Original article

Detection of multiple QTL with epistatic effects under a mixed inheritance model in an outbred population

Akira Narita, Yoshiyuki Sasaki*

Laboratory of Animal Breeding and Genetics, Division of Applied Biosciences,
Graduate School of Agriculture, Kyoto University, Kyoto 606-8502, Japan

Abstract – A quantitative trait depends on multiple quantitative trait loci (QTL) and on the interaction between two or more QTL, named epistasis. Several methods to detect multiple QTL in various types of design have been proposed, but most of these are based on the assumption that each QTL works independently and epistasis has not been explored sufficiently. The objective of the study was to propose an integrated method to detect multiple QTL with epistases using Bayesian inference *via* a Markov chain Monte Carlo (MCMC) algorithm. Since the mixed inheritance model is assumed and the deterministic algorithm to calculate the probabilities of QTL genotypes is incorporated in the method, this can be applied to an outbred population such as livestock. Additionally, we treated a pair of QTL as one variable in the Reversible jump Markov chain Monte Carlo (RJMCMC) algorithm so that two QTL were able to be simultaneously added into or deleted from a model. As a result, both of the QTL can be detected, not only in cases where either of the two QTL has main effects and they have epistatic effects between each other, but also in cases where neither of the two QTL has main effects but they have epistatic effects. The method will help ascertain the complicated structure of quantitative traits.

Bayesian inference / multiple QTL / epistasis / outbred population / mixed inheritance model

1. INTRODUCTION

It may be more realistic that interlocus interactions (epistasis) between two or more quantitative trait loci (QTL), as well as the main effects of QTL themselves, play an important role in expressing a quantitative trait, and the importance of exploring epistatic effects has been discussed recently [17, 30].

---

* Corresponding author: sasaki@kais.kyoto-u.ac.jp

Then, several statistical methods for the detection of epistatic QTL, using a maximum-likelihood [25] or least squares approach [6], were proposed. Since these are multiple-dimensional search approaches, however, many computations are involved. Additionally, these are based on the assumption that the QTL number is fixed, and it may not be appropriate to apply these methods to data where the QTL number is unknown. Then, several improved methods have been proposed to detect multiple epistatic QTL, incorporating a genetic algorithm [1], extending the composite interval mapping [12, 30], and adopting a one-dimensional search [10]. At this stage, genome-wide levels of significance and the confidence interval of estimates have to be calculated by using, for example, the permutation test [2] and bootstrap method [24], in which millions of computations are required.

Alternatively, the Bayesian approach *via* Markov chain Monte Carlo (MCMC) algorithms such as the Gibbs sampler [4] and the Metropolis-Hastings (MH) algorithm [8, 15], has been drawing attention as a new and promising approach and was initially used to map QTL by Thaller and Hoeschele [20, 21]. Additionally, owing to the advent of the Reversible jump Markov chain Monte Carlo (RJMCMC) algorithm [5], the number of QTL itself has become estimable, and several methods using the approach were developed [9, 18, 19, 22, 26]. Yi and Xu [27] developed a revolutionary Bayesian method to map multiple epistatic QTL in a population derived from two inbred lines, and by simulation studies. They showed that the method detects QTL accurately when either of the two QTL has main effects and they have epistatic effects, but when neither of the two QTL has main effects but they have epistatic effects, the pair of QTL may not be able to be detected by the method. Additionally, their method is applicable only to an F_2 population derived from two inbred lines.

The objective of this article was to propose an integrated method using Bayesian inference *via* MCMC algorithms for the detection of multiple QTL with epistasis under the mixed inheritance model in an F_2 population derived from two divergent breeds, such as cattle and swine. In this method, by treating a pair of QTL as one variable in the RJMCMC algorithm, the proposed method will work well, not only in cases where either of the two QTL has main effects and they have epistatic effects, but also in cases where neither of the two QTL has main effects but they have epistatic effects. We briefly evaluated the propriety of the proposed method by simulation studies. By using this method, we can simultaneously estimate the number, locations and effects of QTL, whether epistasis exists among the QTL or not, and also the extent to which the polygenic effect affects a quantitative trait.

2. MATERIALS AND METHODS

2.1. Mixed inheritance model

In this article, we assumed an F_2 population generated from two breeds such as swine, cattle and so on, which were the upward- and downward-selected breeds in which different marker alleles were not completely fixed while different alleles at QTL were fixed in each breed.

Here, y represents a vector for the phenotypic values of a quantitative trait. The vector y can be described as

$$y = X\beta + \sum_{i=1}^{l} Q_i \delta_i + \sum_{j=1}^{m} W_j \gamma_j + Zu + e \tag{1}$$

where,

$$Q_i = \begin{bmatrix} p_{ik}(QQ) - p_{ik}(qq) & p_{ik}(Qq) \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}, \quad \delta_i' = \begin{bmatrix} a_i & d_i \end{bmatrix},$$

$$W_j' = \begin{bmatrix} p_{j1k}(QQ) - p_{j1k}(qq) \\ p_{j1k}(Qq) \\ p_{j2k}(QQ) - p_{j2k}(qq) \\ p_{j2k}(Qq) \\ p_{j1k}(QQ)\,p_{j2k}(QQ) + p_{j1k}(qq)\,p_{j2k}(qq) - p_{j1k}(QQ)\,p_{j2k}(qq) \\ -p_{j1k}(qq)\,p_{j2k}(QQ) \\ p_{j1k}(QQ)\,p_{j2k}(Qq) - p_{j1k}(qq)\,p_{j2k}(Qq) \\ p_{j1k}(Qq)\,p_{j2k}(QQ) - p_{j1k}(Qq)\,p_{j2k}(qq) \\ p_{j1k}(Qq)\,p_{j2k}(Qq) \end{bmatrix} \begin{matrix} \\ \cdots \\ \\ \cdots \\ \\ \cdots \\ \\ \cdots \\ \end{matrix},$$

and $\gamma_j' = \begin{bmatrix} a_{j1} & d_{j1} & a_{j2} & d_{j2} & aa_j & ad_j & da_j & dd_j \end{bmatrix}$.

In these equations, β, u ($\sim N(0, A\sigma_u^2)$) and e ($\sim N(0, I\sigma_e^2)$) are vectors of the overall mean and/or covariates, polygenic effect and environmental effects, respectively, where A is the numerator relationship matrix, and σ_u^2 and σ_e^2 are polygenic and environmental variances, respectively. X and Z are design matrices relating β and u to y, respectively. l and m represent the numbers of

Table I. The symbols used to denote all the parameters in this article.

Symbol	Meanings
μ	Overall mean
σ_e^2	Environmental variance
σ_u^2	Polygenic variance
l	Number of QTL with the only main effect
m	Number of pair of epistatic QTL
θ	Location of QTL
a	Additive effect
d	Dominance effect
aa	Epistatic effect (additive-additive)
ad	Epistatic effect (additive-dominance)
da	Epistatic effect (dominance-additive)
dd	Epistatic effect (dominance-dominance)

QTL and pairs of epistatic QTL, respectively. $\boldsymbol{\delta}_i$ is the vector of additive genetic and dominance effects of the ith QTL, and $\boldsymbol{\gamma}_j$ is the vector of additive and dominance effects of the first and second QTL in the jth pair, and epistatic effects of the pair. $\boldsymbol{Q}_i$ is the matrix of probabilities of the ith putative QTL genotypes, and $\boldsymbol{W}_j$ is the matrix of probabilities of two epistatic QTL genotypes and coefficients of epistatic effects in the jth pair, in which $p_{ik}(QQ)$ and $p_{ik}(qq)$ are the conditional probabilities that the individual k has two alleles derived from breeds 1 and 2 for the ith QTL, respectively. $p_{ik}(Qq)$ is the conditional probability that the individual k is heterozygous. In this article, instead of sampling QTL genotypes as one of the parameters, these probabilities are calculated deterministically by the method proposed by Haley *et al.* [7], using only the marker genotypes. These effects are based on the model proposed by Cockerham [3], except that the QTL genotype probabilities are used instead of the QTL genotypes themselves.

For the convenience of the readers, the symbols used to denote all the parameters in this article and their meanings are shown in Table I.

2.2. Statistical analyses

In this study, the Bayesian approach *via* the Gibbs sampler, the MH algorithm and the RJMCMC algorithm were used. The posterior distributions for all the parameters involved can be generated from these algorithms.

2.2.1. Updating all the parameters except the number of QTL

For the overall mean $\boldsymbol{\beta}$, main and epistatic effect of QTL $\boldsymbol{\delta}$ and $\boldsymbol{\gamma}$, polygenic effects $\boldsymbol{u}$, and polygenic variance σ_u^2, the Gibbs sampler was used. In this algorithm, the new value of a parameter is sampled from the distribution conditioned on the current values of all the other parameters and the phenotypic values.

A new value of $\boldsymbol{\beta}$ is sampled from the normal distribution

$$
N \left\{ \frac{\frac{\eta \sigma_e^2}{\tau^2} + X' \left(y - \sum_{i=1}^{l} Q_i \delta_i - \sum_{j=1}^{m} W_j \gamma_j - Zu \right)}{\frac{\sigma_e^2}{\tau^2} + X'X}, \frac{1}{\frac{\sigma_e^2}{\tau^2} + X'X} \right\},
$$

where η and τ^2 are prior mean and variance for the overall mean, respectively.

New values of $\boldsymbol{\delta}$ and $\boldsymbol{\gamma}$ are sampled from the normal distribution

$$
N \left\{ \left(T^{-2} \sigma_e^2 + Q_i' Q_i \right)^{-1} Q_i' \left(y - X\beta - \sum_{i' \neq i}^{l} Q_{i'} \delta_{i'} - \sum_{j=1}^{m} W_j \gamma_j - Zu \right), \right.
$$
$$
\left. \sigma_e^2 \left(T^{-2} \sigma_e^2 + Q_i' Q_i \right)^{-1} \right\}
$$

and

$$
N \left\{ \left(T^{-2} \sigma_e^2 + W_j' W_j \right)^{-1} W_j' \left(y - X\beta - \sum_{i=1}^{l} Q_i \delta_i - \sum_{j' \neq j}^{m} W_{j'} \gamma_{j'} - Zu \right), \right.
$$
$$
\left. \sigma_e^2 \left(T^{-2} \sigma_e^2 + W_j' W_j \right)^{-1} \right\},
$$

respectively.

The prior for the genetic effects is assumed to be distributed as $N(0, \tau^2)$, and $\boldsymbol{T}$ represents a diagonal matrix whose diagonal elements are τ. $\boldsymbol{Q}_i$ and $\boldsymbol{W}_j$ are determined based on the genotypes of flanking markers.

New values of $\boldsymbol{u}$ are sampled from the normal distribution

$$N\left\{\left(\boldsymbol{Z}'\boldsymbol{Z} + \boldsymbol{A}^{-1}\frac{\sigma_e^2}{\sigma_u^2}\right)^{-1}\boldsymbol{Z}'\left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \sum_{i=1}^{l}\boldsymbol{Q}_i\boldsymbol{\delta}_i - \sum_{j=1}^{m}\boldsymbol{W}_j\boldsymbol{\gamma}_j\right),\right.$$

$$\left.\sigma_e^2\left(\boldsymbol{Z}'\boldsymbol{Z} + \boldsymbol{A}^{-1}\frac{\sigma_e^2}{\sigma_u^2}\right)^{-1}\right\}.$$

A new value of σ_u^2 is sampled from the inverted χ^2-distribution

$$\frac{\boldsymbol{u}'\boldsymbol{A}^{-1}\boldsymbol{u}}{\chi^2(q-2)},$$

where q is the order of $\boldsymbol{A}$.

For QTL locations θ and environmental variance σ_e^2, the MH algorithm has been used. At the initial step of the algorithm, an arbitrary value for a parameter k, noted by $k^{[1]}$, is determined and after implementation of the same procedures for all the other parameters in turn, the subsequent value for k is sampled from a symmetric uniform distribution on the interval $[k^{[1]} - d, k^{[1]} + d]$, where d is a predetermined tuning parameter, updated with a probability min $(1, \lambda)$. λ represents the ratio of likelihood when newly proposed values are substituted for previous values. If new proposals are accepted, these unknown parameters are updated, but if not, the previous values remain unchanged. By repeating this procedure many times, posterior distributions for θ and σ_e^2 were generated. Once θ was sampled, the conditional probabilities $p_{ik}(QQ)$, $p_{ik}(qq)$ and $p_{ik}(Qq)$ were also determined.

2.2.2. Updating the number of QTL

Updating the number of QTL and epistatic QTL pairs, noted by l and m, respectively, needs a change in the dimension of the linear model, so the RJMCMC algorithm was used. The prior distribution of the number of QTL is the Poisson distribution with a mean μ_{l+m}, and with $l_{\max}$ and $m_{\max}$ being set as the upper limits of QTL number and QTL pair number, respectively. In this article, we treated an epistatic QTL pair in the same way as a QTL with only a main effect. Therefore, if a QTL has both a main effect and epistasis with other QTL, the QTL is treated as one of a QTL pair and its main effect belongs to the vector γ. The new proposals can be one of the five following with the probability p_a, p_{ae}, p_m, p_d, and p_{de}: to add one QTL (noted by proposal 1),

to add one pair of QTL (noted by 2), to leave the QTL number unchanged (noted by 3), to delete one QTL (noted by 4), and to delete one pair of QTL (noted by 5). In this study, p_a, p_{ae}, p_m, p_d, and p_{de} are all equal, being one-fifth. When both l and m are 0, p_d is 0 and when l and m are $l_{\max}$ and $m_{\max}$, respectively, p_a is 0. When the new proposal is 1, the location and the effects of a new QTL are sampled and the likelihood is recalculated, and the proposal is accepted with the probability

$$\min\left\{1, \lambda \cdot \frac{p(l+m+1) \cdot p(\boldsymbol{\delta}_{l+1})}{p(l+m)} \cdot \frac{\dfrac{p_d}{l+m+1}}{p_a \cdot p(\boldsymbol{\delta}_{l+1}|\boldsymbol{y}, \phi, \boldsymbol{Q}_{l+1})}\right\}, \qquad (2)$$

where $\varphi = (l, m, \theta, \boldsymbol{\delta}, \boldsymbol{\gamma}, \boldsymbol{Q}, \boldsymbol{W}, \sigma_u^2, \sigma_e^2)$, and the likelihood ratio λ is as follows;

$$\lambda = \frac{\prod\limits_{k=1}^{n} \dfrac{1}{\sqrt{2\pi\sigma_e^2}} \cdot \exp\left\{-\dfrac{\left(\boldsymbol{y} - \boldsymbol{X}\beta - \sum\limits_{i=1}^{l} \boldsymbol{Q}_i\boldsymbol{\delta}_i - \sum\limits_{j=1}^{m} \boldsymbol{W}_j\boldsymbol{\gamma}_j - \boldsymbol{Q}_{l+1}\boldsymbol{\delta}_{l+1} - \boldsymbol{Z}\boldsymbol{u}\right)^2}{2\sigma_e^2}\right\}}{\prod\limits_{k=1}^{n} \dfrac{1}{\sqrt{2\pi\sigma_e^2}} \cdot \exp\left\{-\dfrac{\left(\boldsymbol{y} - \boldsymbol{X}\beta - \sum\limits_{i=1}^{l} \boldsymbol{Q}_i\boldsymbol{\delta}_i - \sum\limits_{j=1}^{m} \boldsymbol{W}_j\boldsymbol{\gamma}_j - \boldsymbol{Z}\boldsymbol{u}\right)^2}{2\sigma_e^2}\right\}}.$$

When the new proposal is 2, the proposal is accepted with the probability

$$\min\left\{1, \lambda \cdot \frac{p(l+m+1) \cdot p(\boldsymbol{\gamma}_{m+1})}{p(l+m)} \cdot \frac{\dfrac{p_d}{l+m+1}}{p_a \cdot p(\boldsymbol{\gamma}_{m+1}|\boldsymbol{y}, \phi, \boldsymbol{W}_{m+1})}\right\}. \qquad (3)$$

The updating step follows that of proposal 1, except that new locations and effects are sampled for two QTL.

When the proposal is 3, only the other parameters in the model are updated. When a proposal is 4, one of the QTL in the model is excluded with equal probability. The proposal is accepted with probability

$$\min\left\{1, \lambda \cdot \frac{p(l+m-1)}{p(l+m) \cdot p(\boldsymbol{\delta}^*)} \cdot \frac{p_a \cdot p(\delta^*|\boldsymbol{y}, \phi^*)}{\dfrac{p_d}{l+m}}\right\}, \qquad (4)$$

where

$$
\lambda = \frac{\displaystyle\prod_{k=1}^{n} \frac{1}{\sqrt{2\pi\sigma_e^2}} \cdot \exp\left\{-\frac{\left(\boldsymbol{y} - \boldsymbol{X\beta} - \sum_{i=1}^{l} \boldsymbol{Q}_i\boldsymbol{\delta}_i - \sum_{j'\neq j}^{m} \boldsymbol{W}_{j'}\boldsymbol{\gamma}_{j'} - \boldsymbol{Zu}\right)^2}{2\sigma_e^2}\right\}}{\displaystyle\prod_{k=1}^{n} \frac{1}{\sqrt{2\pi\sigma_e^2}} \cdot \exp\left\{-\frac{\left(\boldsymbol{y} - \boldsymbol{X\beta} - \sum_{i=1}^{l} \boldsymbol{Q}_i\boldsymbol{\delta}_i - \sum_{j=1}^{m} \boldsymbol{W}_j\boldsymbol{\gamma}_j - \boldsymbol{Zu}\right)^2}{2\sigma_e^2}\right\}},
$$

and $\varphi*$ denotes φ without $\boldsymbol{\delta}_j$, i.e., the effect of the deleted QTL.

For proposal 5, one of the pairs of QTL in the model is excluded with equal probability. The proposal is accepted with a probability

$$
\min\left\{1, \lambda \cdot \frac{p\,(l+m-1)}{p\,(l+m) \cdot p\,(\boldsymbol{\gamma}*)} \cdot \frac{p_a \cdot p\,(\boldsymbol{\gamma} * |\boldsymbol{y}, \phi*)}{\dfrac{p_d}{l+m}}\right\}. \tag{5}
$$

The complete sampling outline is as follows:

(a) updating μ using the Gibbs sampler;
(b) updating l or m using the RJMCMC algorithm;
(c) updating θ using the MH algorithm;
(d) updating $\boldsymbol{\delta}$ and $\boldsymbol{\gamma}$ using the Gibbs sampler;
(e) updating $\boldsymbol{u}$ using the Gibbs sampler;
(f) updating σ_u^2 using the Gibbs sampler;
(g) updating σ_e^2 using the MH algorithm.

2.3. Simulation study

As mentioned above, an F_2 population generated from upward- and downward-selected breeds was assumed. Ten sires and 400 dams were picked up from each breed, noted by breed 1 and 2, respectively. Each dam has one progeny when they are randomly mated, and that the 400 progeny, noted by F_1, are male or female with a probability of 0.5. F_2 individuals were generated by intercrossing the F_1 individuals randomly. At that time, each dam has three progeny, therefore the number of individuals in the F_2 population is expected to be 600, but in most cases, the number of females in the F_1 generation is not
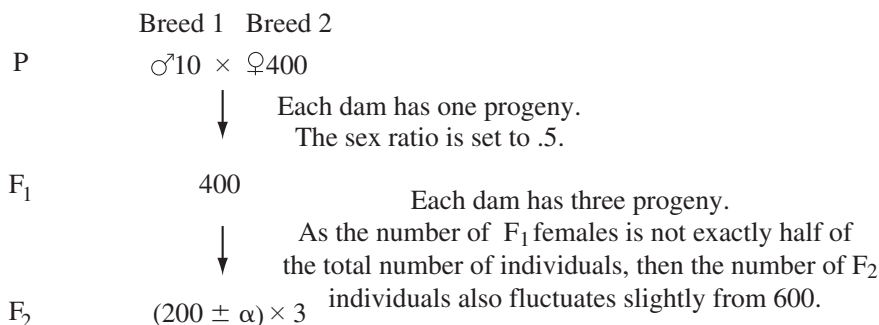
Breed 1 Breed 2

P $\male 10 \times \female 400$

Each dam has one progeny.
The sex ratio is set to .5.

F_1 400

Each dam has three progeny.
As the number of F_1 females is not exactly half of
the total number of individuals, then the number of F_2
individuals also fluctuates slightly from 600.

F_2 $(200 \pm \alpha) \times 3$

Figure 1. Diagram of the structure of a simulated crossbred population.

exactly half of the total number of individuals, then the number of the individuals in the F_2 generation fluctuates slightly from 600. Details are shown in Figure 1.

The number of alleles in each breed was set to three, and two of them were common in both breeds. Each allele frequency was as follows: In breed 1, the frequencies of alleles 1, 2, and 3 were 0.8, 0.1, and 0.1, respectively. In breed 2, the frequencies of alleles 1, 2, and 4 were 0.1, 0.8, and 0.1, respectively.

In this study, we implemented simulation studies for four cases. In case 1, no QTL were set and only a polygenic effect exists. In case 2, one QTL with only a main effect and one pair of QTL with both main and epistatic effects between each other were present. In case 3, one QTL with only a main effect and two pairs of QTL with an epistatic effect were set, and for one of the two pairs, neither QTL has the main effect. In case 4, which is the most complicated situation, the same number of QTL were located at the same position as in case 3. One QTL has epistatic effects with the other two, and none of the three QTL has the main effect. The other two QTL have the respective main effects. The overall mean of the phenotypic values was set to 100, and it was assumed that there were two chromosomes, each of which were 100 cM long and both had eleven marker loci, one every 10 cM. The polygenic and environmental variances were 25 and 75, respectively. These situations and values were common in all cases. The locations and effects of QTL are presented in Table II. Phenotypic values were available only for the F_2 population and the pedigree information was used for the three generations.

The chain was run 50 000 rounds totally, and in order to eliminate the effect of a serial correlation, the chain was thinned and saved one sample per 10 cycles, so 5000 samples were finally saved per replicate. In total, 25 replicates were carried out for the respective cases, where the means, medians and modes

Table II. The true values of the parameters set in respective cases.

Parameter[a]	True value				
case 1					
l/m	0/0				
case 2					
l/m		1/1			
	QTL1	QTL2		QTL3	
θ	[1, 37.0][b]	[1, 88.0]		[2, 83.0]	
a	4.0	4.0		4.0	
d	3.0	3.0		3.0	
aa		7.0			
ad		5.0			
da		0.0			
dd		0.0			
case 3					
l/m		1/2			
	QTL1	QTL2	QTL3	QTL4	QTL5
θ	[2, 52.0]	[1, 37.0]	[2, 18.0]	[1, 88.0]	[2, 83.0]
a	5.0	4.0	0.0	0.0	0.0
d	2.0	3.0	0.0	0.0	0.0
aa		7.0		7.0	
ad		0.0		5.0	
da		0.0		0.0	
dd		0.0		0.0	
case 4					
l/m		2/2			
	QTL1	QTL2	QTL3	QTL4	QTL5
θ	[1, 37.0]	[2, 52.0]	[1, 88.0]	[2, 83.0]	[2, 18.0]
a	4.0	5.0	0.0	0.0	0.0
d	3.0	2.0	0.0	0.0	0.0
aa				7.0	7.0
ad				5.0	0.0
da				0.0	0.0
dd				0.0	0.0

[a] See Table I.
[b] The former figure represents the chromosome number, and the latter represents the position (cM) of the QTL.

of posterior distributions were recorded. The prior distributions for μ was set to the normal distribution N (100, 8^2), and for δ and γ, the prior was set to N (0, 8^2). The mean of the prior Poisson distribution of the number of QTL and QTL pair number μ_{l+m} was 3, and in case 3, to check the influence of μ_{l+m} on the estimates, two additional different prior means (μ_{l+m} = 2 and 5) were used. For the updating of θ and σ_e^2 via the MH algorithm, the tuning parameter d was set to 2.5. For σ_e^2, the range from which the values were sampled was from 0 to the phenotypic value. The upper limits of QTL and QTL pair numbers, noted by l_{max} and m_{max}, were set to 4 and 3, respectively. The calculations started with the null model ($l^{[1]} = m^{[1]} = 0$), but for all the other parameters, the initial values were decided arbitrarily. In case 3, the different initial values, $l^{[1]} = l_{max}$ and $m^{[1]} = m_{max}$, were also used to investigate the influence of this factor.

3. RESULTS

Table III represents the posterior distributions of the QTL number in each case. Table IV represents the modes of the posterior distributions of all the other parameters in these cases. Since both the mean and the median were hardly different from the mode, the former are not shown here.

In case 1, neither the QTL with the main effect nor pairs of epistatic QTL were detected, which is consistent with true situations. The result indicates that the method can avoid type II errors in the absence of QTL and QTL pairs. In case 2, QTL pairs with both strong main and epistatic effects, which are relatively easy to detect, were mapped accurately. In cases 3 and 4, it is shown that the probabilities that the same number of QTL and QTL pairs as the true ones existed were extremely high (from 0.832 to 0.902), i.e., that the estimated QTL number corresponded to the true one. Especially, it should be noted that the proposed method succeeded in detecting epistatic QTL pairs without main effects. To help the understanding of the readers, Figure 2 shows histograms of the posterior distributions for locations of epistatic QTL in one of the replicates of cases 3 and 4.

Additionally, even if as in case 4, QTL that have epistatic effects with two QTL at the same time, or maybe more, were also detectable. For the other parameters, though only the environmental variance component was considerably overestimated, it is clear that the overall mean, the positions and effects of QTL, and polygenic variance were also accurately estimated. For comparison, one of the simulated data set in case 3 was analyzed using conventional interval mapping. Figure 3 shows that the QTL with the main effect, such as

A. Narita, Y. Sasaki

Table III. The posterior distributions of the QTL (l) and QTL pair number (m) in each case.

l	m 0	1	2	3
case 1				
0	**0.851** (0.080)[a]	0.000 (0.001)	0.000 (0.000)	0.000 (0.000)
1	0.143 (0.076)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
2	0.006 (0.005)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
3	0.000 (0.001)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
4	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
case 2				
0	0.000 (0.000)	0.001 (0.004)	0.006 (0.017)	0.000 (0.000)
1	0.000 (0.000)	**0.901** (0.034)	0.003 (0.010)	0.000 (0.000)
2	0.000 (0.000)	0.082 (0.027)	0.000 (0.001)	0.000 (0.000)
3	0.001 (0.004)	0.004 (0.003)	0.000 (0.000)	0.000 (0.000)
4	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
case 3 ($\mu_{l+m} = 2$)[b]				
0	0.000 (0.000)	0.000 (0.001)	0.013 (0.030)	0.000 (0.001)
1	0.000 (0.000)	0.002 (0.003)	**0.902** (0.036)	0.002 (0.006)
2	0.000 (0.000)	0.004 (0.007)	0.074 (0.029)	0.000 (0.000)
3	0.000 (0.000)	0.001 (0.002)	0.002 (0.002)	0.000 (0.000)
4	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
case 3 ($\mu_{l+m} = 3$)				
0	0.000 (0.000)	0.001 (0.001)	0.008 (0.014)	0.000 (0.000)
1	0.000 (0.000)	0.003 (0.003)	**0.902** (0.033)	0.001 (0.008)
2	0.000 (0.000)	0.001 (0.002)	0.081 (0.031)	0.000 (0.000)
3	0.000 (0.000)	0.000 (0.000)	0.003 (0.004)	0.000 (0.000)
4	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
case 3 ($\mu_{l+m} = 5$)				
0	0.000 (0.000)	0.000 (0.001)	0.005 (0.011)	0.000 (0.000)
1	0.000 (0.000)	0.002 (0.004)	**0.832** (0.065)	0.004 (0.014)
2	0.000 (0.000)	0.003 (0.005)	0.139 (0.058)	0.000 (0.001)
3	0.000 (0.000)	0.001 (0.001)	0.013 (0.010)	0.000 (0.000)
4	0.000 (0.000)	0.000 (0.000)	0.000 (0.001)	0.000 (0.000)
case 4				
0	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.005 (0.024)
1	0.000 (0.000)	0.001 (0.003)	0.012 (0.022)	0.007 (0.022)
2	0.001 (0.002)	0.016 (0.024)	**0.888** (0.053)	0.001 (0.003)
3	0.001 (0.002)	0.002 (0.003)	0.064 (0.036)	0.000 (0.000)
4	0.000 (0.000)	0.000 (0.000)	0.002 (0.002)	0.000 (0.000)

[a] The figure is the average of 25 replicates.
[b] The figure in parenthesis represents the standard deviation.
[c] μ_{l+m} is the prior mean for l and m.

Table IV. The posterior distributions of the overall mean, the locations and effects of QTL, polygenic and environmental variances.

Parameter[a]	Mode of the posterior distribution				
case 1					
$\mu/\sigma_u^2/\sigma_e^2$	100.18[b] (0.89)[c]		25.57 (9.84)		73.95 (7.54)
case 2					
$\mu/\sigma_u^2/\sigma_e^2$	100.36 (1.09)		20.57 (15.94)		83.54 (11.81)
	QTL1		QTL2		QTL3
θ	[1, 36.72][d] (4.33)		[1, 88.72] (2.05)		[2, 82.64] (2.38)
a	4.04 (0.92)		3.84 (1.13)		3.68 (1.19)
d	3.05 (1.14)		2.65 (1.21)		3.04 (1.38)
aa			6.87(0.92)		
ad			5.17 (1.55)		
da			0.53 (1.85)		
dd			0.02 (1.85)		
case 3 ($\mu_{l+m}=2$)					
$\mu/\sigma_u^2/\sigma_e^2$	100.16 (1.30)		25.59 (9.75)		81.37 (7.88)
	QTL1	QTL2	QTL3	QTL4	QTL5
θ	[2, 51.51](2.55)	[1, 37.86] (2.89)	[2, 19.07](3.68)	[1, 88.85](2.69)	[2, 81.04](3.17)
a	5.05 (0.70)	3.87 (1.14)	0.26 (1.12)	0.03 (0.96)	0.14 (0.95)
d	2.01 (0.92)	2.65 (1.40)	−0.07 (1.36)	−0.16 (1.17)	0.23 (1.46)
aa		7.12 (0.74)		6.70 (0.79)	
ad		0.45 (1.86)		4.72 (1.43)	
da		−0.38 (1.50)		−0.05 (1.14)	
dd		0.23 (2.16)		−0.30 (1.94)	
case 3 ($\mu_{l+m}=3$)					
$\mu/\sigma_u^2/\sigma_e^2$	100.18 (1.29)		24.60 (9.76)		81.98 (8.06)
	QTL1	QTL2	QTL3	QTL4	QTL5
θ	[2, 51.28](2.60)	[1, 37.95](2.85)	[2, 18.81](3.50)	[1, 88.85] (2.73)	[2, 80.94](3.29)
a	5.06 (0.74)	3.86 (1.26)	0.20 (1.10)	0.06 (0.96)	0.16 (0.90)
d	1.96 (0.88)	2.68 (1.28)	−0.06 (1.35)	−0.11 (1.12)	0.31 (1.50)
aa		7.11 (0.83)		6.67 (0.84)	
ad		0.38 (1.80)		4.68 (1.46)	
da		−0.45 (1.62)		0.00 (1.09)	
dd		0.23 (2.11)		−0.19 (1.97)	
case 3 ($\mu_{l+m}=5$)					
$\mu/\sigma_u^2/\sigma_e^2$	100.16 (1.33)		25.05 (9.12)		81.04 (7.16)
	QTL1	QTL2	QTL3	QTL4	QTL5
θ	[2, 51.42](2.67)	[1, 37.86](2.86)	[2, 18.90](3.44)	[1, 88.81](2.69)	[2, 80.99] (3.22)
a	4.97 (0.75)	3.92 (1.14)	0.30 (1.14)	0.03 (0.97)	0.10 (0.98)
d	2.00 (0.87)	2.66 (1.37)	−0.06 (1.33)	−0.05 (1.13)	0.34 (1.53)
aa		7.08 (0.79)		6.66 (0.80)	
ad		0.38 (1.77)		4.77 (1.48)	
da		−0.35 (1.59)		−0.06 (1.14)	
dd		0.33 (2.17)		−0.30 (1.95)	
case 4					
$\mu/\sigma_u^2/\sigma_e^2$	99.66 (1.65)		25.60 (12.05)		79.27 (11.34)
	QTL1	QTL2	QTL3	QTL4	QTL5
θ	[1, 37.44](4.43)	[2, 51.14](3.02)	[1, 88.64](2.45)	[2, 83.25](3.31)	[2, 17.59] (4.50)
a	3.94 (0.59)	5.12 (0.74)	−0.14 (0.98)	0.11 (1.34)	−0.09 (1.22)
d	2.87 (0.95)	2.24 (1.01)	−0.08 (1.23)	0.01 (2.30)	−0.18 (1.80)
aa			6.75 (0.72)	7.32 (1.07)	
ad			5.22 (1.58)	−0.3l(1.49)	
da			0.31 (1.35)	0.11 (1.69)	
dd			−0.05 (1.59)	0.25 (2.44)	

[a] See Table I.

[b] The figure is the average of 25 replicates.

[c] The figure in parenthesis represents the standard deviation.

[d] The figures in brackets represent the chromosome number and the estimated position (cM) of the QTL.

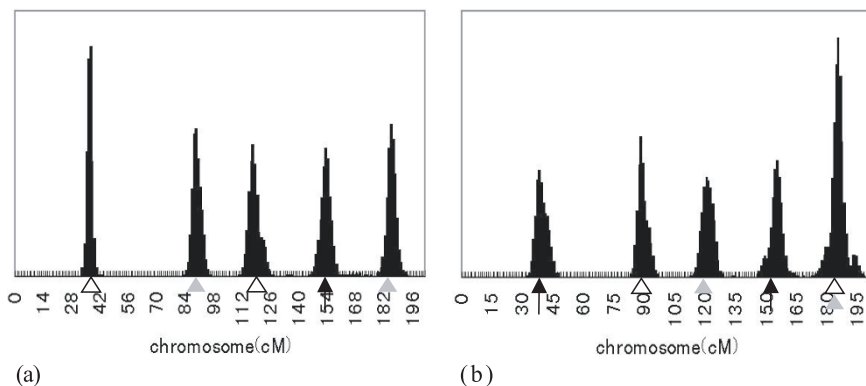(a)   (b)

Figure 2. Histograms of the posterior distributions for locations of epistatic QTL. The results in one of the replicates of case 3 (a) and case 4 (b) are shown, respectively. Two triangles of the same color indicate the true positions that are epistatic, and the arrows indicate the positions of QTL with only main effects. For convenience of explanation, the two chromosomes are presented as if they were connected to each other. The segment from the point of 100 cM to the right end represents chromosome 2.
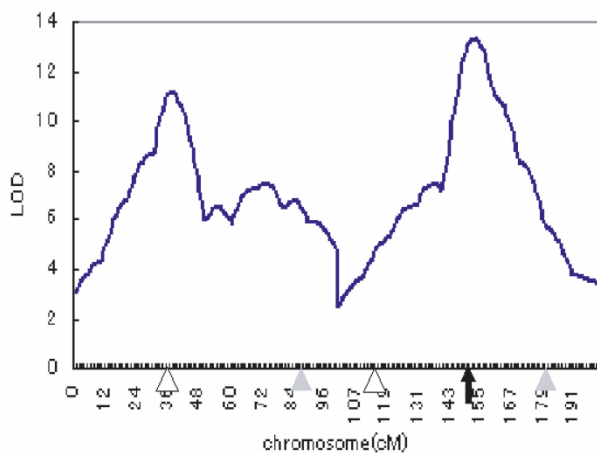
Figure 3. LOD scores plotted against the whole genome using conventional interval mapping. The simulated data set used was the same as in Figure 2 (a). Two triangles of the same color indicate the true positions that are epistatic, and the arrows indicate the positions of QTL with only main effects. For convenience of explanation, the two chromosomes are presented as if they were connected to each other. The segment from the point of 100 cM to the right end represents chromosome 2.

(a)   (b)
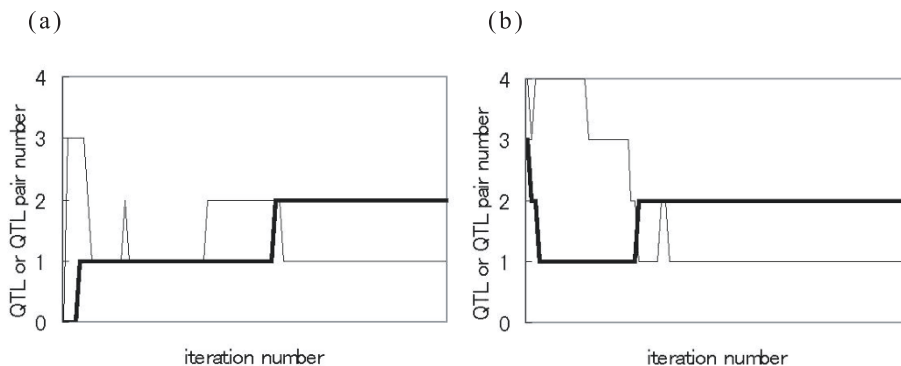
Figure 4. The traces of samples for QTL number (l) and QTL pair number (m) against the iteration number. (a) and (b) show the results when $l^{[1]} = m^{[1]} = 0$ and $l^{[1]} = l_{\max}$ and $m^{[1]} = m_{\max}$, respectively. The thin line and the bold line indicate QTL number and QTL pair number, respectively.

QTL1 and QTL2, were detectable under the single-QTL model, but epistatic QTL without main effects could not be detected.

In case 3, it is also shown that since μ_{l+m} was larger, the posterior distributions of l and m tended to shift to slightly larger values but the influence was not so strong as to alter their modes. For all the other parameters, quite similar results were yielded regardless of prior means. This indicates that the method is robust, if not perfect, to prior means of QTL number.

We also checked the influence of starting values, including the mixing behavior of the RJMCMC algorithm. The results are shown in Figure 4. Whether starting with $l^{[1]} = m^{[1]} = 0$ or $l^{[1]} = l_{\max}$ and $m^{[1]} = m_{\max}$, though models which were quite different from the true one were sampled in the burn-in period, along iterations of the sampler, l and m gradually got closer to the true values, and finally the correct model was reached. The appropriate mixing behavior of the RJMCMC algorithm was observed for l, whereas for m, the sampler did not frequently move to the other models. However, the extent was not so serious, and this is not surprising because ten parameters were involved for a pair of QTL, whereas for a single QTL only three were involved. This problem can be overcome by sufficient iterations of sampling and by lengthening the interval at which the samples are stored.

4. DISCUSSION

Generally, it may be more likely that a quantitative trait is controlled by multiple QTL with different modes of inheritance and effects of various sizes,

which may have some epistatic effects on one another. To ascertain the complicated mechanism of quantitative traits, conventional methods considering only one QTL at a time, such as interval mapping [13] and composite interval mapping [11, 29], have seemed to be inadequate and these may increase the risk of both type I and type II errors. Therefore, it is necessary to develop a more sophisticated method, in which the multiple QTL are simultaneously taken into account. Since Bayesian approaches via MCMC algorithms have been generally used as an alternative to maximum-likelihood and least squares methods, several methods to detect multiple QTL have been developed in an F_2 population derived from two inbred lines such as mice and Drosophila [18], in a crossbred population such as swine [19], in a granddaughter design such as dairy cattle [22], and in a population with a more complex pedigree [9].

These methods are all based on the assumption that each QTL contribute independently to a quantitative trait. The estimation of the number of QTL itself has been emphasized too much, while epistatic effects among them have hardly been explored. Yi and Xu [27] developed an epoch-making method to map multiple epistatic QTL using the Bayesian inference *via* MCMC algorithms.

However, the authors stated that their method may not be adequate in cases where neither of two QTL has the main effect but they have the epistatic effect. Traits on which neither of two epistatic QTL has the main effect have been discussed in several articles, and epistatic loci that do not have the main effect do not seem to be negligible components in quantitative traits. van Wezel [23] showed that mice homozygous for the STS allele at the D18Mit57 marker and having at least one BALB/c allele at D17Mit72, and mice homozygous for the STS allele at the D17Mit72 marker and homozygous for BALB/c allele at the D18Mit57 marker were more susceptible to colon tumors, and stated that the two loci had the strong reciprocal interaction, but neither had the significant main effect. Li *et al*. [14] reported that, on three grain yield components in rice, the majority of epistatic markers did not appear to have individual effects. Then, by treating a pair of epistatic QTL as one variable in the RJMCMC algorithm, we have made it possible for two QTL to be added to or deleted from the model simultaneously. As a result, not only pairs of QTL that have both the main and the epistatic effects but also those that have no main effects but have significant epistatic effects are also detectable.

Moreover, the method proposed by Yi and Xu is designed for an inbred population, which can be obtained easily in crops and experimental animals. Because of a long generation interval, the risk of inbreeding depression, and great cost for feeding and management, it is very difficult to produce an

inbred line by repeating consanguine mating in livestock. This is the reason why the method cannot be directly applied to complicated data in livestock, therefore, additional careful considerations and appropriate algorithms are required. Then, as is often the case with livestock such as cattle and swine, the model used assumes that putative QTL are diallelic with alternative alleles fixed in two parental breeds but alleles at marker loci are not completely fixed. We have also incorporated the deterministic approach to calculate the probabilities of QTL genotypes [7], which has little effect on the accuracy of estimation and save us considerable computing time [16, 27].

Additionally, by setting the mixed inheritance model in which a random genetic effect, along with a multiple QTL with a moderate or large effect, is included, the method is able to take into account QTL that are segregating within each breed and a large number of polygenes with minute and individually undetectable effects.

In this study, an integrated system for detecting multiple epistatic QTL in an outbred population under the mixed inheritance model has been established. Recently, to overcome the drawback of the previous method [27], Yi *et al*. [28] have developed a more elaborate and robust method which can add or delete QTL two by two, and unlike our method, two QTL added into the model at the same time are not necessarily deleted together. In addition to removing insignificant effects of QTL (whether main or epistatic) from the model and to retaining the appropriate mixing behavior, effect indicators, which take a value one if QTL have significant main effects and zero otherwise, were used as additional parameters in the method. However, their method is principally designed for populations with only two genotypes, *e.g.*, a backcross population, double haploid lines and recombinant inbred lines, and of course does not take a polygenic effect into account. They also stated that for an F_2 population in which higher-order epistatic effects are involved, the number of parameters and computational load will largely increase. In our method, epistatic QTL without significant main effects can be detected in a relatively simple way in an F_2 population, without any additional variables.

Of course, by some brief modifications, our method can be extended for different types of design, such as the half-sib population. The method is also robust to prior distributions and starting values, and the RJMCMC algorithm mixes reasonably. The mixing behavior of the RJMCMC algorithm depends on the way in which the effects of QTL added newly to the model are generated [27]. Especially, in the case of a large number of unknown parameters, sampling values for new parameters from uniform distributions, without any regards for the current values of the other parameters, resulted in a seriously

low acceptance rate, *i.e.*, "sticky" (data not shown). By conditioning the probability density function for a parameter on the current values of the other parameters and the phenotypic values, the mixing behavior is improved and appropriate convergence to the posterior distributions can be attained. As mentioned above, the results obtained by simulation experiments were almost consistent with the true values, and it gave researchers evidence that the system proposed in this study will work well in real cases.

REFERENCES

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

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

[3] Cockerham C.C., An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present, Genetics 39 (1954) 859–882.

[4] Geman S., Geman D., Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, IEEE Trans. Pattn. Anal. Mach. Intell. 6 (1984) 721–741.

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

[6] Haley C.S., Knott S.A., A simple regression method for mapping quantitative trait loci in line crosses using flanking markers, Heredity 69 (1992) 315–324.

[7] Haley C.S., Knott S.A., Elsen J.M., Mapping quantitative trait loci in crosses between outbred lines using least squares, Genetics 136 (1994) 1195–1207.

[8] Hastings W.K., Monte Carlo sampling methods using Markov chains and their applications, Biometrika 57 (1970) 97–109.

[9] Heath S.C., Markov chain Monte Carlo segregation and linkage analysis for oligogenic models, Am. J. Hum. Genet. 61 (1997) 748–760.

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

[11] Jansen R.C., Interval mapping of multiple quantitative trait loci, Genetics 135 (1993) 205–211.

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

[13] Lander E.S., Botstein D., Mapping mendelian factors underlying quantitative traits using RFLP linkage maps, Genetics 121 (1989) 185–199.

[14] Li Z., Pinson S.R., Park W.D., Paterson A.H., Stansel J.W., Epistasis for three grain yield components in rice (*Oryza sativa* L.), Genetics 145 (1997) 453–465.

[15] Metropolis N., Rosenbluth A.W., Rosenbluth M.N., Teller A.H., Teller E., Equation of state calculations by fast computing machines, J. Chem. Phys. 21 (1953) 1087–1091.

[16] Pong-Wong R., George A.W., Woolliams J.A., Haley C.S., A simple and rapid method for calculating identity-by-descent matrices using multiple markers, Genet. Sel. Evol. 33 (2001) 453–471.

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

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

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

[20] Thaller G., Hoeschele I., A Monte Carlo method for Bayesian analysis of linkage between single markers and quantitative trait loci: I. Methodology, Theor. Appl. Genet. 93 (1996) 1161–1166.

[21] Thaller G., Hoeschele I., A Monte Carlo method for Bayesian analysis of linkage between single markers and quantitative trait loci: II. A stimulation study, Theor. Appl. Genet. 93 (1996) 1167–1174.

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

[23] van Wezel T., Stassen A.P., Moen C.J., Hart A.A., van del Valk M.A., Demant P., Gene interaction and single gene effects in colon tumour susceptibility in mice, Nat. Genet. 14 (1996) 468–470.

[24] Visscher P.M., Thompson R., Haley C.S., Confidence intervals in QTL mapping by bootstrapping, Genetics 143 (1996) 1013–1020.

[25] Wang D.L., Zhu J., Li Z.K., Paterson A.H., Mapping QTLs with epistatic effects and QTL×environment interactions by mixed linear model approaches, Theor. Appl. Genet. 99 (1999) 1255–1264.

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

[27] Yi N., Xu S., Mapping quantitative trait loci with epistatic effects, Genet. Res. 79 (2002) 185–198.

[28] Yi N., Xu S., Allison D.B., Bayesian model choice and search strategies for mapping interacting quantitative trait loci, Genetics 165 (2003) 867–883.

[29] Zeng Z.B., Precision mapping of quantitative trait loci, Genetics 136 (1994) 1457–1468.

[30] Zeng Z.B., Kao C.H., Basten C.J., Estimating the genetic architecture of quantitative traits, Genet. Res. 74 (1999) 279–289.