target distribution with an unfixed dimension. Bayesian methods, implemented via the reversible jump MCMC algorithm, have been developed to map QTLs in backcross, F2 and full-sib families for normally distributed traits (Satagopan & Yandell, 1998; Stephens & Fisch, 1998; Sillanpää & Arjas, 1998, 1999) and for complex binary traits based on a threshold model (Yi & Xu, 2000*a*). For complicated pedigrees, the reversible jump MCMC methods for mapping QTLs are also available (Heath, 1997; Uimari & Hoeschele, 1997; Xu & Yi, 2000; Yi & Xu, 2001). Although multiple QTLs at the whole genome level have been taken into consideration in the above Bayesian methods, epistatic effects have been absent.

Theoretically, it may be straightforward to include epistatic effects in the analysis in a Bayesian framework. However, one major difficulty in analysing epistasis is the generation or deletion of many parameters when implementing the reversible jump step. In this paper, we propose a Bayesian method to map multiple QTLs with pairwise locus epistasis for normally distributed and binary traits. A simple and efficient sampling algorithm is derived to implement the reversible jump in the MCMC algorithm. The method is developed for arbitrary mating designs derived from two inbred lines. As in our previous works (Yi & Xu, 2000*a*, *b*), complex binary traits are modelled under the threshold model of quantitative traits.

2. Genetic model

We consider any mapping population derived from two inbred lines, P_1 and P_2. The two inbred lines are crossed to produce a hybrid generation, F1; subsequent generations are obtained by selfing, sibmating or backcrossing to the parental or F1 generations. Let the two lines differ by l loci affecting the trait under investigation. Denote the alleles carried by P_1 and P_2 by B_q and b_q, respectively, at the qth QTL ($q = 1, 2, \ldots, l$). For l putative QTLs, there are 3^l different possible QTL genotypes in the mapping population.

We consider two types of traits: normally distributed and dichotomously distributed traits. The latter are also called binary traits. For a normally distributed trait, the observed phenotypic value of individual j, y_j, can be described by the following linear model (e.g. Bulmer, 1980):

$$\begin{aligned} y_j = b_0 &+ \sum_{q=1}^{l}(x_{jq}-1)a_q + \sum_{q=1}^{l}x_{jq}(2-x_{jq})d_q \\ &+ \sum_{q<q'}^{l}(x_{jq}-1)(x_{jq'}-1)aa_{qq'} \\ &+ \sum_{q<q'}^{l}(x_{jq}-1)x_{jq'}(2-x_{jq'})ad_{qq'} \\ &+ \sum_{q<q'}^{l}x_{jq}(2-x_{jq})(x_{jq'}-1)da_{qq'} \\ &+ \sum_{q<q'}^{l}x_{jq}(2-x_{jq})x_{jq'}(2-x_{jq'})dd_{qq'} + e_j, \\ & j = 1, 2, \ldots, n, \end{aligned} \tag{1}$$

where b_0 is the overall mean; l is the number of QTLs on the genome; a_q and d_q are the additive and dominance effects, respectively, at the qth QTL; $aa_{qq'}$, $ad_{qq'}$, $da_{qq'}$ and $dd_{qq'}$ are the epistatic effects between the qth and q'th QTL, called additive-by-additive, additive-by-dominance, dominance-by-additive and dominance-by-dominance effects, respectively; x_{jq} denotes the number of B_q alleles at the qth QTL for individual j ($x_{jq} = 0$, 1, or 2); e_j is the residual error assumed to be i.i.d. $N(0, \sigma_e^2)$. The residual error

https://doi.org/10.1017/S0016672301005511 Published online by Cambridge University Press

includes the environmental error and higher-order epistasis. Note that the model includes only the two-locus interactions and the higher-order interactions have been ignored.

Given the number of QTLs, model (1) contains $2l$ possible additive and dominance effects, and $2l(l-1)$ possible epistatic effects, with a total of $2l^2$ QTL effects. Hereafter, we use b_k and w_{jk} to denote the QTL effects (additive, dominance and epistatic effects) and their corresponding coefficients for $k = 1, 2, \ldots, 2l^2$; $j = 1, 2, \ldots, n$. Therefore, model (1) can be rewritten as

$$y_j = \sum_{k=0}^{2l^2} w_{jk} b_k + e_j = \mathbf{w}_j^T \mathbf{b} + e_j, \quad j = 1, 2, \ldots, n, \qquad (2)$$

where $\mathbf{w}_j = (w_{j0}, w_{j1}, \ldots, w_{j(2l^2)})^T$; $\mathbf{b} = (b_0, b_1, \ldots, b_{2l^2})^T$; $w_{j0} = 1$ for all j, corresponding to the coefficient of the overall mean.

For complex binary traits, the observed phenotype can be defined in a binary fashion, i.e. $s_j = 1$ if individual j is affected, and $s_j = 0$ otherwise. Complex binary traits are conventionally analysed using the threshold model (Lynch & Walsh, 1998). The threshold model assumes that an underlying normally distributed variable (liability), denoted by y_j, determines the binary observation. The link between y_j and s_j is through a threshold t, i.e. $s_j = 1$ if $y_j > t$, and $s_j = 0$ otherwise. The liability can be modelled by equation (2). The threshold model is over-parameterized so that some constraints must be imposed. The constraints are usually taken as $t = 0$ and $\sigma_e^2 = 1$ (Albert & Chib, 1993).

3. Bayesian mapping

(i) *Bayesian probability model*

In Bayesian analysis we treat all quantities under consideration as random variables, be they observed data, unknown parameters or missing data. A full probability model, i.e. a joint distribution for all quantities, is set up to combine the sampling distribution of observed data and the prior distribution for the unknowns.

In QTL mapping analysis, we observe the phenotype $\mathbf{y} = \{y_i\}_{i=1}^n$ for continuous traits or $\mathbf{s} = \{s_i\}_{i=1}^n$ for binary traits and the marker data $\mathbf{M}$. For the threshold model, the values of the underlying liability are not observed, and thus treated as missing values. Our aim here is to make joint inference about the number of QTLs l, their locations $\lambda = \{\lambda_q\}_{q=1}^l$ and their effects, including additive, dominance and epistatic effects $\mathbf{b} = (b_1, \ldots, b_{2l^2})^T$. The position of the qth QTL, λ_q, is represented by the distance of the QTL from one end of the chromosome. The locations of markers on chromosomes are fixed *a priori*. For convenience of description, the allelic inheritance patterns of marker loci are assumed to be known, although they are sampled for missing and partially informative markers.

The vector $\mathbf{x}_j = (x_{j1}, x_{j2}, \ldots, x_{jl})$ specifies the unordered genotypes for l QTLs of individual j. Genotypes of QTLs are not observed, and thus x_{jq}'s are missing data. For complicated mating designs, it is not convenient to sample x_{jq}'s directly. Instead, we use the segregation indicators to derive x_{jq}'s indirectly. Note that x_{jq}'s can be decomposed into two components, i.e., $x_{jq} = x_{jq}^p + x_{jq}^m$. The components x_{jq}^p and x_{jq}^m denote the number of B_q alleles at the paternal gamete and maternal gamete of the qth QTL in individual j respectively ($x_{jq}^p = 0$ or 1, and $x_{jq}^m = 0$ or 1). The vector $(x_{j1}^p, x_{j1}^m, \ldots, x_{jl}^p, x_{jl}^m)$ then represents the ordered genotype at the l QTLs for individual j. x_{jq}^p and x_{jq}^m can be derived using a recursive approach as follows. Assume that individuals are entered into the pedigree in a chronological order so that the parents are evaluated before their progeny. Define z_{jq}^p and z_{jq}^m as the paternal and maternal segregation (meiosis) indicators, respectively, for individual j at the qth QTL. These indicator variables are defined as $z_{jq}^p = 1$ if the paternal allele of individual j inherits the paternal allele of its father and $z_{jq}^p = 0$ otherwise; similarly, $z_{jq}^m = 1$ if the maternal allele of individual j inherits the paternal allele of its mother and $z_{jq}^m = 0$ otherwise. Let individuals j_1 and j_2 be the father and mother of individual j, respectively, then x_{jq}^p and x_{jq}^m can be expressed as

$$x_{jq}^p = z_{jq}^p x_{j_1 q}^p + (1 - z_{jq}^p) x_{j_1 q}^m$$

and

$$x_{jq}^m = z_{jq}^m x_{j_2 q}^p + (1 - z_{jq}^m) x_{j_2 q}^m,$$

respectively. Therefore, the coefficients in models (1) and (2) are completely determined by the segregation indicators.

Hereafter, we use θ to denote all the unknown parameters and missing data, i.e. $\theta = (l, \lambda, \mathbf{b}, \mathbf{Z}, \sigma_e^2)$, where $\mathbf{Z} = \{z_{jq}^p, z_{jq}^m\}_{j=1,\, q=1}^{n,\, l}$ and $\sigma_e^2 = 1$ for binary traits. Combining the sampling distribution for observed data and the prior distribution for unobservable variables, the joint distribution of all variables is

$$p(\theta, \mathbf{y}) = p(\mathbf{y} \mid \theta) \cdot p(\theta) \qquad (3)$$

for normally distributed traits and

$$p(\theta, \mathbf{y}, \mathbf{s}) = p(\mathbf{s} \mid \theta, \mathbf{y}) p(\mathbf{y} \mid \theta) \cdot p(\theta) \qquad (4)$$

for binary traits. Here we have suppressed the notation for conditional on the observed marker data. Hereafter, we use the generic symbols $p(\cdot)$ and $p(\cdot \mid \cdot)$ to represent the density and conditional density, respectively, where the actual form of the distribution depends not on p but on the argument.

https://doi.org/10.1017/S0016672301005511 Published online by Cambridge University Press

The likelihood function of the observed phenotype in (3) or the conditional distribution of the liability in (4), $p(\mathbf{y}\,|\,\theta)$, can be factorized as follows:

$$p(\mathbf{y}\,|\,\theta) = \prod_{j=1}^{n} p(y_j\,|\,\theta) = \prod_{j=1}^{n} (2\pi\sigma_e^2)^{-\frac{n}{2}} \exp\left\{ -\frac{(y_j - \mathbf{w}_j^T\mathbf{b})^2}{2\sigma_e^2} \right\}. \quad (5)$$

Note that the residual variance σ_e^2 is set to one for $p(\mathbf{y}\,|\,\theta)$ in (4).

The likelihood of the observed binary data in (4) is

$$p(\mathbf{s}\,|\,\mathbf{y}, \theta) = \prod_{i=1}^{n} p(s_i\,|\,y_i) = \prod_{i=1}^{n} \{1(y_i > 0)1(s_i = 1) + 1(y_i < 0)1(s_i = 0)\}, \quad (6)$$

where $1(X \in A)$ is the indicator function, taking a value of one if $X \in A$ is true and zero otherwise.

(ii) *Prior distributions*

In Bayesian analysis, we need to specify the prior density for the parameters and the distributions of the missing values given the parameters. The prior distributions for different types of unknowns are usually assumed to be independent *a priori*. Therefore, the joint prior distribution is

$$p(\theta) = p(l) \cdot p(\lambda\,|\,l) \cdot p(\mathbf{b}\,|\,l) \cdot p(\sigma_e^2) \cdot p(\mathbf{Z}\,|\,l, \lambda). \quad (7)$$

The prior of the number of QTLs is chosen to be Poisson with a predetermined Poisson mean μ, or Uniform between 0 and a prespecified integer $l_{\max}$. The mean μ or the maximum integer $l_{\max}$ is chosen to reflect the prior belief that there are a small number of QTLs, which can be separated from the polygenic background. The QTL positions have a joint prior of $p(\lambda) = \prod_{q=1}^{l} p(\lambda_q)$, where each $p(\lambda_q)$ is Uniform across the whole genome when no information regarding the locations is available.

For purpose of conjugacy, the priors for the QTL effects are assumed to be independently normal so that $\mathbf{b} \sim \mathrm{N}(\mathbf{b}_0, \mathbf{B}_0)$, where $\mathbf{b}_0 = \mathbf{1}\eta$, $\mathbf{B}_0 = \mathbf{I}\tau^2$, $\mathbf{1}$ is the column vector of identity, $\mathbf{I}$ is an identity matrix, and η and τ^2 are prior mean and variance, respectively, for each element of vector $\mathbf{b}$. For normally distributed traits, the prior for σ_e^2 is assumed to be a scaled inverted chi-square distribution with known hyper-parameter values of ν_0 and σ_0^2 so that $\sigma_e^2 \sim \text{Inv-}\chi^2(\nu_0, \sigma_0^2)$.

Since the inheritance state of a QTL depends only on those of the two flanking loci (markers or QTLs), the conditional distribution of QTL segregation indicator matrix, $p(\mathbf{Z}\,|\,l, \lambda)$, can be factorized into $p(\mathbf{Z}\,|\,l, \lambda) = \prod_{q=1}^{l} p(\mathbf{Z}_q\,|\,\mathbf{Z}_q^L, \mathbf{Z}_q^R, \lambda_q)$. Here $\mathbf{Z}_q$ denotes the segregation indicators for all individuals at the qth QTL, i.e. $\mathbf{Z}_q = \{z_{jq}^p, z_{jq}^m\}_{j=1}^n$, and $\mathbf{Z}_q^L$ and $\mathbf{Z}_q^R$ the segregation indicators for the left and right flanking loci (markers or QTLs) of the qth QTL, respectively.

(iii) *Reversible jump MCMC algorithm*

In Bayesian analysis, inferences about the parameters of interest are based on the joint posterior distribution of all the unknowns. Since the joint posterior distribution does not have a standard form, MCMC samplers are used to generate samples from the joint posterior distribution (Hastings, 1970; Geman & Geman, 1984; Green, 1995). The MCMC algorithm usually makes use of the full conditional distribution of some unknowns given the current values of all others, and thus is implemented in an alternating conditional sampling fashion. When the fully conditional distribution is the kernel of a standard density, e.g. normal distribution, the Gibbs sampler is applied to draw samples for that distribution (Geman & Geman, 1984). Otherwise, sampling needs to be done by using the Metropolis–Hastings algorithm (Hastings, 1970), or its extension, reversible jump algorithm (Green, 1995).

For continuous traits, the complete sampling scheme consists of the following update steps:

(a) updating the overall mean and QTL effects $\mathbf{b}$;
(b) updating residual variance σ_e^2;
(c) updating QTL locations λ and segregation indicators $\mathbf{Z}$;
(d) updating the number of QTLs l: adding a new QTL to the model or removing an existing QTL from the model.

For binary traits, we should add an update step to sample the liability for all individuals, and cancel the step for updating residual variance because the residual variance is assumed to be unity in the threshold model. With the underlying liability replaced by the realized sample, other unknowns in the threshold model can be updated using the methods for normally distributed traits.

One complete pass over these update steps defines a cycle of iteration. Starting from an initial point, the algorithm proceeds to update each of the groups of the unknowns in turn until a certain criterion of convergence is reached. Discarding samples of the first few thousand cycles (burn-in period) and thereafter saving one realization in every hundred cycles, we get a random sample from the joint posterior distribution for post-Bayesian analysis.

Except for updating the number of QTLs, all other updating steps are conventional because they do not alter the dimension of the vector of all unknowns, and thus can be implemented using Gibbs samplers or traditional Metropolis–Hastings algorithms. With the

https://doi.org/10.1017/S0016672301005511 Published online by Cambridge University Press

conjugate prior, the full conditional posterior distribution for b_k is normal, i.e.

$$b_k \mid (\mathbf{y}, \theta_{-b_k}) \sim \mathrm{N}\left(\frac{\eta/\tau^2 + \sum_{j=1}^{n} w_{jk}(y_j - \sum_{k' \neq k}^{2l^2} w_{jk'} b_{k'})/\sigma_e^2}{1/\tau^2 + \sum_{j=1}^{n} w_{jk}^2/\sigma_e^2}, \frac{1}{1/\tau^2 + \sum_{j=1}^{n} w_{jk}^2/\sigma_e^2} \right), \quad k = 0, 1, \ldots, 2l^2, \quad (8)$$

where θ_{-b_k} represents all elements of θ except b_k; η and τ^2 are prior mean and variance for b_k, respectively.

Given a scaled inverted chi-square prior distribution, i.e. $\sigma_e^2 \sim \text{Inv-}\chi^2(\nu_0, \sigma_0^2)$, the full conditional posterior distribution of σ_e^2 for normally distributed traits is

$$\sigma_e^2 \mid (\mathbf{y}, \theta_{-\sigma_e^2}) \sim \text{Inv-}\chi^2 \times \left(\nu_0 + n, \frac{\nu_0 \sigma_0^2 + \sum_{j=1}^{n} (y_j - \sum_{k=0}^{2l^2} w_{jk} b_k)^2}{\nu_0 + n} \right), \quad (9)$$

where $\theta_{-\sigma_e^2}$ means all elements of θ except σ_e^2. Therefore, the overall mean, the QTL effects and the residual variance can be easily sampled.

Under the threshold model, the liability y_j, conditional on θ and s_j, also has a simple form. The random variables $y_1, \ldots, y_n$ are independent truncated normal, i.e. $y_j \mid (\theta, s_j)$ is distributed $\mathrm{N}(\mathbf{w}_j^T \mathbf{b}, 1)$ truncated at the left by 0 or at the right by 0, depending on whether $s_j = 1$ or $s_j = 0$. The algorithm for simulating a truncated normal variable described by Devroye (1986) is used to sample the liability.

QTL locations λ and segregation indicators $\mathbf{Z}$ are updated on a locus-by-locus basis. Since the distribution of $\mathbf{Z}_q$ is highly dependent of the location λ_q, it is desirable to jointly update the location and the segregation indicators for each QTL. However, the joint posterior distribution for $\mathbf{Z}_q$ and λ_q has a nonstandard form, i.e.

$$p(\lambda_q, \mathbf{Z}_q \mid \mathbf{y}, \theta_{-(\lambda_q, \mathbf{Z}_q)}) \propto p(\mathbf{y} \mid \theta) p(\lambda_q) p(\mathbf{Z}_q \mid \mathbf{Z}_q^L, \mathbf{Z}_q^R, \lambda_q),$$

where $\theta_{-(\lambda_q, \mathbf{Z}_q)}$ means all elements of θ except λ_q and $\mathbf{Z}_q$. The Metropolis–Hastings algorithm is used to draw samples from this distribution. The jump rule is chosen as follows. First, a new location λ_q^* is sampled from Uniform$[\lambda_q - d, \lambda_q + d]$, where d is a predetermined tuning parameter. Then, new segregation indicators for the jth individual, denoted by $\mathbf{z}_{jq} = \{z_{jq}^m, z_{jq}^p\}$, are generated from the conditional distribution $p(\mathbf{z}_{jq}^* \mid y_j, \theta_{-(\lambda_q, \mathbf{Z}_q)}, \lambda_q^*, \mathbf{z}_{1q}^*, \ldots, \mathbf{z}_{(j-1)q}^*)$. This conditional distribution is discrete and thus is easily sampled (Yi & Xu, 2001). We denote the proposal distributions for λ_q^* and $\mathbf{Z}_q^*$ by $q(\lambda_q^*; \lambda_q)$ and $q(\mathbf{Z}_q^*)$, respectively, where $q(\lambda_q^*; \lambda_q)$ is a uniform density on $[\lambda_q - d, \lambda_q + d]$ and $q(\mathbf{Z}_q^*) = \prod_{j=1}^{n} p(\mathbf{z}_{jq}^* \mid y_j, \theta_{-(\lambda_q, \mathbf{Z}_q)}, \lambda_q^*, \mathbf{z}_{1q}^*, \ldots, \mathbf{z}_{(j-1)q}^*)$. The proposal λ_q^* and $\mathbf{Z}_q^*$ are then accepted with probability $\min\{1, r\}$, where

$$r = \frac{p(\lambda_q^*, \mathbf{Z}_q^* \mid \mathbf{y}, \theta_{-(\lambda_q, \mathbf{Z}_q)}) q(\lambda_q; \lambda_q^*) q(\mathbf{Z}_q)}{p(\lambda_q, \mathbf{Z}_q \mid \mathbf{y}, \theta_{-(\lambda_q, \mathbf{Z}_q)}) q(\lambda_q^*; \lambda_q) q(\mathbf{Z}_q^*)}. \quad (10)$$

Updating the number of QTLs results in a change in the dimension and thus needs a reversible jump step. Instead of drawing a QTL number randomly, the reversible jump step is facilitated by proposing to add or drop a QTL in the model. In a given cycle with l QTLs, there are two types of move: one increasing the number of QTLs to $l+1$ and the other reducing it to $l-1$. Let p_a and $p_d = 1 - p_a$ be the probabilities of choosing either type of move. These proposal probabilities can be arbitrarily chosen as long as they satisfy the conditions: $p_a = 0$ if $l = l_{\max}$ and $p_d = 0$ if $l = 0$. Here we choose $p_a = p_d = 0{\cdot}5$ if $0 < l < l_{\max}$.

When addition of a new QTL is proposed, a new location, new segregation indicators and new effects will be generated for the proposed new QTL. The new effects include the additive, dominance effects for the new QTL, and all interactions (epistatic effects) between the new QTL and all existing QTLs. Denote the new location by λ^*, the new segregation indicators by $\mathbf{Z}^* = \{z_{j(l+1)}^p, z_{j(l+1)}^m\}_{j=1}^n$ and the new effects by $\mathbf{b}^* = (a_{l+1}, d_{l+1}, aa_{1(l+1)}, \ldots, aa_{l(l+1)}, ad_{1(l+1)}, \ldots, ad_{l(l+1)}, da_{1(l+1)}, \ldots, da_{l(l+1)}, dd_{1(l+1)}, \ldots, dd_{l(l+1)})^T$. The efficiency of the reversible jump step greatly depends on the proposal distribution of the parameters. These parameters are generated as follows:

1. Sample the location λ^* from a uniform distribution over the whole genome;

2. Sample the segregation indicator $\mathbf{z}_j^* = \{z_{j(l+1)}^p, z_{j(l+1)}^m\}$ for individual j according to

$$p(\mathbf{z}_j^* \mid \lambda^*, \mathbf{z}_j^L, \mathbf{z}_j^R), \quad j = 1, \ldots, n,$$

where $\mathbf{z}_j^L$ and $\mathbf{z}_j^R$ are the segregation indicators for the left and the right flanking loci of the new location λ^* (markers or QTLs), respectively. The proposal probability for $\mathbf{Z}^*$ is

$$p(\mathbf{Z}^*) = \prod_{j=1}^{n} p(\mathbf{z}_j^* \mid \lambda^*, \mathbf{z}_j^L, \mathbf{z}_j^R).$$

The corresponding values for the coefficients of the effects $\mathbf{b}^*$, denoted as $\mathbf{W}^*$, are then calculated from the sampled $\mathbf{Z}^*$.

3. Note that the full conditional distribution for the new effects $\mathbf{b}^*$ is multivariate normal $\mathbf{b}^* \mid \mathbf{y}, \theta, \mathbf{W}^* \sim \mathrm{N}(\mathbf{b}^*, \mathbf{B}^*)$, where $\mathbf{b}^* = (\mathbf{B}_0^{*-1} + \mathbf{W}^{*T}\mathbf{W}^*)^{-1}[\mathbf{B}_0^{*-1}\mathbf{b}_0^* + \mathbf{W}^{*T}(\mathbf{y} - \mathbf{W}^{\mathrm{T}}\mathbf{b})]$ and $\mathbf{B}^* = (\mathbf{B}_0^{*-1} + \mathbf{W}^{*T}\mathbf{W}^*)^{-1}\sigma_e^2$ with $\mathbf{b}_0^*$ and $\mathbf{B}_0^*$ as the prior mean and prior variance for

https://doi.org/10.1017/S0016672301005511 Published online by Cambridge University Press

b*, respectively. The elements of vector **b*** are sequentially sampled from the conditional distributions,

$p(b_1^*|\mathbf{y}, \theta, \mathbf{W}^*), p(b_2^*|\mathbf{y}, \theta, \mathbf{W}^*, b_1^*),$
$\ldots, p(b_{2+4l}^*|\mathbf{y}, \theta, \mathbf{W}^*, b_1^*, \ldots, b_{1+4l}^*).$

These conditional distributions are univariate normal

$$b_k^*|\mathbf{y}, \theta, \mathbf{W}^*, b_1^*, \ldots, b_{k-1}^* \sim \mathrm{N}\left(\frac{\eta/\tau^2+\sum_{j=1}^n w_{jk}^*(y_j-\mathbf{w}_j^T\mathbf{b}-\sum_{k'=1}^{k-1}w_{jk'}^*b_{k'}^*)/\sigma_e^2}{1/\tau^2+\sum_{j=1}^n w_{jk}^{*2}/\sigma_e^2}, \frac{1}{1/\tau^2+\sum_{j=1}^n w_{jk}^{*2}/\sigma_e^2}\right),$$
$$k = 1, \ldots, 2+4l, \quad (11)$$

where w_{jk}^* is the jkth element of the coefficient matrix **W***.

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

$$r = \frac{p(\mathbf{y}|\theta^*)}{p(\mathbf{y}|\theta)} \cdot \frac{p(l+1)\cdot p(\mathbf{b}^*)}{p(l)} \cdot \frac{\frac{p_d}{l+1}}{p_a \cdot p(\mathbf{b}^*|\mathbf{y}, \theta, \mathbf{W}^*)}, \quad (12)$$

where $\theta^* = (\theta, \lambda^*, \mathbf{b}^*, \mathbf{Z}^*)$ with l in θ replaced by $(l+1)$.

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

$$r = \frac{p(\mathbf{y}|\theta^*)}{p(\mathbf{y}|\theta)} \cdot \frac{p(l)}{p(l-1)\cdot p(\mathbf{b}^*)} \cdot \frac{p_a \cdot p(\mathbf{b}^*|\mathbf{y}, \theta^*, \mathbf{W}^*)}{\frac{p_d}{l}}, \quad (13)$$

where θ^* is θ with the items corresponding to the deleted QTL removed, and **b*** and **W*** are the effects and the coefficients of the deleted QTL.

4. Simulation studies

(i) *Designs of the simulation experiments*

Two inbred lines were crossed to produce the hybrid generation F1. A total of 250 F2 individuals were obtained by selfing F1. These F2 individuals were crossed back to the F1 to produce 250 individuals. The mapping population consists of 503 individuals, including two inbred parents and one F1 hybrid. One chromosome of length 100 cM was simulated. Eleven co-dominant markers were evenly placed on the chromosome with marker intervals of 10 cM each. A quantitative trait y was modelled as being controlled by two or three QTLs residing on the simulated chromosome and a random environmental deviate distributed as $N(0, \sigma_e^2)$ where $\sigma_e^2 = 1.0$ was assumed. We designed three simulation experiments. The true locations, additive and dominance effects of the simulated QTLs, and the epistatic effects between the simulated QTLs are given in Table 1. In design I, the first QTL has additive and dominance effects that account for 12·3% and 8·4% of the phenotypic variation of the trait, respectively, whereas the second QTL exhibits no marginal effect. However, the two loci exhibit an additive-by-additive epistasis that accounts for an additional 12·76% of the phenotypic variance. Design II is a classical complementary model of digenic epistasis, where the two simulated QTLs exhibit not only marginal effects but also all epistatic effects. In design III, the first and the second QTLs show additive, dominance, additive-by-additive and additive-by-dominance effects, whereas the third QTL has no marginal effect but exhibits additive-by-additive and additive-by-dominance interactions with the first and second QTLs. It is noted that there are a total of eight non-existing effects in design III (one additive, one dominance and six epistatic effects).

In addition to the normal trait, we also generated a binary phenotype for each individual using the normally distributed phenotypic value as the underlying liability. The binary phenotype took a value of 1 if the liability $y \geqslant 0$, and a value of 0 otherwise. The overall means were set at $-0{\cdot}5$ for design I, $-0{\cdot}7$ for design II and $-0{\cdot}9$ for design III. The binary trait incidences were 52%, 49% and 51% for the three designs, respectively. We analysed both the normally distributed trait and the binary trait using the proposed method.

Two different models were used to analyse the simulated data. The non-epistatic model included only the additive and dominance effects, ignoring all epistasis. The second model (the epistatic model) includes all two-locus epistatic effects. In all analyses, the same starting values and prior distributions were used. The MCMC algorithm started with no QTL. The starting values for the overall mean and the residual variance were 0·0 and 1·0, respectively. The truncated Poisson prior was used for the number of QTL, with a mean of $\mu = 2$ and a maximum number of $l_{\max} = 6$. The prior for the overall mean was distributed as N(0, 2). For the normal data, a flat but bound prior was chosen for the residual variance. The priors for all QTL effects were chosen to be N(0, 1) for most analyses. The prior variance was slightly smaller than the simulated phenotypic variance. To check the influence of the prior variances of the QTL effects on

https://doi.org/10.1017/S0016672301005511 Published online by Cambridge University Press

Table 1. *The true locations, additive and dominance effects of the simulated QTLs, and the epistatic effects between the simulated QTLs. The heritability of each type of effect is defined as the proportion of the phenotypic variance explained by that effect* (*in parentheses*)

Design:	I		II		III		
Location:	25	55	25	55	25	55	85
a	0·6000 (0·1227)	0·0000 (0·0000)	0·4000 (0·0569)	0·4000 (0·0587)	0·3000 (0·0259)	0·3000 (0·0258)	0·0000 (0·0000)
d	0·7000 (0·0838)	0·0000 (0·0000)	0·4000 (0·0305)	0·4000 (0·0306)	0·4000 (0·0222)	0·4000 (0·0222)	0·0000 (0·0000)
aa	0·8500 (0·1276)		0·4000 (0·0284)		0·6000[a] (0·0543)	0·7000[b] (0·0740)	0·7000[c] (0·0694)
ad	0·0000 (0·0000)		0·4000 (0·0217)		0·7000 (0·0356)	0·7000 (0·0643)	0·7000 (0·0544)
da	0·0000 (0·0000)		0·4000 (0·0235)		0·0000 (0·0000)	0·0000 (0·0000)	0·0000 (0·0000)
dd	0·0000 (0·0000)		0·4000 (0·0276)		0·0000 (0·0000)	0·0000 (0·0000)	0·0000 (0·0000)
b_0	−0·5000		−0·7000			−0·9000	
σ_e^2	1·0000		1·0000			1·0000	

[a] Epistatic effect between the first QTL and the second QTL.
[b] Epistatic effect between the first QTL and the third QTL.
[c] Epistatic effect between the second QTL and the third QTL.

Table 2. *Estimate of the posterior distribution of the QTL number and its expectation*

Design	Model	Data type	Estimated distribution for $l =$ 0	1	2	3	4	5	6	Estimated expectation
I	Non-epistasis	Normal	0·0000	0·8628	0·1309	0·0061	0·0002	0·0000	0·0000	1·1437
		Binary	0·0000	0·9269	0·0709	0·0021	0·0001	0·0000	0·0000	1·0754
	Epistasis	Normal	0·0000	0·0132	0·9821	0·0046	0·0001	0·0000	0·0000	1·9916
		Binary	0·0000	0·0204	0·9789	0·0007	0·0000	0·0000	0·0000	1·9803
II	Non-epistasis	Normal	0·0000	0·0002	0·9021	0·0943	0·0034	0·0000	0·0000	2·1009
		Binary	0·0000	0·0372	0·8709	0·0883	0·0036	0·0000	0·0000	2·0583
	Epistasis	Normal	0·0000	0·0001	0·9996	0·0003	0·0000	0·0000	0·0000	2·0002
		Binary	0·0000	0·0812	0·9166	0·0021	0·0001	0·0000	0·0000	1·9211
		Normal[a]	0·0000	0·0001	0·9900	0·0081	0·0011	0·0007	0·0000	2·0123
		Normal[b]	0·0000	0·0001	0·9998	0·0001	0·0000	0·0000	0·0000	2·0000
III	Non-epistasis	Normal	0·0000	0·0004	0·1797	0·5234	0·2594	0·0351	0·0020	3·1551
		Binary	0·0000	0·0411	0·4788	0·3863	0·0862	0·0075	0·0001	2·5405
	Epistasis	Normal	0·0000	0·0000	0·0001	0·9961	0·0038	0·0000	0·0000	3·0039
		Binary	0·0000	0·0000	0·0161	0·9823	0·0016	0·0000	0·0000	2·9855

[a] The prior variances for all QTL effects are 0·5.
[b] The prior variances for all QTL effects are 2·0.

the performance of MCMC, we also analysed the normal data with two different prior variances (0·5 and 2·0) for design II in the epistatic model. Finally, the tuning parameters of proposal distributions in the Metropolis–Hastings sampling were chosen to be 2·0 cM for QTL locations.

The proposed MCMC sampler was run for 10^6 cycles in each of the MCMC analyses after discarding the first 2000 cycles for the burn-in period. On a Sun SPARC 5 workstation, each analysis took approximately 9 hours. The chains were thinned (by saving one iteration in every 50 cycles) to reduce serial correlation in the stored samples so that the total number of observations kept in the post-Bayesian sample was 20000 for each parameter. The stored samples were used to infer the statistical properties of the parameters of interest.

(ii) Results

The approximate posterior distributions for the number of QTL are presented in Table 2. For design

https://doi.org/10.1017/S0016672301005511 Published online by Cambridge University Press

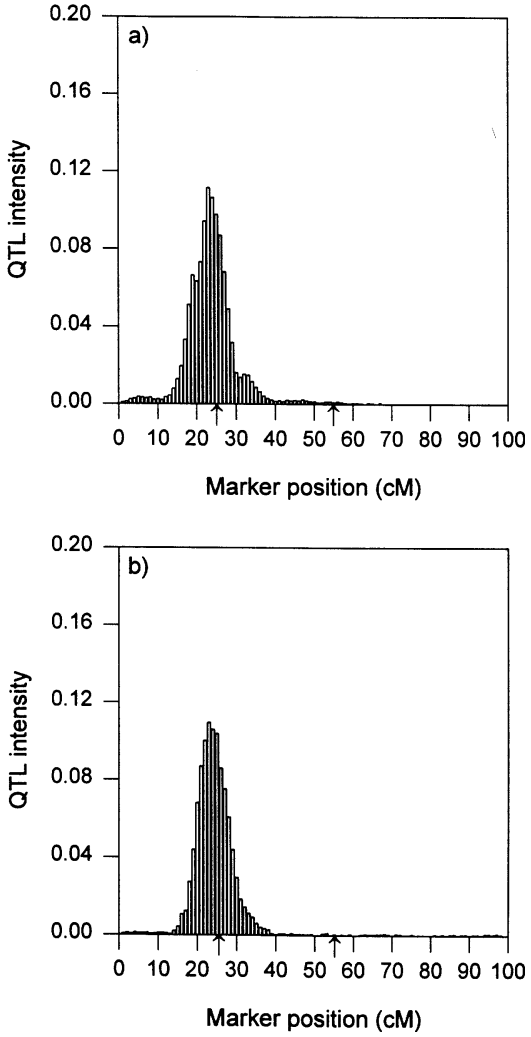

Fig. 1. Analysis of design I under the non-epistatic model. Histograms of the posterior QTL intensity for (*a*) normal data and (*b*) binary data. The true positions of the two QTLs are indicated by the arrows on the horizontal axes.

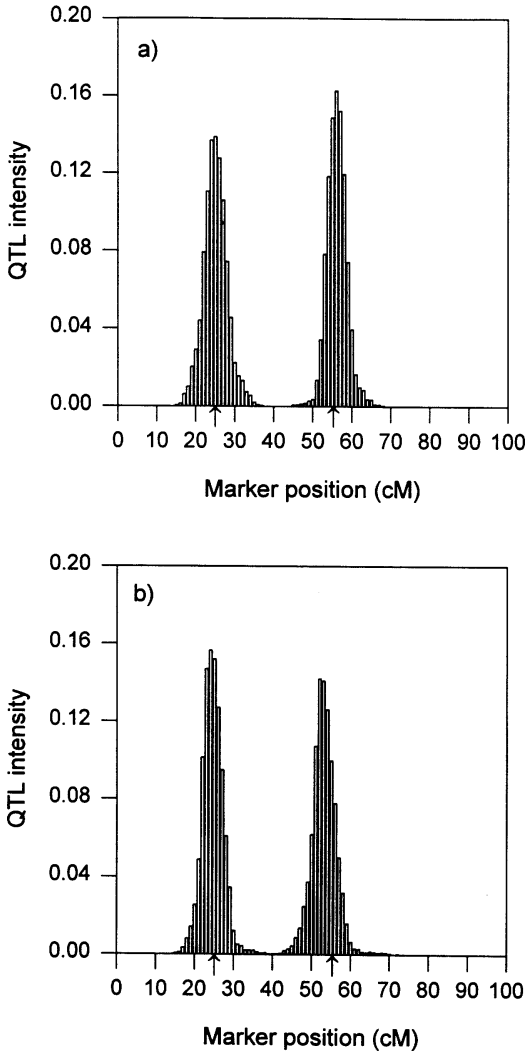

Fig. 2. Analysis of design I under the epistatic model. Histograms of the posterior QTL intensity for (*a*) normal data and (*b*) binary data. The true positions of the two QTLs are indicated by the arrows on the horizontal axes.

I under the non-epistatic model, the posterior modes are 1 for both types of data, whereas the true number of QTLs is 2. The QTL intensity profiles for design I under the non-epistatic model are given in Fig. 1. The major peaks of the profiles for both the normal and the binary data occur at 23 cM. Although the first QTLs were located reasonably for both types of data, the second QTLs were not detected. The fact that the second simulated QTL remains undetected from this analysis is expected because this QTL influences the trait only through interaction with the first QTL and exhibits no marginal effect.

For design I under the epistatic model, the posterior modes for the number of QTLs are 2 for both types of data, which coincides with the true number of QTLs. The posterior expectations are also close to the true number of QTLs (Table 2). These analyses strongly support a model with two QTLs in the chromosome for both types of data. The QTL intensity profiles for these analyses are depicted in Fig. 2*a* for the normal data and Fig. 2*b* for the binary data. The QTL intensity profiles are concentrated around the true locations of the simulated QTLs. The first peak of QTL intensities occurs at 25 cM for normal data and 24 cM for binary data, while the second peak occurs at 56 cM for normal data and 52 cM for binary data. The estimates of QTL locations are close to the true values. The results indicate that the epistatic effect model allows the detection of QTLs with no marginal but epistatic effects.

For design II, the two simulated QTLs were both detected in all analyses. The posterior modes for the posterior distributions of the QTL number overlaps with the true number for the two types of data under the non-epistatic and the epistatic models (Table 2). These results are expected because both QTLs were

https://doi.org/10.1017/S0016672301005511 Published online by Cambridge University Press

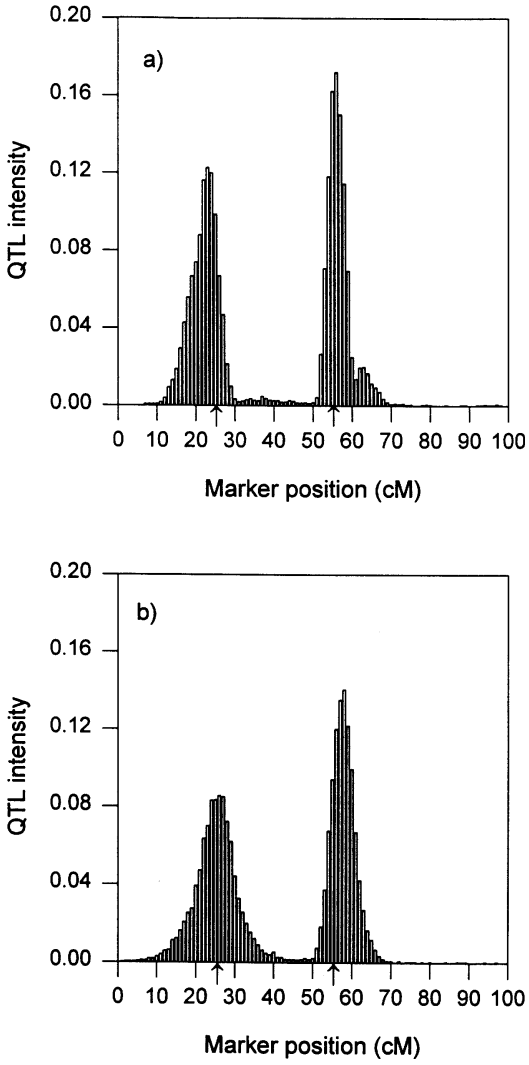

Fig. 3. Analysis of design II under the non-epistatic model. Histograms of the posterior QTL intensity for (*a*) normal data and (*b*) binary data. The true positions of the two QTLs are indicated by the arrows on the horizontal axes.

simulated to have marginal effects in design II. However, it has been observed from Table 2 that the estimated posterior variance of the number of QTL under the epistatic model is smaller than that under the non-epistatic model. The QTL intensities are shown in Fig. 3 and 4 for the non-epistatic and the epistatic models, respectively. The fact that each graph has two major peaks apparently supports two QTLs residing at this chromosome. In addition, the graphs are concentrated around the simulated locations, indicating that the locations of the two QTLs are estimated reasonably. Comparing Fig. 3 and Fig. 4, however, it can be seen that the posterior variance of the locations under the epistatic model is smaller than that under the non-epistatic model. These results indicate that allowing for epistasis when it is present should improve statistical power to detect QTLs and the precision of their localization.

In design III, three QTLs were simulated at 25 cM, 55 cM and 85 cM, respectively. The first and the second QTLs, which both show marginal effects, were detected in the non-epistatic model analysis for the normal data and the binary data (Fig. 5). The third QTL, which exerts no marginal effects, was estimated very inaccurately for the normal data and remained undetected in the binary data analysis. With the inclusion of epistatic effects, however, all three simulated QTLs were detected. In the epistatic model analysis, the posterior modes for the number of QTLs coincide with the true number of QTLs for both types of data, and the posterior expectations are essentially equal to the true number of QTLs (Table 2). It can be seen, from Table 2, that the estimated posterior variance of the number of QTLs under the epistatic model is much smaller than that under the non-epistatic model. The QTL intensity profiles for epistatic model analyses are depicted in Fig. 6*a* for the normal data and Fig. 6*b* for the binary data. The QTL intensity profiles are concentrated around the true locations of the three simulated QTLs.

The chromosome regions with sufficiently high posterior QTL intensity are given in Table 3 for the non-epistatic model and Table 4 for the epistatic model. We used only the posterior samples, in which QTL locations fall into these regions, to estimate the QTL effects and the QTL locations. From Table 3, it can be observed that the estimates of the dominance effects are rather inaccurate, particularly in design III. Under the epistatic model, however, the estimates of the QTL effects are reliable in most cases. The estimates of the QTL locations are close to the simulated values for all cases. As expected, all parameters of interest were estimated more accurately in normal data analysis than in binary data analysis. The estimates and standard errors for the overall mean and the residual variance are also given in Tables 3 and 4. From Table 3, we can see that the residual variances were slightly overestimated in designs I and II, and seriously overestimated in design III, for the normal data when ignoring the epistatic effects. This result is expected because the variation of the epistasis is absorbed into the residual error when ignoring the epistatic effects. It is also observed that the overall means were overestimated in design I and design III. Under the epistatic model, the estimates of the residual variance and the overall mean are close to the simulated values for all analyses (Table 4).

Plots of the changes in the number of QTLs against the number of the iterations for all analyses were drawn using the posterior sample of the QTL number (not shown here). These plots showed that the MCMC algorithm mixes well over the number of QTL, changing frequently but being centralized around the posterior mode of the QTL number. Under the non-epistatic model, the observed acceptance proportions

https://doi.org/10.1017/S0016672301005511 Published online by Cambridge University Press

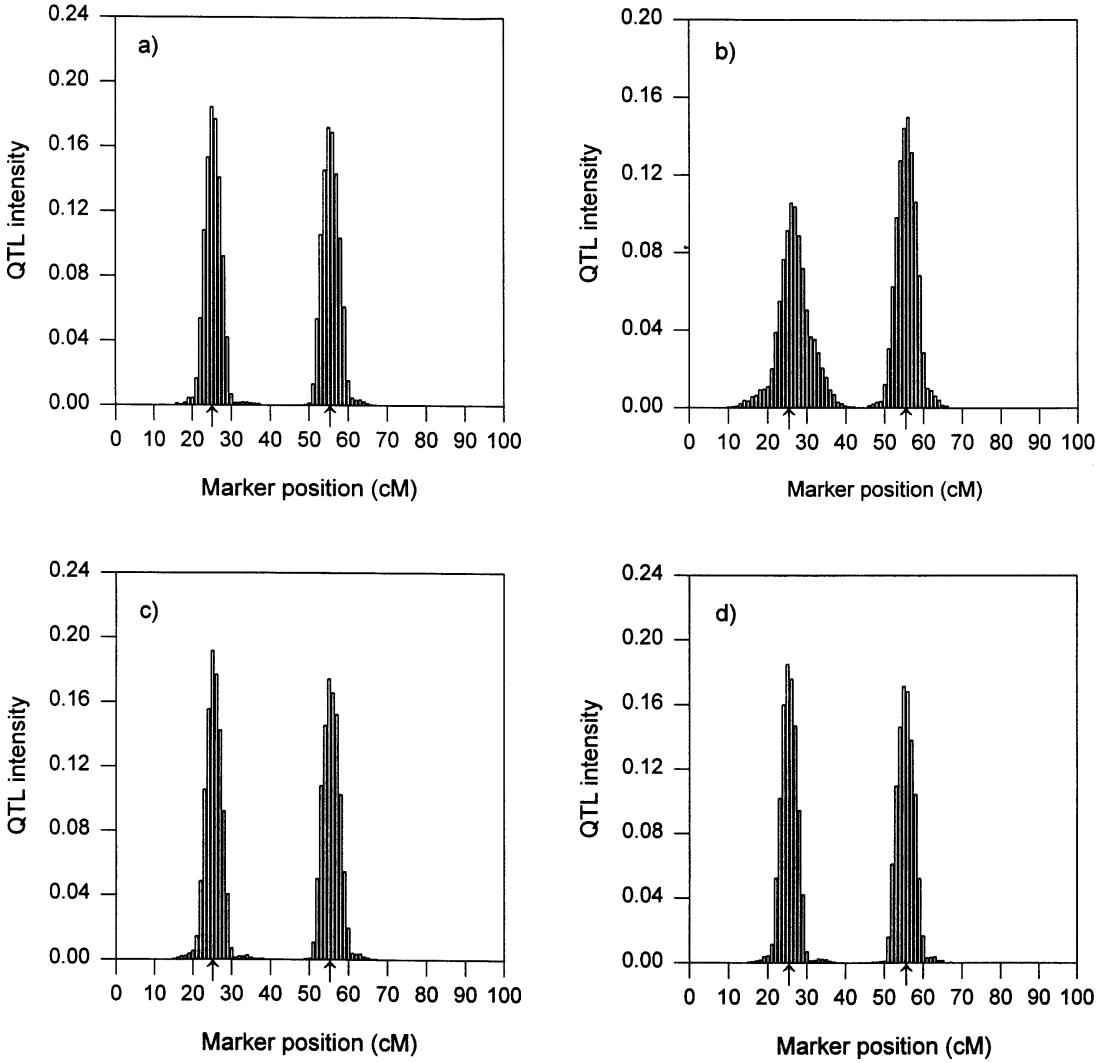

Fig. 4. Analysis of design II under the epistatic model. Histograms of the posterior QTL intensity for (*a*) normal data for prior variance $\tau^2 = 1{\cdot}0$, (*b*) binary data for prior variance $\tau^2 = 1{\cdot}0$, (*c*) normal data for prior variance $\tau^2 = 0{\cdot}5$, and (*d*) normal data for prior variance $\tau^2 = 2{\cdot}0$. The true positions of the two QTLs are indicated by the arrows on the horizontal axes.

for both adding and deleting QTLs were approximately 7% for both the normal and the binary data in all designs. The acceptance rates for the analyses of the epistatic model in design I and II were approximately 3%. This is expected because the epistatic model has many more parameters than the non-epistatic model. The acceptance rates for the analyses of the epistatic model in design III were slightly lower.

In Bayesian analysis it is important to investigate the influences of the choice of initial values and prior distributions for the unknowns on the performance of the MCMC algorithm. We found that our algorithm is quite robust to the initial values of the number of QTLs. For example, when we started with $l_0 = 0$, the number of QTLs l increased to the simulated value after several hundred iterations, and then changed frequently around the true value, whereas when we started with $l_0 = 6$, the number of QTLs quickly decreased to 0 and then increased to the true value, and subsequently stabilized around the true value. We also tried different initial values for the overall mean, the residual variance and other parameters and found little influence on the MCMC performance.

We also investigated the sensitivity to the choice of prior variances for QTL effects. For design II, the posterior distributions of the QTL number and the QTL intensities for three different prior variances are given in Table 2 and Fig. 4, respectively, for the epistatic model analysis of the normal data. The posterior mode of the number of QTLs does not seem to be affected by the choice of prior variance. The posterior probabilities and the posterior mean were slightly affected by the choice of prior variance (Table 2). In general, reducing the prior variance favours a higher number of QTLs and a larger posterior mean. The QTL intensity profiles are compared for three different prior variances (Fig. 4). The prior variance does not seem to affect the estimates of the locations

https://doi.org/10.1017/S0016672301005511 Published online by Cambridge University Press

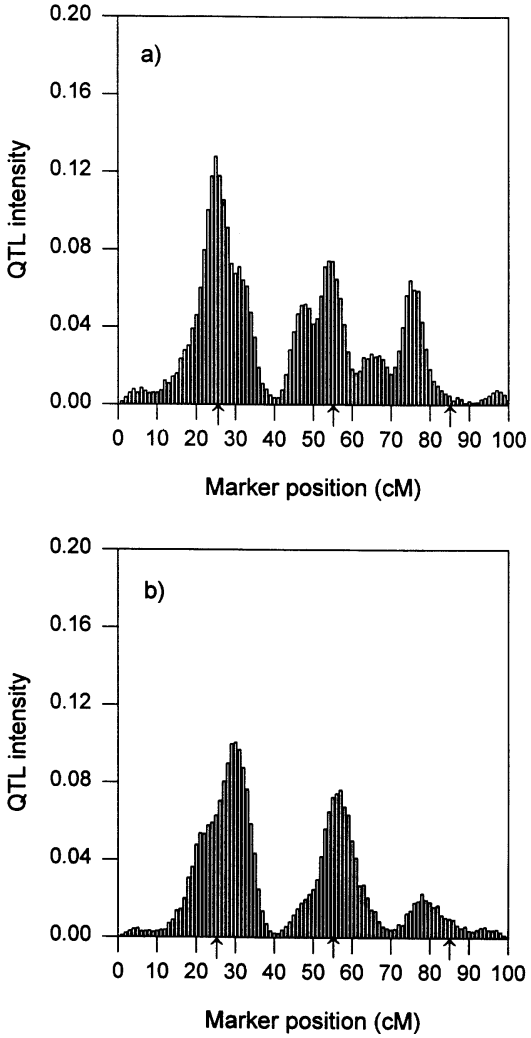

Fig. 5. Analysis of design III under the non-epistatic model. Histograms of the posterior QTL intensity for (*a*) normal data and (*b*) binary data. The true positions of the two QTL are indicated by the arrows on the horizontal axes.

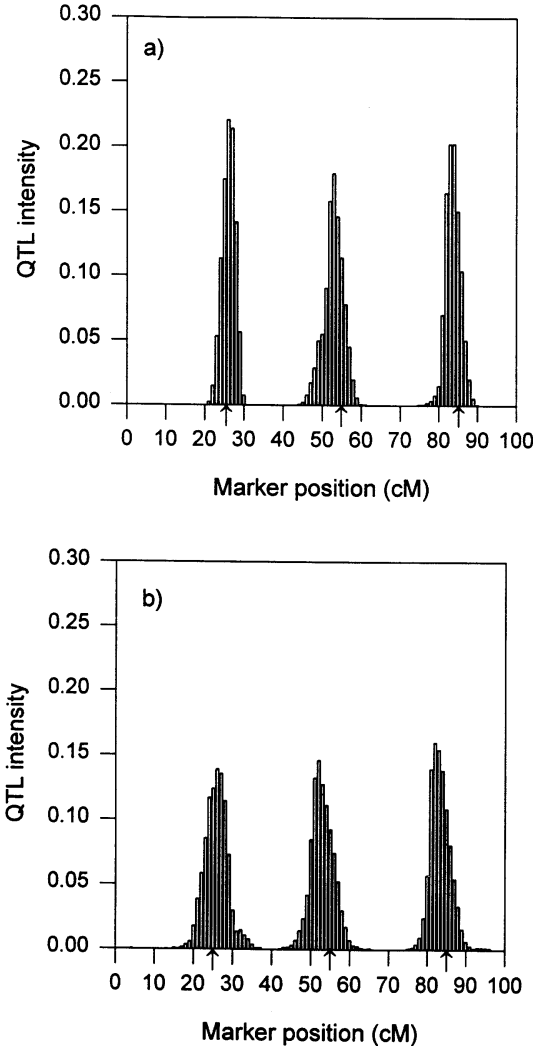

Fig. 6. Analysis of design III under the epistatic model. Histograms of the posterior QTL intensity for (*a*) normal data and (*b*) binary data. The true positions of the two QTLs are indicated by the arrows on the horizontal axes.

of QTLs. The estimates for the population mean, the residual variance and the QTL effects are essentially the same for the three choices of prior variance (Table 4).

5. Discussion

Epistasis, an important genetic component underlying many complex traits, has not been extensively explored in QTL analysis. In this study, we have developed a Bayesian approach for mapping epistatic QTLs for both normally distributed and binary traits in arbitrary mating designs derived from two inbred lines. In QTL mapping studies, estimating the number of QTLs is of major importance. To further understand the genetic architecture of a complex trait, it is also important to know which of all the possible main and interaction effects are contributing to the genetic variance. Our Bayesian method can make a joint statistical inference about the number, locations, marginal and epistatic effects of QTL. In the proposed reversible jump MCMC algorithm, the number of QTLs is updated via two reversible jump steps: adding one QTL into the model or removing one QTL from the model. Such an algorithm is expected to be useful in situations where there is at least a significant marginal effect of one QTL. The proposed method has been successfully applied to three simulated designs. However, the algorithm may not be adequate in cases where none of the QTL has marginal effects but they exhibit an epistatic effect. In such situations, we need to add two QTLs into the model simultaneously. Such an additional reversible jump step can easily be incorporated into our procedure.

In this study we included all possible QTL effects in the model and applied the reversible jump only to the number of QTLs. Our method shows that sampling of new parameters under the epistatic model can be

https://doi.org/10.1017/S0016672301005511 Published online by Cambridge University Press

Table 3. *Non-epistatic model: highest posterior QTL intensity interval, Bayesian mean estimates of QTL locations and allelic and dominance effects. Posterior standard errors of the estimates are given in parentheses*

Design	Data type	Interval	Sum of intensity	Mode of location	Mean of location	a	d	b_0	σ_e^2
I	Normal	15–34	0·985	23	23·101	0·642	0·271	0·032	1·299
					(1·150)	(0·151)	(0·148)	(0·071)	(0·174)
	Binary	16–33	0·998	23	23·664	0·587	0·294	−0·113	
					(1·168)	(0·168)	(0·184)	(0·350)	
II	Normal	14–30	0·925	23	22·704	0·408	0·296		
					(1·149)	(0·149)	(0·147)	−0·486	1·224
		52–65	0·966	56	56·161	0·569	0·549	(0·079)	(0·067)
					(2·596)	(0·115)	(0·138)		
	Binary	14–35	0·944	26	24·662	0·438	0·295		
					(1·197)	(0·197)	(0·207)	−0·529	
		52–65	0·979	58	57·378	0·619	0·611	(0·124)	
					(2·680)	(0·165)	(0·184)		
III	Normal	16–35	0·999	25	25·722	0·609	−0·029	0·126	1·708
					(1·286)	(0·286)	(0·383)	(0·115)	(0·104)
		44–60	0·809	54	51·847	0·433	−0·184		
					(4·119)	(0·223)	(0·257)		
	Binary	16–35	0·999	30	27·088	0·455	−0·184	0·267	
					(1·233)	(0·233)	(0·360)	(0·116)	
		44–66	0·941	57	55·499	0·348	−0·347		
					(4·696)	(0·174)	(0·168)		

performed using the reversible jump MCMC algorithm. The model determination in terms of the QTL effects was inferred by simply examining the posterior estimates of the QTL effects. In all analyses of three simulated designs, Bayesian credibility intervals do not include 0 for all existing QTL marginal effects and epistatic effects and include 0 for all non-existing effects, which strongly supports the epistatic models for the simulated data. However, we found that the accurate estimation of epistatic effects depends strongly on the size of mapping population. With insufficient sample sizes, therefore, more formal statistical tools, e.g. Bayes factor, may be required to assess the model. Another possibility to determine the model in terms of QTL effects is to include the number of epistatic effects as a random variable. Although this strategy will add another level of complexity to the analysis, it deserves further study.

Conditional on marker information, QTL segregation indicators are highly dependent of the QTL position. Therefore, the two groups of variables must be updated jointly using the Metropolis–Hastings algorithm. This joint updating strategy has been adopted in our previous work (Yi & Xu, 2001). However, we found in this study that extra steps to redraw the QTL meiosis indicators could be omitted. This omission does not affect the behaviour of the MCMC algorithm, but does speeds it up. A similar algorithm has also been used in the reversible jump MCMC under the identity-by-descent-based variance component model, in which the corresponding IBD matrices are formed when the QTL position is updated (Yi & Xu, 2000*b*). It has been shown, from our extensive simulation studies, that the joint updating scheme can greatly improve the mixing of the MCMC.

The mixing behaviour of the reversible jump algorithm strongly depends on the method of generating new parameters when a new QTL is proposed. Sampling new parameters in an arbitrary fashion and making no reference to the current values of other parameters may result in a rather low acceptance rate. The problem can be quite serious when the number of new parameters to be sampled is large, as seen in our epistatic model. Previously, the proposed new QTL effects have always been sampled from their prior distributions (Stephens & Fisch, 1998; Sillanpää & Arjas, 1998, 1999; Yi & Xu 2000*a*, 2001). As such, the mixing behaviour is highly sensitive to the prior chosen. To facilitate a better fit to the data, the current effects of old QTLs should be modified when a new QTL is added to the model (Satagopan & Yandell, 1998). However, this is computationally infeasible in the epistatic model due to the need for inverting a high-dimensional matrix. In this study, we generated the new QTL effects from the conditional posterior distribution $p(\mathbf{b}^* \mid \mathbf{y}, \theta, \mathbf{W}^*)$ via a sequential sampling scheme. This has eliminated the need for inverting a large matrix. By combining the current values of other parameters and the phenotypic values in the proposal distribution, the newly generated parameters fit the data better and thus reduce the dependence on the choice of the prior variance of QTL effects. Another advantage of the proposed algorithm is that the Jacobian involved in the

https://doi.org/10.1017/S0016672301005511 Published online by Cambridge University Press

Table 4. *Epistatic model: highest posterior QTL intensity interval, Bayesian mean estimates of QTL locations and allelic and dominance effects. Posterior standard errors of the estimates are given in parentheses*

Design	Data type	Interval	Sum of intensity	Mode of location	Mean of location	a	d	aa	ad	da	dd	b_0	σ_e^2
I	Normal	18–32	0·961	25	24·502	0·727	0·438	0·759	0·132	−0·182	0·164	−0·366	1·002
					(1·235)	(0·235)	(0·249)	(0·223)	(0·262)	(0·265)	(0·298)	(0·221)	(0·067)
		48–63	0·984	56	55·605	−0·038	−0·076						
					(2·371)	(0·243)	(0·256)						
	Binary	18–32	0·985	24	24·091	0·529	0·561	1·036	0·114	0·128	0·204	−0·723	
					(1·382)	(0·382)	(0·383)	(0·370)	(0·438)	(0·439)	(0·438)	(0·371)	
		47–61	0·926	52	52·783	0·153	−0·156						
					(2·574)	(0·386)	(0·394)						
II	Normal	18–32	0·988	25	24·814	0·304	0·298	0·325	0·568	0·530	0·484	−0·564	0·934
					(1·148)	(0·148)	(0·186)	(0·147)	(0·188)	(0·196)	(0·238)	(0·145)	(0·064)
		50–61	0·986	55	55·053	0·404	0·357						
					(2·031)	(0·146)	(0·190)						
	Binary	18–35	0·863	26	26·428	0·313	0·363	0·435	0·503	0·538	0·291	−0·730	
					(1·291)	(0·291)	(0·311)	(0·287)	(0·366)	(0·358)	(0·214)	(0·289)	
		50–63	0·973	56	55·204	0·584	0·657						
					(2·401)	(0·289)	(0·351)						
	Normal[a]	18–32	0·886	25	24·835	0·305	0·303	0·209	0·559	0·519	0·470	−0·562	0·944
					(1·144)	(0·144)	(0·181)	(0·145)	(0·187)	(0·193)	(0·228)	(0·141)	(0·064)
		50–63	0·994	55	55·012	0·398	0·359						
					(2·096)	(0·147)	(0·183)						
	Normal[b]	18–32	0·987	25	24·871	0·283	0·294	0·234	0·585	0·556	0·492	−0·561	0·952
					(1·154)	(0·154)	(0·191)	(0·151)	(0·193)	(0·201)	(0·241)	(0·150)	(0·064)
		50–63	0·990	55	55·014	0·391	0·346						
					(2·122)	(0·153)	(0·194)						
III	Normal	20–34	0·999	26	25·576	0·311	0·395	0·652[c]	0·337	0·118	−0·082	−0·810	1·081
					(1·334)	(0·133)	(0·181)	(0·144)	(0·278)	(0·277)	(0·210)	(0·159)	(0·071)
		46–61	0·990	53	52·481	0·303	0·296	0·727[d]	0·679	0·068	−0·118		
					(2·451)	(0·164)	(0·212)	(0·134)	(0·164)	(0·168)	(0·201)		
		77–90	0·997	84	83·257	0·022	−0·099	0·599[e]	0·432	0·276	0·213		
					(1·851)	(0·130)	(0·159)	(0·149)	(0·317)	(0·323)	(0·291)		
	Binary	20–34	0·964	26	25·451	0·405	0·295	0·661[c]	0·266	0·166	0·120	−0·576	
					(1·206)	(0·206)	(0·253)	(0·217)	(0·208)	(0·264)	(0·288)	(0·241)	
		46–61	0·961	52	52·516	0·413	0·294	0·678[d]	0·402	0·033	−0·093		
					(2·719)	(0·224)	(0·274)	(0·203)	(0·281)	(0·274)	(0·264)		
		77–90	0·983	82	82·872	0·034	−0·133	0·673[e]	0·303	−0·061	0·165		
					(2·370)	(0·198)	(0·245)	(0·205)	(0·217)	(0·286)	(0·261)		

[a] The prior variances for all QTL effects are 0·5.
[b] The prior variances for all QTL effects are 2·0.
[c] Epistatic effect between the first QTL and the second QTL.
[d] Epistatic effect between the first QTL and the third QTL.
[e] Epistatic effect between the second QTL and the third QTL.

acceptance probability can be easily evaluated because adding or deleting a QTL does not affect parameters of other QTLs.

The reversible jump MCMC algorithm is computationally intensive. To obtain reliable results from the MCMC output, we have to run a sufficiently long chain to ensure convergence of MCMC, and choose a subsample from this chain to reduce the serial correlation. Recently, Brooks & Giudici (1998) proposed some criteria to evaluate the convergence of the chain. In real data analyses, these diagnostic tools should be used to assess convergence of reversible jump MCMC simulations.

We thank Dr Claus Vogl for helpful comments on the manuscript. This research was supported by the National Institutes of Health Grant GM55321 and the US Department of Agriculture National Research Initiative Competitive Grants Program 00-35300-9245 to S.X.

References

Albert, J. H. & Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association* **88**, 669–679.

Brooks, S. P. & Giudici, P. (1998). Convergence assessment for reversible jump MCMC simulations. Manuscript available at http://www.statslab.cam.ac.uk/~mcmc/pages/list.html

Bulmer, M. G. (1980). *The Mathematical Theory of Quantitative Genetics*. Oxford: Clarendon Press.

Carlborg, O., Andersson, L. & Kinghorn, B. (2000). The use of a genetic algorithm for simultaneous mapping of multiple interacting quantitative trait loci. *Genetics* **155**, 2003–2010.

Churchill, G. A. & Doerge, R. W. (1994). Empirical threshold values for quantitative trait mapping. *Genetics* **138**, 963–971.

Devroye, L. (1986). *Non-uniform Random Variable Generation*. New York: Springer.

Geman, S. & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**, 721–741.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82**, 711–732.

Haley, C. S. & Knott, S. A. (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 for oligogenic models. *American Journal of Human Genetics* **61**, 748–760.

Jannink, J. & Jansen, R. (2001). Mapping epistatic quantitative trait loci with one-dimensional genome searches. *Genetics* **157**, 445–454.

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

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

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

Lark, K. G., Chase, K., Adler, F. R., Mansur, L. M. & Orf, J. J. (1995). Interactions between quantitative trait loci in soybean in which trait variation at one locus is conditional upon a specific allele at another. *Proceedings of the National Academy of Sciences of the USA* **92**, 4656–4660.

Lynch, M. & Walsh, B. (1998). *Genetics and Analysis of Quantitative Traits*. Sunderland, MA: Sinauer Associates.

Routman, E. J. & Cheverud, J. M. (1997). Gene effects on a quantitative trait: two-locus epistatic effect measured at microsatellite markers and at estimated QTLs. *Evolution* **51**, 1654–1662.

Satagopan, J. M. & Yandell, B. S. (1998). Bayesian model determination for quantitative trait loci. Manuscript available at ftp://ftp.stat.wisc.edu/pub/yandell/rev jump.html

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

Sillanpää, M. J. & Arjas, E. (1998). Bayesian mapping of multiple quantitative trait loci from incomplete inbred line cross data. *Genetics* **148**, 1373–1388.

Sillanpää, M. J. & Arjas, E. (1999). Bayesian mapping of multiple quantitative trait loci from incomplete outbred offspring data. *Genetics* **151**, 1605–1619.

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

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

Visscher, P. M., Thomson, R. & Haley, C. S. (1996). Confidence intervals in QTL mapping by bootstrapping. *Genetics* **143**, 1013–1020.

Wang, D. L., Zhu, J., Li, Z. K. & Paterson, A. H. (1999). Mapping QTLs with epistatic effects and QTL × environment interactions by mixed linear model approaches. *Theoretical and Applied Genettics* **99**, 1255–1264.

Xu, S. & Yi, N. (2000). Mixed model analysis of quantitative trait loci. *Proceedings of the National Academy of Sciences of the USA* **97**, 14542–14547.

Yi, N. & Xu, S. (2000*a*). Bayesian mapping of quantitative trait loci for complex binary traits. *Genetics* **155**, 1391–1403.

Yi, N. & Xu, S. (2000*b*). Bayesian mapping of quantitative trait loci under the identity-by-descent-based variance component model. *Genetics* **156**, 411–422.

Yi, N. & Xu, S. (2001). Bayesian mapping of quantitative trait loci under complicated mating designs. *Genetics* **157**, 1759–1771.

Yu, S. B., Li, J. X., Xu, C. G., Tan, Y. F., Gao, Y. J., Li, X. H., *et al.* (1997). Importance of epistasis as the genetic basis of heterosis in an elite rice hybrid. *Proceedings of the National Academy of Sciences of the USA* **94**, 9226–9231.

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

Zeng, Z.-B., Kao, C.-H. & Basten, C. J. (2000). Estimating the genetic architecture of quantitative traits. *Genetical Research* **74**, 279–289.