

# Bayesian mapping of QTL in outbred F2 families allowing inference about whether F0 grandparents are homozygous or heterozygous at QTL

T Hayashi and T Awata

Laboratory of Animal Genome, National Institute of Agrobiological Sciences, 2 Ikenodai, Tsukuba-shi, Ibaraki 305-0901, Japan

In this paper, we propose a new Bayesian method for QTL analysis in outbred F2 families based on Markov chain Monte Carlo (MCMC) estimation allowing inference about whether each of F0 founders (grandparents) is homozygous or heterozygous at QTL. This, in turn, allows us to select a model accurately explaining observations of phenotypes for F2 individuals. The proposed method performs the fitting a statistical model of the two possible QTL states in each F0 grandparent, that is, homozygous and heterozygous at QTL, and gives a posterior distribution for the QTL states in each F0 grandparent. We confine ourselves to the discrimination of two QTL states, homozygous or heterozygous, for each of the F0 grandparents without taking into consideration

whether common alleles are shared by F0 grandparents. The statistical model includes allelic effects and dominance effects for each QTL. The number of parameters representing allelic effects and dominance effects is therefore changed depending on the QTL states. A Reversible Jump MCMC technique is used for transition between the models of different dimensions. The effectiveness of the proposed method was investigated using simulation experiments. It was practicable to estimate the QTL states of F0 grandparents as well as the number, the locations and the effects of QTL segregating in an outbred F2 family.

*Heredity* (2005) **94**, 326–337. doi:10.1038/sj.hdy.6800638  
Published online 26 January 2005

**Keywords:** Bayesian method; MCMC; QTL analysis; outbred population; F2 family

## Introduction

In inbred organisms such as rice and laboratory mice, F2 families derived from crossing two lines are suitable experimental populations for mapping QTL expressed differently between two lines. F2 families have also been constructed for detecting QTL in outbred organisms, such as livestock and trees, starting with crosses between individuals sampled from each of two genetically diverged populations.

Interval mapping methods (Lander and Botstein, 1989; Haley *et al.*, 1994) have been successfully applied to QTL analyses of F2 families in outbred organisms as well as inbred organisms. In pigs, many analyses with interval mapping have been performed for F2 families derived from a cross between two breeds and successful detection of several QTL affecting economically important traits have been reported (eg Andersson *et al.*, 1994; Walling *et al.*, 2000; Bidanel *et al.*, 2001). In fruit trees or forest trees, F2 populations originated from a cross between two selected individuals have also been utilized for interval mapping and QTL of useful traits such as fruit quality or adaptive properties were detected (Dirlewanger *et al.*, 1999; Frewen *et al.*, 2000).

The model of interval mapping analysis usually applied to outbred F2 families assumes that founder

individuals (grandfathers and grandmothers in the F0 generation) of the families were fixed for different alleles at QTL. However, a high degree of heterozygosity is usually observed within a population of outbred organisms and many QTL are considered to be segregating in a population. There is reduced efficiency in detecting QTL at which the F0 grandparents are heterozygous. These particular QTL may, therefore, remain undetected.

Pérez-Enciso and Varona (2000) proposed a method allowing the detection of differences between means of QTL effects in two outbred populations as well as the variation of QTL effects within each of populations. Thus, not only QTL segregating between two outbred populations but also QTL segregating within each of populations could be detected by their method. However, since estimation of QTL variance within each of founder populations is required in their mixed model, their method is not appropriate for F2 families derived from a small number of F0 founders (particularly designs with only two founders).

Recently, Bayesian methods based on the Markov chain Monte Carlo (MCMC) algorithm have been developed for QTL analyses in outbred pedigrees as well as inbred ones (Satagopan *et al.*, 1996; Uimari and Hoeschele, 1997; Sillanpää and Arjas, 1998, 1999; Bink and Van Arendonk, 1999; Uimari and Sillanpää, 2001; Yi and Xu, 2001, 2002; Bink *et al.*, 2002; Jannink and Wu, 2003). Utilizing reversible jump MCMC (RJ-MCMC) technique (Green, 1995), it becomes possible to estimate the number of QTL involved in a quantitative trait as well as the locations and effects of QTL. Both homozygous QTL and heterozygous QTL in founder individuals can be detected. For outbred F2 families, Sillanpää

Correspondence: T Hayashi, Laboratory of Animal Genome, National Institute of Agrobiological Sciences, 2 Ikenodai, Tsukuba-shi, Ibaraki 305-0901, Japan. E-mail: hayatk@affrc.go.jp

Received 8 December 2003; accepted 1 October 2004; published online 26 January 2005

and Arjas (1999) devised a Bayesian estimation of QTL using an RJ-MCMC algorithm. Their method assumed that the number of possible QTL genotypes segregating in an F2 family was known and incorporated the parameters of effects for each QTL genotypes in the model. It is, however, usually difficult to infer the number of QTL genotypes in advance of analysis. Another Bayesian method applicable for outbred F2 population was developed by Yi and Xu (2001). Their model assumed that all of the alleles that the F0 grandparents carried were different from each other, and parameterized allelic effects for all of the founder alleles and interaction (dominance) effects between two alleles. Accordingly, the number of QTL genotypes was predetermined and a large number of parameters were included in their model. The pedigree-analysis method proposed by Uimari and Sillanpää (2001) and Bink *et al* (2002) is also applicable to the analysis of outbred F2 families. In their method, however, only the case of biallelic QTL was treated.

When some of F0 grandparents are homozygous at QTL, a model with fewer parameters is more appropriate for explaining the observed F2 phenotypes. A more suitable model would take into account whether each of the F0 grandparents was homozygous at each QTL. The inference about QTL genotype for each of the F0 founders would also provide useful information for breeding programs based on these individuals.

In this paper, we propose a new Bayesian method implementing these innovations, which gives a posterior distribution for the QTL states in each F0 grandparent. We confine ourselves to the discrimination of two QTL states, homozygous or heterozygous, for each of the F0 grandparents. Accordingly, the analysis does not take into account whether common alleles are shared by the F0 grandparents. We parameterize allelic effects and dominance effects for each QTL. The total number of parameters states, thus, RJ-MCMC technique therefore depends on the F0 QTL genotypes. Consequently, RJ-MCMC is used for transition between the models of different dimensions.

Some Bayesian methods incorporate the estimation of marker haplotypes into the MCMC procedure (Sillanpää and Arjas, 1999; Bink and Van Arendonk, 1999; Yi and Xu, 2001). However, the step is very time-consuming and marker haplotypes remain ambiguous in the case of low marker informativeness in founder individuals. In the proposed method, estimation of the marker haplotypes is, therefore, carried out separately from the MCMC procedure. The QTL genotypes for individuals were sampled from the conditional distributions, given the estimated haplotypes.

A Bayesian method allowing the estimation of the number of alleles at QTL present in the experimental families was recently developed by Jannink and Wu (2003). Their method was applied to the analyses of multiple families derived from crosses among inbred lines. The number of alleles segregating among inbred lines was inferred assuming that all of inbred lines were homozygous at QTL. Their model is not, therefore, applicable to the design we consider here: outbred F2 families, in which F0 grandparents may have been heterozygous at QTL. In the following, we focus mainly on a three-generation family that consists of F2 individuals, F1 parents and two outbred F0 grandparents. We

assume that phenotypic values of a trait for all of F2 individuals are available and, in addition, marker genotypes for all of individuals in the family.

## Materials and methods

### Analyzed family

We assume that a male and a female chosen from different populations (breeds or lines) have been crossed to produce  $m$  F1 progeny and  $n$  F2 progeny. Such data are often available for linkage and QTL study in forest trees (Bradshaw Jr and Stettler, 1995; Wu, 1998; Dirlewanger *et al*, 1999; Costa *et al*, 2000; Bliss *et al*, 2002). In the QTL analyses of pigs, a three-generation family is often constructed, however, multiple males and females are used as founders. The proposed method can be extended to the F2 families from more than two F0 grandparents.

### Statistical model

We consider the ordered genotypes at QTL for each individual by discriminating its paternal allele and maternal allele even when an individual is homozygous. The QTL genotypes of F1 and F2 individuals are indicated in terms of the founder origin of alleles F1 and F2 individuals carry. There are, thus, four origins of alleles, denoted by 1, 2, 3 and 4 corresponding to each of four alleles of two F0 grandparents. The origins 1 and 2 indicate paternal allele and maternal allele of a grandfather and 3 and 4 indicate paternal allele and maternal allele of a grandmother, respectively. When no ancestral information for the F0 grandparents is available, which is typical, origins (1,2) and (3,4) are arbitrarily allocated, one to each grandparent.

We assume that the number of QTL affecting the trait is  $N_Q$ . For the  $q$ th QTL allele of origin  $j$  ( $j = 1, 2, 3, 4$ ) in the F2 population, the allelic effect is denoted by  $\alpha_{qj}$ , and the interaction effect (dominance effect) between a pair of alleles of origin  $j$  and  $k$  by  $\delta_{qjk}$  ( $j < k$ ). For identifiability of allelic effects, the restriction,  $\sum_{j=1}^4 \alpha_{qj} = 0$ , is imposed for each QTL.

Let  $\mathbf{y}$  represent an  $n \times 1$  vector indicating the observed phenotypic values of a quantitative trait for  $n$  individuals of the F2 population. We can apply a linear model for  $\mathbf{y}$ , in which QTL on all chromosomes are considered simultaneously, as follows:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \sum_{q=1}^{N_Q} \mathbf{g}_q + \mathbf{e} \quad (1)$$

where  $\mathbf{X}$  is a known design matrix for a vector of nongenetic effect  $\mathbf{b}$  (including the overall mean),  $\mathbf{g}_q$  ( $q = 1, 2, \dots, N_Q$ ) is an  $n \times 1$  vector for contribution of the  $q$ th QTL in F2 individuals and  $\mathbf{e}$  is the vector of residual (environmental) effect following an  $n$ -variate normal distribution with mean vector  $\mathbf{0}$  and covariance matrix  $\sigma_e^2 \mathbf{I}_n$  ( $\mathbf{I}_n$  is the  $n \times n$  identity matrix). The  $i$ th element of  $\mathbf{g}_q$ ,  $g_{qi}$ , indicates the genetic contribution of the  $q$ th QTL for the  $i$ th F2 individual. When the grandparental origins of two alleles at the  $q$ th QTL are  $j$  and  $k$  ( $j, k = 1, 2, 3, 4$ ),  $g_{qi}$  is expressed as

$$g_{qi} = \alpha_{qj} + \alpha_{qk} + \delta_{qjk} \quad (2)$$

One of the purposes of the proposed method is to discriminate whether the F0 grandparents are homo-

zygous or heterozygous at QTL, which we will refer to as the ‘QTL state’ in a grandparent. There are four possible patterns for the combination of QTL states in a grandfather and a grandmother; (i) the two grandparents are homozygous for alternative alleles (Q1Q1 × Q3Q3), (ii) the grandfather is heterozygous and the grandmother is homozygous (Q1Q2 × Q3Q3), (iii) the grandfather is homozygous and the grandmother is heterozygous (Q1Q1 × Q3Q4) and (iv) the two grandparents are heterozygous (Q1Q2 × Q3Q4). In the proposed method, it does not matter whether two grandparents share common alleles or not. Thus, a mating type such as Qq × qq (backcross type) is classified into pattern (ii) and a type such as Qq × Qq (F2 intercross type) is classified into pattern (iv). The combination of QTL states at the  $q$ th QTL in grandparents are denoted by  $S_q$ , which takes values one of 1, 2, 3 and 4 corresponding to (i), (ii), (iii) and (iv) above.

We obtain  $\alpha_{q1} = \alpha_{q2}$  for a homozygous grandfather and  $\alpha_{q3} = \alpha_{q4}$  for a homozygous grandmother at the  $q$ th QTL. Allelic effects,  $\alpha_{qj}$ , and interactions,  $\delta_{qjkr}$ , are reparameterized according to each of the four QTL states as shown in Table 1, where subscript ‘ $q$ ’ is omitted. As shown in Table 1, for  $S_q = 1$ , allelic effects are expressed as  $\alpha_{q1} = \alpha_{q2} = \alpha_{q1}$  and  $\alpha_{q3} = \alpha_{q4} = \alpha_{q3} (= -\alpha_{q1})$  and interactions are written as  $\delta_{q13} = \delta_{q14} = \delta_{q23} = \delta_{q24} = \delta_{q13}$  and  $\delta_{q12} = \delta_{q34} = 0$ . Moreover, replacing  $\alpha_{q1}$  and  $d_{q13}$  by  $a/2$  and  $d$ , respectively, the genetic contributions of the QTL are written as  $a$ ,  $d$  and  $-a$  for genotypes Q1Q1, Q1Q3 and Q3Q3, respectively, which is a conventional notation of QTL effects for inbred F2 populations. Taking the restriction on allelic effects into account, the numbers of independent parameters related to the effects of a QTL are 2, 5, 5 and 9 for  $S_q = 1, 2, 3$  and 4, respectively, as shown in Table 1. The dimension of a model (1) is changed depending on the QTL states in the grandparents.

The effects of the  $q$ th QTL are expressed by a vector  $\mathbf{a}_q$ , the components of which are independent parameters for allelic and interaction effects at the  $q$ th QTL and specified corresponding to  $S_q$ . We obtain  $\mathbf{a}_q = (a_{q1}, d_{q13})$ ,  $(a_{q1}, a_{q2}, d_{q12}, d_{q13}, d_{q23})$ ,  $(a_{q1}, a_{q3}, d_{q13}, d_{q14}, d_{q34})$  or  $(a_{q1}, a_{q2}, a_{q3}, d_{q12}, d_{q13}, d_{q14}, d_{q23}, d_{q24}, d_{q34})$  for  $S_q = 1, 2, 3$  and 4, respectively, from Table 1. We write genetic effects of all QTL as  $\mathbf{a} = (\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_{N_Q})$  and the QTL states in all QTL from the two F0 grandparents as  $\mathbf{S} = (S_1, S_2, \dots, S_{N_Q})$ .

Let  $G_{qi}$  denote the ordered genotype at the  $q$ th QTL for the  $i$ th F2 individual, which is expressed by the grandparental origins of two alleles the  $i$ th F2 individual carries at the  $q$ th QTL. For example,  $G_{qi} = 12$  means that origins of the paternal allele and maternal allele the  $i$ th

F2 individual carries at the  $q$ th QTL are 1 and 2, respectively. Genotypes at the  $q$ th QTL for all of  $n$  individuals are expressed as  $\mathbf{G}_q = (G_{q1}, G_{q2}, \dots, G_{qn})$  and genotypes at all of QTL are as  $\mathbf{G} = (\mathbf{G}_1, \mathbf{G}_2, \dots, \mathbf{G}_{N_Q})$ . Given the number of QTL,  $N_Q$ , QTL states of the two grandparents,  $\mathbf{S}$ , and QTL genotypes,  $\mathbf{G}$ , the distribution of  $\mathbf{y}$ ,  $p(\mathbf{y} | N_Q, \mathbf{S}, \mathbf{G}, \mathbf{a}, \mathbf{b}, \sigma_e^2)$ , is written as a normal distribution from equation (1) as follows:

$$p(\mathbf{y} | N_Q, \mathbf{S}, \mathbf{G}, \mathbf{a}, \mathbf{b}, \sigma_e^2) = (\sigma_e^2)^{-n/2} \times \exp \left\{ -\frac{1}{2\sigma_e^2} \left( \mathbf{y} - \mathbf{Xb} - \sum_{q=1}^{N_Q} \mathbf{g}_q \right)' \left( \mathbf{y} - \mathbf{Xb} - \sum_{q=1}^{N_Q} \mathbf{g}_q \right) \right\} \quad (3)$$

where  $\mathbf{g}_q = (g_{q1}, g_{q2}, \dots, g_{qn})'$  is a vector of genetic effects at the  $q$ th QTL for F2 individuals, obtained from (2) and Table 1 depending on the value of  $\mathbf{S}$ .

### Sampling of QTL genotype

**Estimation of marker haplotype:** In the proposed method, marker haplotypes (linkage phase among alleles at linked markers) are estimated for all of individuals in the three-generation family in advance of the MCMC cycles (ie, it is excluded from MCMC cycles). Marker haplotypes are estimated following the method of Knott *et al* (1996), where linkage phases of markers are determined for parental individuals so that the recombination events between adjacent markers on the gametes transmitted to their offspring are minimized. For individuals of each generation, marker haplotypes are determined generation by generation based on the observed marker genotypes. Firstly, marker haplotypes of F0 grandparents are determined by the observed marker genotypes of their F1 progeny. Secondly, the marker haplotype of each F1 individual is determined by the observed marker genotypes of its F2 progeny and confirmed by the haplotypes of its F0 parents. Lastly, marker haplotypes of F2 individuals are determined by the haplotypes of their F1 parents.

In the method of Knott *et al* (1996), marker haplotypes of sires were determined, based on the observed marker genotypes of half-sib progeny, where markers at which sires were homozygous were uninformative and omitted from consideration. In a three-generation family, even markers at which F0 grandparents are homozygous are informative for detection of QTL and included in consideration of marker haplotypes unless an F0 grandfather and an F0 grandmother are homozygous with a same allele at the markers.

For sampling QTL genotypes, the reconstructed marker haplotype of each individual is translated into a combination of ordered genotypes expressed in terms of parental allelic origins at markers included in the haplotype. Consider, for example, three linked markers A, B and C, in which it is assumed that an F0 male and an F0 female have marker haplotypes A1B1C1/A2B2C2 and A3B3C3/A4B4C4, respectively, and their F1 offspring has marker haplotypes A1B1C2/A3B4C4. Then, for the F1 individual, the combination of ordered genotypes with parental allelic origins at A–B–C is written as 13–14–24, where allelic origins 1 and 2 (3 and 4) indicate paternal allele and maternal allele of an F0 father (mother). Hereafter, ‘haplotype’ means the combination of ordered genotypes expressed with the parental allelic

**Table 1** Correspondence between effects of alleles and interactions indicated by grandparental haplotype origin and actual model parameters for each of QTL states in F0 grandparents

S	Pattern	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha_4$	$\delta_{12}$	$\delta_{13}$	$\delta_{14}$	$\delta_{23}$	$\delta_{24}$	$\delta_{34}$	$n_p$
1	Q1Q1 × Q3Q3	$a_1$	$a_1$	$a_3$	$a_3$	0	$d_{13}$	$d_{13}$	$d_{13}$	0	0	2
2	Q1Q2 × Q3Q3	$a_1$	$a_2$	$a_3$	$a_3$	$d_{12}$	$d_{13}$	$d_{13}$	$d_{23}$	$d_{23}$	0	5
3	Q1Q1 × Q3Q4	$a_1$	$a_1$	$a_3$	$a_4$	0	$d_{13}$	$d_{14}$	$d_{13}$	$d_{14}$	$d_{34}$	5
4	Q1Q2 × Q3Q4	$a_1$	$a_2$	$a_3$	$a_4$	$d_{12}$	$d_{13}$	$d_{14}$	$d_{23}$	$d_{24}$	$d_{34}$	9

$n_p$  is the number of independent parameters for effects of each QTL state.

origins at markers concerned as described in the above example.

Ordered genotypes of F1 progeny at the markers homozygous in either of F0 grandparents cannot be uniquely expressed in terms of parental allelic origins. Consider a marker B, at which an F0 grandfather has ordered genotype B1B1, an F0 grandmother has ordered genotype B3B4 and their F1 offspring has ordered genotype B1B3. Then, the ordered genotype of the F1 individual could be expressed in terms of parental allelic origins as 13 or 23, since an allele B1 has two possible parental origins, 1 and 2. Accordingly, when including markers homozygous in F0 grandparents in haplotype, F1 individuals have more than one possible haplotype. Consider, for example, a haplotype of three markers, A, B and C, which are located on a chromosome in this order with recombination values  $r_{AB}$  between A and B and  $r_{BC}$  between B and C, respectively. When the genotypes of an F0 grandfather and an F0 grandmother are A1B1C1/A2B1C2 and A3B3C3/A4B4C4, respectively, and their F1 offspring has genotype A1B1C1/A3B3C3, the two possible haplotypes of the F1 individual, which can be denoted by a combination of ordered genotypes at A–B–C, are expressed as 13-13-13 and 13-23-13. In this case, the ratio of the probabilities for the occurrences of two possible haplotypes is  $(1-r_{AB})(1-r_{BC})/r_{AB}r_{BC}$ .

The estimated haplotypes for all individuals in the family are denoted by  $\mathbf{H}$ . For the individuals whose marker haplotypes are not uniquely determined, as in the above example, all of possible haplotypes of the individuals are included in  $\mathbf{H}$  with the probabilities of occurrence of each haplotype.

**Probability distribution for sampling QTL genotype:** Let  $\lambda_q$  denote the location of the  $q$ th QTL and  $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_{N_Q})$  denote the locations of all QTL. For updating the QTL genotype  $\mathbf{G}$  in the MCMC procedure for F2 individuals, probability distributions of QTL genotypes of F1 individuals and F2 individuals given  $\mathbf{H}$  are required in order to sample a value of  $\mathbf{G}$ . For the F1 individuals, the ordered genotypes expressed in terms of founder alleles of the  $q$ th QTL are expressed as  $F_q = (F_{q1}, F_{q2}, \dots, F_{qm})$ , and genotypes at all of QTL are  $\mathbf{F} = (F_1, F_2, \dots, F_{N_Q})$ . Given  $\mathbf{H}$ , the joint distribution of QTL genotypes of F1 and F2 individuals,  $p(\mathbf{G}, \mathbf{F} | \mathbf{H}, \boldsymbol{\lambda})$ , has the form

$$p(\mathbf{G}, \mathbf{F} | \mathbf{H}, \boldsymbol{\lambda}) = \prod_{q=1}^{N_Q} [p(\mathbf{F}_q | \mathbf{H}, \boldsymbol{\lambda}) \prod_{i=1}^n p(G_{qi} | F_{qi1}, F_{qi2}, \mathbf{H}, \boldsymbol{\lambda})] \quad (4)$$

where  $F_{qi1}$  and  $F_{qi2}$  are the genotypes at the  $q$ th QTL for the F1 mother and father of the  $i$ th F2 individual. In equation (4), we can write  $p(\mathbf{F}_q | \mathbf{H}, \boldsymbol{\lambda}) = \prod_{j=1}^m p(F_{qj} | \mathbf{H}, \boldsymbol{\lambda})$ , where  $p(F_{qj} | \mathbf{H}, \boldsymbol{\lambda})$  is obtained in terms of the recombination values among QTL and markers linked to QTL.

In the proposed method, we calculate  $p(F_{qj} | \mathbf{H}, \boldsymbol{\lambda})$  following the method of Haley *et al* (1994). Consider a simple example as follows. There are two F0 individuals and an F1 offspring and two linked markers are denoted by A and B. The haplotypes of A and B are written as 12-12 and 34-34 for the two F0 parents and we assume that the genotype of an F1 progeny is (in terms of parental

allelic origins) 13-14. Suppose that a QTL (denoted by Q) is located between A and B with recombination values  $r_{AQ}$  and  $r_{QB}$  for intervals A–Q and Q–B. Ordered genotypes at A–Q–B of two F0 parents are written as 12-12-12 and 34-34-34 in terms of parental allelic origins. There are four possible QTL genotypes expressed by founder allelic origins, that is, 13, 14, 23 and 24, for the F1 individual, thus, possible ordered genotypes of A–Q–B are 13-13-14, 13-14-14, 13-23-14 and 13-24-14. Denoting the probability that the QTL genotype of the F1 individual is  $ij$  by  $p_{ij}$ , we obtain, using the method of Haley *et al* (1994),  $p_{13} = K(1-r_{AQ})^2 r_{QB}(1-r_{QB})$ ,  $p_{14} = Kr_{AQ}(1-r_{AQ})(1-r_{QB})^2$ ,  $p_{23} = Kr_{AQ}(1-r_{AQ})r_{QB}^2$  and  $p_{24} = Kr_{AQ}r_{QB}(1-r_{QB})$ , where  $K$  is a constant.

In this example, the probability of the QTL genotype for the F1 individual being  $ij$ ,  $p(F_{qj} = kl | \mathbf{H}, \boldsymbol{\lambda})$  can be expressed as

$$p(F_{qj} = kl | \mathbf{H}, \boldsymbol{\lambda}) = \frac{p_{kl}}{\sum_{gh \in \{13, 14, 23, 24\}} p_{gh}}$$

where  $\sum_{gh \in \{13, 14, 23, 24\}}$  means summation over all possible QTL genotypes.

When more than one haplotype is possible for an F1 individual, the above formula for  $p(F_{qj} = kl | \mathbf{H}, \boldsymbol{\lambda})$  should be modified accordingly. For the case of two possible haplotypes  $H_1$  and  $H_2$ ,  $p(F_{qj} = kl | \mathbf{H}, \boldsymbol{\lambda})$  can be written as

$$p(F_{qj} = kl | \mathbf{H}, \boldsymbol{\lambda}) = \frac{P(H_1)p_{kl}^{(1)} + P(H_2)p_{kl}^{(2)}}{P(H_1) \sum_{gh \in \{13, 14, 23, 24\}} p_{gh}^{(1)} + P(H_2) \sum_{gh \in \{13, 14, 23, 24\}} p_{gh}^{(2)}}$$

where  $P(H_i)$  is the probability of occurrence of  $i$ th haplotype and  $p_{kl}^{(i)}$  is the probability that the QTL genotype is  $kl$  given haplotype  $H_i$  ( $i = 1, 2$ ).

The probability of  $G_{qi}$  given  $F_{qi1}$  and  $F_{qi2}$ ,  $p(G_{qi} | F_{qi1}, F_{qi2}, \mathbf{H}, \boldsymbol{\lambda})$ , is obtained in a similar way. Let the paternal allele and maternal allele of an F1 father (mother) of an F2 individual be denoted by 1 and 2 (3 and 4), respectively. The probability that an F2 individual receives alleles of  $k$  and  $l$  from its F1 parents given  $\mathbf{H}$  is calculated in the same way as  $p(F_{qj} = kl | \mathbf{H}, \boldsymbol{\lambda})$  ( $kl = 13, 14, 23, 24$ ). Considering the F0 grandparental origin of the QTL alleles carried by each of F1 parents, the grandparental origin of the QTL alleles of the F2 individual is obtained. For example, when QTL genotypes of two F1 parents are expressed as  $F_{qi1} = 23$  and  $F_{qi2} = 14$  and two QTL alleles of an F2 progeny are expressed in terms of F1 parental origin as 2 and 4, QTL alleles of the F2 individual can be related to F0 grandparental origins as  $G_{qi} = 34$ . When updating the QTL genotype of each F1 and F2 individual in the MCMC procedure, a QTL genotype is sampled from prior probability distribution (Equation (4)).

#### Statistical procedure for QTL estimation

**Prior and posterior distribution for the model parameters and QTL genotype:** The model parameters and QTL genotypes to be inferred via MCMC algorithm are written as  $\boldsymbol{\theta} = (N_Q, \mathbf{a}, \boldsymbol{\lambda}, \mathbf{S}, \mathbf{G}, \mathbf{E}, \mathbf{b}, \sigma_e^2)$ . As described below, marker haplotype,  $\mathbf{H}$ , is estimated in advance of MCMC cycles, thus,  $\mathbf{H}$  is treated as known variable and is not updated in the MCMC cycles. The joint prior density function for  $\boldsymbol{\theta}$  can be calculated as the product of

the prior density for each of the components

$$p(\boldsymbol{\theta}|\mathbf{H}) = p(N_Q)p(\mathbf{a}|N_Q, \mathbf{S})p(\boldsymbol{\lambda}|N_Q)p(\mathbf{S}|N_Q) \times p(\mathbf{G}, \mathbf{F}|\mathbf{H}, \boldsymbol{\lambda})p(\mathbf{b})p(\sigma_e^2) \quad (5)$$

The prior distribution of the number of QTL,  $p(N_Q)$ , is assumed to be a truncated Poisson distribution with mean  $\mu$  and a predetermined maximum number  $N_{Q\max}$ . The formula (4) gives the prior density for  $\mathbf{G}$  and  $\mathbf{F}$ ,  $p(\mathbf{G}, \mathbf{F}|\mathbf{H}, \boldsymbol{\lambda})$ . Other priors are assumed to be uniform distributions over the possible values of the parameters. The joint posterior density function is written as

$$p(\boldsymbol{\theta}|\mathbf{y}, \mathbf{H}) = p(\boldsymbol{\theta}|\mathbf{H})p(\mathbf{y}|N_Q, \mathbf{S}, \mathbf{G}, \mathbf{a}, \mathbf{b}, \sigma_e^2) \quad (6)$$

**MCMC sampling:** The model parameters and QTL genotypes  $\boldsymbol{\theta} = (N_Q, \mathbf{a}, \boldsymbol{\lambda}, \mathbf{S}, \mathbf{G}, \mathbf{F}, \mathbf{b}, \sigma_e^2)$  are estimated using MCMC algorithm. The MCMC algorithm consists of the following steps:

a. updating the effects at each QTL

$$\mathbf{a} = (\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_{N_Q});$$

b. updating the fixed effects  $\mathbf{b}$  and residual variance  $\sigma_e^2$ ;

c. updating the QTL locations  $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_{N_Q})$ ;

d. updating the QTL genotypes  $\mathbf{F}$  and  $\mathbf{G}$  for each of F1 and F2 individuals;

e. updating the QTL states in grandparents  $\mathbf{S} = (S_1, S_2, \dots, S_{N_Q})$  and changing the parameterization of model (1) accordingly;

f. adding one new QTL to the model or removing one existing QTL from the model.

The fixed effects  $\mathbf{b}$  and residual variance  $\sigma_e^2$  were updated (step b) using Gibbs sampling (Geman and Geman, 1984). The full conditional distributions for the elements of  $\mathbf{b}$ , general mean and sex effect were normal distributions and that for  $\sigma_e^2$  was an inverse  $\chi^2$  distribution with degrees of freedom equal to  $n-2$ . Updating of other parameters was carried out using Metropolis-Hastings sampling (Metropolis *et al.*, 1953; Hastings, 1970). Neither the estimation of marker haplotypes nor handling of missing genotype data is included in MCMC cycles.

**Updating QTL effects:** The QTL effects  $\mathbf{a}$  are updated locus by locus and component by component. To update each of allelic effects and interactions, a random variable  $u$  is simulated independently from a symmetric uniform distribution around zero. The new proposal value is obtained as the sum of the existing value and  $u$ . Denoting the new proposal value of  $\alpha_{qj}$  by  $\alpha_{qj}^*$ , we take  $\alpha_{qj}^* = \alpha_{qj} + u$ . The new proposal value of  $\mathbf{a}$ , denoted by  $\mathbf{a}^*$  one component of which is replaced by the new value, is accepted with probability

$$\min \left\{ 1, \frac{p(\mathbf{y}|N_Q, \mathbf{S}, \mathbf{G}, \mathbf{a}^*, \mathbf{b}, \sigma_e^2)p(\mathbf{a}^*|N_Q, \mathbf{S})}{p(\mathbf{y}|N_Q, \mathbf{S}, \mathbf{G}, \mathbf{a}, \mathbf{b}, \sigma_e^2)p(\mathbf{a}|N_Q, \mathbf{S})} \right\} \quad (7)$$

**Updating QTL locations:** As seen in the method of Sillanpää and Arjas (1998, 1999) and Yi and Xu (2001), we do not fix the order of QTL when updating the QTL locations. The locations of QTL are modified for one QTL at a time. For the  $q$ th QTL, a new proposal value  $\lambda_q^*$  is sampled from a symmetric uniform distribution centered

at the previous value  $\lambda_q$ . Accordingly, a new genotype at the  $q$ th QTL is proposed for each of F1 and F2 individuals with probability  $q(\mathbf{G}_q^*, \mathbf{F}_q^*)$ . The new value  $\lambda_q^*$  and  $(\mathbf{G}_q^*, \mathbf{F}_q^*)$  are accepted jointly with probability

$$\min \left\{ 1, \frac{p(\mathbf{y}|N_Q, \mathbf{S}, \mathbf{G}^*, \mathbf{a}, \mathbf{b}, \sigma_e^2)p(\lambda_q^*|N_Q)p(\mathbf{G}^*, \mathbf{F}^*|\mathbf{H}, \lambda_q^*)q(\mathbf{G}_q, \mathbf{F}_q)}{p(\mathbf{y}|N_Q, \mathbf{S}, \mathbf{G}, \mathbf{a}, \mathbf{b}, \sigma_e^2)p(\lambda_q|N_Q)p(\mathbf{G}, \mathbf{F}|\mathbf{H}, \lambda)q(\mathbf{G}_q^*, \mathbf{F}_q^*)} \right\} \quad (8)$$

where  $\lambda^*$ ,  $\mathbf{G}^*$  and  $\mathbf{F}^*$  are proposal values of  $\lambda$ ,  $\mathbf{G}$  and  $\mathbf{F}$  with the elements corresponding to the  $q$ th QTL being replaced by new values. We assumed that proposal probability  $q(\mathbf{G}_q, \mathbf{F}_q)$  is the conditional probability of the QTL genotypes given genotypes of markers linked to the QTL and the recombination values among the loci and equivalent to  $p(\mathbf{G}, \mathbf{F}|\mathbf{H}, \lambda)$ . Therefore, the acceptance probability (8) is simplified to

$$\min \left\{ 1, \frac{p(\mathbf{y}|N_Q, \mathbf{S}, \mathbf{G}^*, \mathbf{a}, \mathbf{b}, \sigma_e^2)p(\lambda_q^*|N_Q)}{p(\mathbf{y}|N_Q, \mathbf{S}, \mathbf{G}, \mathbf{a}, \mathbf{b}, \sigma_e^2)p(\lambda_q|N_Q)} \right\} \quad (9)$$

**Updating QTL genotype of F1 and F2 individuals:** In this step, QTL genotypes  $\mathbf{G}$  and  $\mathbf{F}$  are updated QTL by QTL and individual by individual without the change of the QTL locations. A new genotype at the  $q$ th QTL of the  $j$ th F1 individual  $F_{qj}^*$  is sampled from  $p(F_{qj}|\mathbf{H}, \lambda)$  based on the  $\lambda_q$  and recombination values between the QTL and linked markers. In the meantime, the QTL genotypes for F2 offspring of the  $j$ th F1 individual are also modified accordingly. Assuming that the  $j$ th F1 individual is mated with the  $k$ th F1 individual to produce the  $i$ th F2 individual, a new QTL genotype is sampled from  $p(G_{qi}|F_{qj}^*, F_{qk}, \mathbf{H}, \lambda)$ . Using the similar argument as used in derivation of equation (9), the probability of accepting the new genotypes at the  $q$ th QTL for the  $j$ th F1 individual and its F2 progenies is simplified as

$$\min \left\{ 1, \frac{p(\mathbf{y}|N_Q, \mathbf{S}, \mathbf{G}^*, \mathbf{a}, \mathbf{b}, \sigma_e^2)}{p(\mathbf{y}|N_Q, \mathbf{S}, \mathbf{G}, \mathbf{a}, \mathbf{b}, \sigma_e^2)} \right\}$$

where  $\mathbf{G}^*$  is  $\mathbf{G}$  with the genotypes of the  $q$ th QTL for F2 progeny of the  $j$ th F1 parent replaced by the proposal values.

After update of the QTL genotypes for each of F1 individuals and its F2 progeny, each of the F2 individuals are updated for the QTL genotypes individual by individual. A proposal value for the QTL genotype of the  $i$ th F2 individual  $G_{qi}^*$  is sampled from the distribution  $p(G_{qi}|F_{qi1}, F_{qi2}, \mathbf{H}, \lambda)$  and accepted with probability

$$\min \left\{ 1, \frac{p(\mathbf{y}|N_Q, \mathbf{S}, \mathbf{G}^*, \mathbf{a}, \mathbf{b}, \sigma_e^2)}{p(\mathbf{y}|N_Q, \mathbf{S}, \mathbf{G}, \mathbf{a}, \mathbf{b}, \sigma_e^2)} \right\}$$

where  $\mathbf{G}^*$  contains all elements of  $\mathbf{G}$  except  $G_{qi}$  replaced by  $G_{qi}^*$ .

**Updating the QTL states in F0 grandparents:** The QTL state in grandparents  $\mathbf{S}$  is updated by QTL using an RJ-MCMC algorithm. For the  $q$ th QTL states,  $S_q$ , a new value  $S_q^*$  is proposed with a probability  $q(S_q^*|S_q)$ . According to  $S_q^*$ , the parameter for QTL effects  $\mathbf{a}_q$  is also modified to  $\mathbf{a}_q^*$ . In updating the QTL states, the proposal QTL states are restricted such that the only one QTL allele is changed in the F0 grandparents. Thus, the possible moves between the previous QTL state and the new QTL state are as follows; (i)  $S_q^* = 1, 2$  or  $3$  for  $S_q = 1$ , (ii)  $S_q^* = 1, 2$  or  $4$  for

$S_q = 2$ , (iii)  $S_q^* = 1, 3$  or  $4$  for  $S_q = 3$ , (iv)  $S_q^* = 2, 3$  or  $4$  for  $S_q = 4$ . When the dimension of  $\mathbf{a}_q^*$  is greater than  $\mathbf{a}_q$ ,  $\mathbf{a}_q^*$  is obtained via a mapping function,  $\mathbf{a}_q^* = f(\mathbf{a}_q, \mathbf{u})$ , where  $\mathbf{u}$  is a random variable used for dimension matching. When the dimension of  $\mathbf{a}_q^*$  is less than  $\mathbf{a}_q$ , a mapping function ( $\mathbf{a}_q^*, \mathbf{u}$ ) =  $g(\mathbf{a}_q)$  is adopted, where  $g$  is an inverse of  $f$ . As an example of transformation  $\mathbf{a}_q^* = f(\mathbf{a}_q, \mathbf{u})$ , with its Jacobian  $|\partial \mathbf{a}_q^* / \partial (\mathbf{a}_q, \mathbf{u})| = 1$ , consider the case of  $S_q = 1$  (Q1Q1  $\times$  Q3Q3) and  $S_q^* = 2$  (Q1Q2  $\times$  Q3Q3) as follows. For  $S_q = 1$  and  $S_q^* = 2$ , the independent components of  $\mathbf{a}_q$  and  $\mathbf{a}_q^*$  are written as  $\mathbf{a}_q = (a_1, d_{13})$  and  $\mathbf{a}_q^* = (a_1^*, a_2^*, d_{12}^*, d_{13}^*, d_{23}^*)$ , respectively. Considering the variable for dimension matching  $\mathbf{u} = (u_1, u_2, u_3)$ , a transformation  $\mathbf{a}_q^* = f(\mathbf{a}_q, \mathbf{u})$  can be expressed as

$$\begin{aligned} a_1^* &= a_1 + 0.5u_1, \\ a_2^* &= a_1 - 0.5u_1, \\ d_{12}^* &= u_2, \\ d_{13}^* &= d_{13} + 0.5u_3, \end{aligned}$$

and

$$d_{23}^* = d_{13} - 0.5u_3,$$

We can easily show that the Jacobian  $|\partial \mathbf{a}_q^* / \partial (\mathbf{a}_q, \mathbf{u})| = 1$ . For other combinations of  $S_q$  and  $S_q^*$ , a transformation between  $\mathbf{a}_q$  and  $\mathbf{a}_q^*$  is obtained in a similar way.

The proposal values  $S_q^*$  and  $\mathbf{a}_q^*$  are accepted with probability

$$\min \left\{ 1, \frac{p(\mathbf{S}^* | N_Q) p(\mathbf{a}^* | N_Q, \mathbf{S}^*) p(\mathbf{y} | N_Q, \mathbf{S}^*, \mathbf{G}, \mathbf{a}^*, \mathbf{b}, \sigma_e^2) q(S_q | S_q^*)}{p(\mathbf{S} | N_Q) p(\mathbf{a} | N_Q, \mathbf{S}) p(\mathbf{y} | N_Q, \mathbf{S}, \mathbf{G}, \mathbf{a}, \mathbf{b}, \sigma_e^2) q(S_q^* | S_q)} \right\}$$

where  $\mathbf{S}^*$  and  $\mathbf{a}^*$  differs from  $\mathbf{S}$  and  $\mathbf{a}$  in that elements  $S_q$  and  $\mathbf{a}_q$  are replaced by the proposal values  $S_q^*$  and  $\mathbf{a}_q^*$ , respectively.

**Updating QTL number:** The QTL number  $N_Q$  is updated by adding one new QTL to a model with probability  $p_a$  or deleting one existing QTL from a model with probability  $p_d$  in the way described in Jannink and Fernando (2004) and Sillanpää *et al* (2004). Therefore, for a proposal of QTL number  $N_Q^*$ , there are three possible values;  $N_Q^* = N_Q + 1$ ,  $N_Q^* = N_Q - 1$  and  $N_Q^* = N_Q$  with probabilities  $p_a$ ,  $p_d$  and  $1 - p_a - p_d$ .

When attempting to add one new QTL, firstly the location and the grandparental QTL states are randomly obtained by sampling from uniform distributions over their possible values. Next, effects of the QTL are determined from uniform distributions and genotypes of F1 and F2 individuals are obtained by equation (4) depending on the haplotypes of markers linked to the QTL and the parental QTL genotypes as explained in the section 'Probability distribution for sampling QTL genotypes'. The new QTL is accepted with probability

$$\min \left\{ 1, \frac{p(\mathbf{y} | N_Q + 1, \mathbf{S}^*, \mathbf{G}^*, \mathbf{a}^*, \mathbf{b}, \sigma_e^2)}{p(\mathbf{y} | N_Q, \mathbf{S}, \mathbf{G}, \mathbf{a}, \mathbf{b}, \sigma_e^2)} \times \frac{\mu p_d}{p_a (N_Q + 1)} \right\}$$

where  $\mathbf{S}^*$ ,  $\mathbf{G}^*$  and  $\mathbf{a}^*$  indicate  $\mathbf{S}$ ,  $\mathbf{G}$  and  $\mathbf{a}$  with the  $N_Q + 1$  elements added corresponding with the new QTL.

For deleting an existing QTL, a random choice is made among the existing QTL. The probability for its deletion is

$$\min \left\{ 1, \frac{p(\mathbf{y} | N_Q - 1, \mathbf{S}^*, \mathbf{G}^*, \mathbf{a}^*, \mathbf{b}, \sigma_e^2)}{p(\mathbf{y} | N_Q, \mathbf{S}, \mathbf{G}, \mathbf{a}, \mathbf{b}, \sigma_e^2)} \times \frac{p_a N_Q}{\mu p_d} \right\}$$

where  $\mathbf{S}^*$ ,  $\mathbf{G}^*$  and  $\mathbf{a}^*$  mean  $\mathbf{S}$ ,  $\mathbf{G}$  and  $\mathbf{a}$  except that the items corresponding to the  $q$ th QTL are replaced by the proposal values.

### A simulation study

**Simulated data:** The proposed method was evaluated empirically by analyzing simulated three-generation families originated by the cross between two F0 grandparents sampled from outbred founder populations. The three-generation families included two F0 grandparents, 40 F1 individuals, consisting of 20 male and 20 female subjects, and 400 F2 individuals in 20 full-sib families produced by crossing each F1 males to a different F1 female, so that the size of each F2 full-sib family was 20. It is possible to actually construct three-generation families of such size in pigs and forest trees, for example.

The simulated genome was composed of two chromosomes each with length 100 cM. We assumed the existence of three QTL, referred to as QTL1, QTL2 and QTL3, respectively. The true positions, QTL states in two grandparents (which means whether each of grandparents is homozygous or heterozygous at the QTL) and allelic effects and interactions are given in Table 2. On each chromosome, there were 11 markers, one located every 10 cM.

Two cases, Case I and Case II, were considered with respect to the marker informativeness in F0 grandparents. In Case I, both grandparents were heterozygous and shared no common alleles at any marker. In Case II, we assumed that both grandparents were homozygous for different alleles at all markers (as would be the case for two inbred lines). We simulated 100 data sets and 10 data sets for Case I and Case II, respectively.

Marker genotypes of F1 and F2 individuals were determined by sampling gametes from their parents taking linkage phase and recombination events into consideration. Observed marker genotypes were allocated by reshuffling the linkage phase, consequently, haplotypes  $\mathbf{H}$  had to be estimated. The residual variance was set at  $\sigma_e^2 = 1.0$ . The fixed effects comprised the overall mean (set at 0) and a sex effect, which increased phenotypic values of males by 0.3. Phenotypic records were simulated for F2 individuals as the sum of allelic effects at three QTL, sex effect and a residual effect (environmental effect).

**Models used for the analysis of simulated data:** The proposed method was used to estimate the number of QTL, the QTL state, the position and the genetic effects of each QTL, as well as the overall mean, sex effect and

**Table 2** Parameter values for the QTL states and each allelic effect at three QTL used in the simulation experiments

	Position of QTL	S	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha_4$	Heritability
QTL1	Chromosome1 (55 cM)	4	0.2	-0.4	0.4	-0.2	0.123
QTL2	Chromosome2 (25 cM)	1	0.3	0.3	-0.3	-0.3	0.111
QTL3	Chromosome2 (75 cM)	2	0.1	0.5	-0.3	-0.3	0.136

The heritability is defined as proportion of the phenotypic variance explained by each QTL. The interaction effects between alleles are set at 0.

residual variance. For comparison with conventional genetic models used for QTL analyses in F2 populations, each data set was analyzed using simplified models. In one, it was assumed that the grandparents were homozygous for alternative alleles; in the other, that they were heterozygous and shared no common alleles ( $S = 4$ ). Hereafter, the model used in our new analysis is referred to as M0, and the simplified models as M1 ( $S = 1$ ) and M4 ( $S = 4$ ).

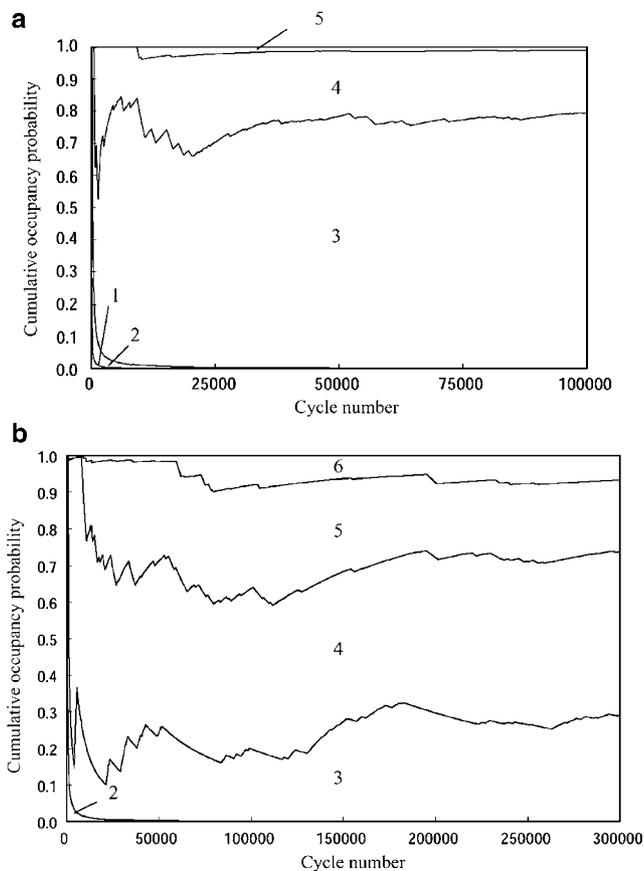
**Estimation procedure for simulated data sets:** Firstly, haplotypes  $\mathbf{H}$  were estimated for all individuals in each of data sets as described above. MCMC cycles were then started for the estimation of parameters. The initial value of the QTL number was set at two, and the corresponding locations were determined by sampling from a uniform distribution over the whole genome length. For the analyses with model M0, the initial state of each QTL in two F0 grandparents was assumed to be  $\mathbf{S} = (S_1, S_2) = (1, 1)$ , indicating that the two grandparents were fixed for different alleles at two QTL. The mean of prior Poisson distribution for the QTL number was set at  $\mu = 2$  and the predetermined maximum number of QTL was set at  $N_{Qmax} = 10$ . The starting values of allelic effects and dominance effect of each QTL and sex effect were randomly chosen from a uniform distribution on  $[-1, 1]$ . The initial values of overall mean and the residual variance were set to the sample mean and the sample variance of the phenotypic values. Prior distributions of all parameters (except  $N_Q$ ,  $\mathbf{G}$  and  $\mathbf{F}$ ) were assumed to be uniform.

For each data set in Case I, MCMC cycles were repeated  $5 \times 10^4$  times and the first  $10^4$  cycles (burn-in) were not used for estimating the parameter values. Sampling was carried out every 10 cycles to reduce serial correlation so that the total number of samples kept was  $4 \times 10^3$ . In each data set in Case II, we repeated MCMC cycles  $2.5 \times 10^5$  times, the first  $1.5 \times 10^5$  cycles of which were discarded. Sampling was carried out every 20 cycles so that the total number of samples kept was  $5 \times 10^3$ . These sampling schemes for Case I and Case II were based on the evaluation of the convergence of MCMC cycles using QTL occupancy probability (Heath, 1997; Uimari and Sillanpää, 2001), as described below.

## Results

The number of cycles required for convergence of MCMC estimation with model M0 was evaluated for Case I and Case II using one simulated data set, on the basis of a plot of cumulative QTL occupation probabilities (Figure 1) as a function of iteration number, that is,  $\Pr(N_Q \leq l | \text{iteration number} = k)$  (where  $1 \leq l \leq 5$ ,  $0 \leq k \leq 10^5$  for Case I and  $2 \leq l \leq 6$ ,  $0 \leq k \leq 3 \times 10^5$  for Case II, respectively) following Heath (1997) and Uimari and Sillanpää (2001). In the analysis of a data set in Case I, QTL occupation probability became stable after  $10^4$  cycles, while more than  $2 \times 10^5$  cycles were required in Case II. Therefore, different burn-in periods and sampling points were adopted as described above. Accordingly, the number of data sets simulated was different between Case I (100 data sets) and Case II (10 data sets).

The summary of the posterior distributions of the QTL number  $N_Q$  are shown in Table 3 for each of the two cases



**Figure 1** Cumulative QTL occupancy probabilities obtained in the analyses of one data set simulated in Case I (a) and in Case II (b). The five areas, one on the top, three between two lines and one on the bottom, in (a) and (b) indicate probabilities of the number of fitted QTL, as denoted by numeral in each area.

over all data sets analyzed with three models, M0, M1 and M4. The mean of the posterior probabilities of each value of  $N_Q$  and the times of posterior modes being consistent with the true value of  $N_Q$  ( $N_Q = 3$ ) in all of analyses are listed in Table 3. In Case I, three QTL affecting the trait were clearly confirmed in the analyses with M0 and M4. In most of data sets in Case I, the posterior probabilities were concentrated about the true value of  $N_Q$  and the posterior modes were consistent with the true value of  $N_Q$  in the analyses using M0 and M4. In Case II, the posterior probabilities of  $N_Q$  were more dispersed. In the analyses with model M1, there was little difference for the posterior probabilities of  $N_Q$  between Case I and Case II. Since all of QTL were assumed to be homozygous for alternative alleles in the grandparents, the difference in the marker informativeness between Case I and Case II was not utilized. In some data sets in which QTL1 was undetected by M1, the posterior modes agreed with  $N_Q = 2$ .

QTL locations were estimated using the posterior QTL intensity function (Sillanpää and Arjas, 1998, 1999) defined as the relative frequency of the cycles in which QTL is detected in small intervals of equal size, say 1 cM, along each chromosome. The means of posterior QTL intensities are presented with standard deviations in Figures 2 and 3.

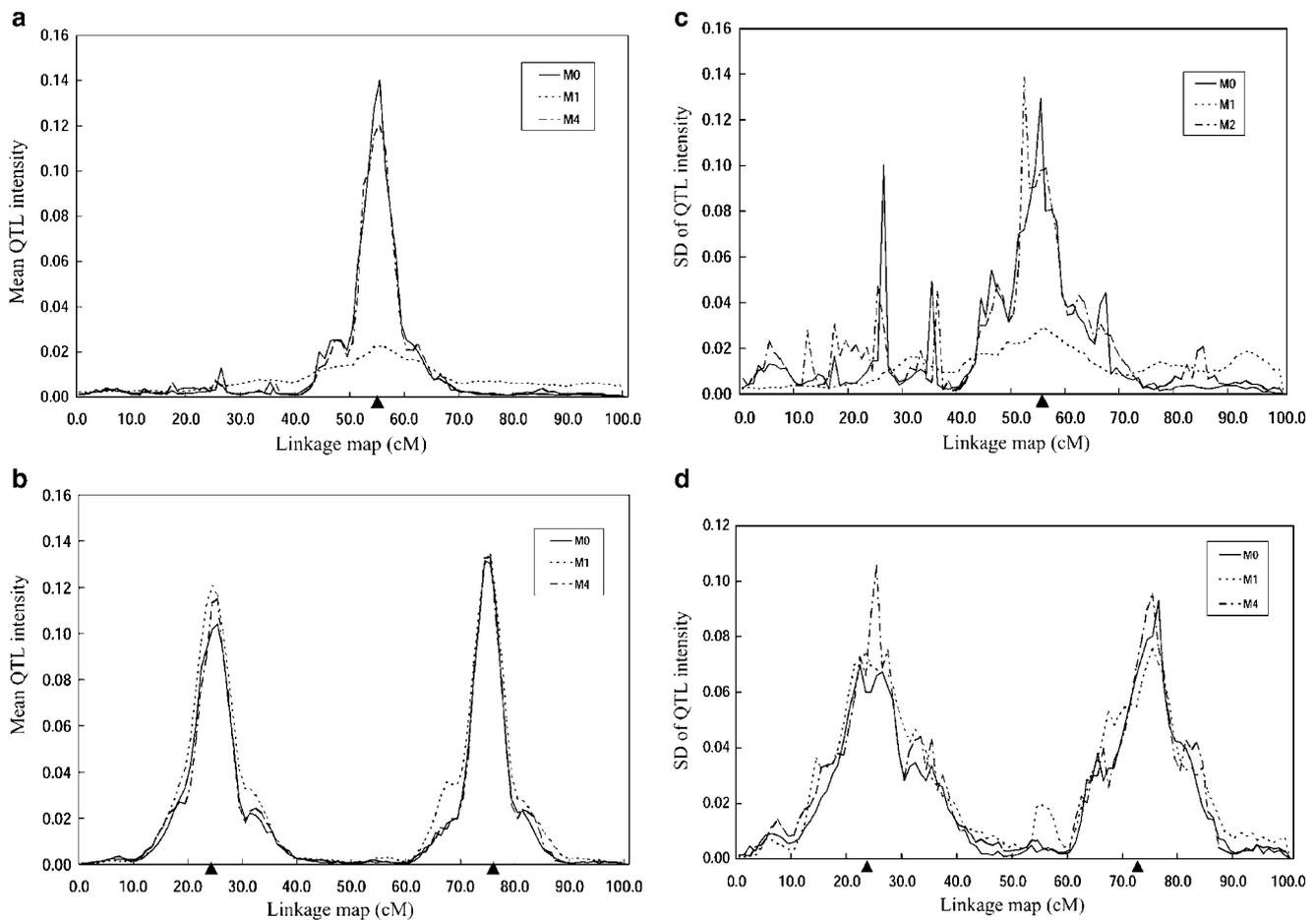
**Table 3** Posterior probabilities of the QTL number and its expectation in the analyses of simulated data sets

Model <sup>a</sup>	Mean of posterior probabilities of $N_Q$							Mean of posterior expectations of $N_Q$	
	0	1	2	3	4	5	$\geq 6$		
Case I	M0	0.0000 (0) <sup>b</sup>	0.0000 (0)	0.0043 (0)	0.7537 (89)	0.2127 (9)	0.0275 (2)	0.0018 (0)	3.366
	M1	0.0000 (0)	0.0048 (0)	0.3978 (47)	0.5761 (53)	0.0211 (0)	0.0002 (0)	0.0000 (0)	2.623
	M4	0.0000 (0)	0.0000 (0)	0.0023 (0)	0.7742 (91)	0.1924 (7)	0.0280 (2)	0.0030 (0)	3.255
Case II	M0	0.0000 (0)	0.0000 (0)	0.0264 (0)	0.3318 (4)	0.3310 (4)	0.1724 (1)	0.1384 (1)	4.146
	M1	0.0000 (0)	0.0047 (0)	0.3545 (4)	0.6275 (6)	0.0132 (0)	0.0001 (0)	0.0000 (0)	2.650
	M4	0.0000 (0)	0.0000 (0)	0.0000 (0)	0.0479 (0)	0.1712 (2)	0.2462 (3)	0.5345 (5)	6.135

We generated 100 data sets in Case I and 10 data sets in Case II, respectively.

<sup>a</sup>Three models with respect to the QTL states in F0 grandparents applied to the analysis of simulated data sets: M0 allows the QTL states variable and M1 and M4 set the QTL states at 1 (both of F0 grandparents are homozygous) and 4 (both of F0 grandparents are heterozygous), respectively.

<sup>b</sup>The times of analyses in which posterior mode was observed at the corresponding value of  $N_Q$ .

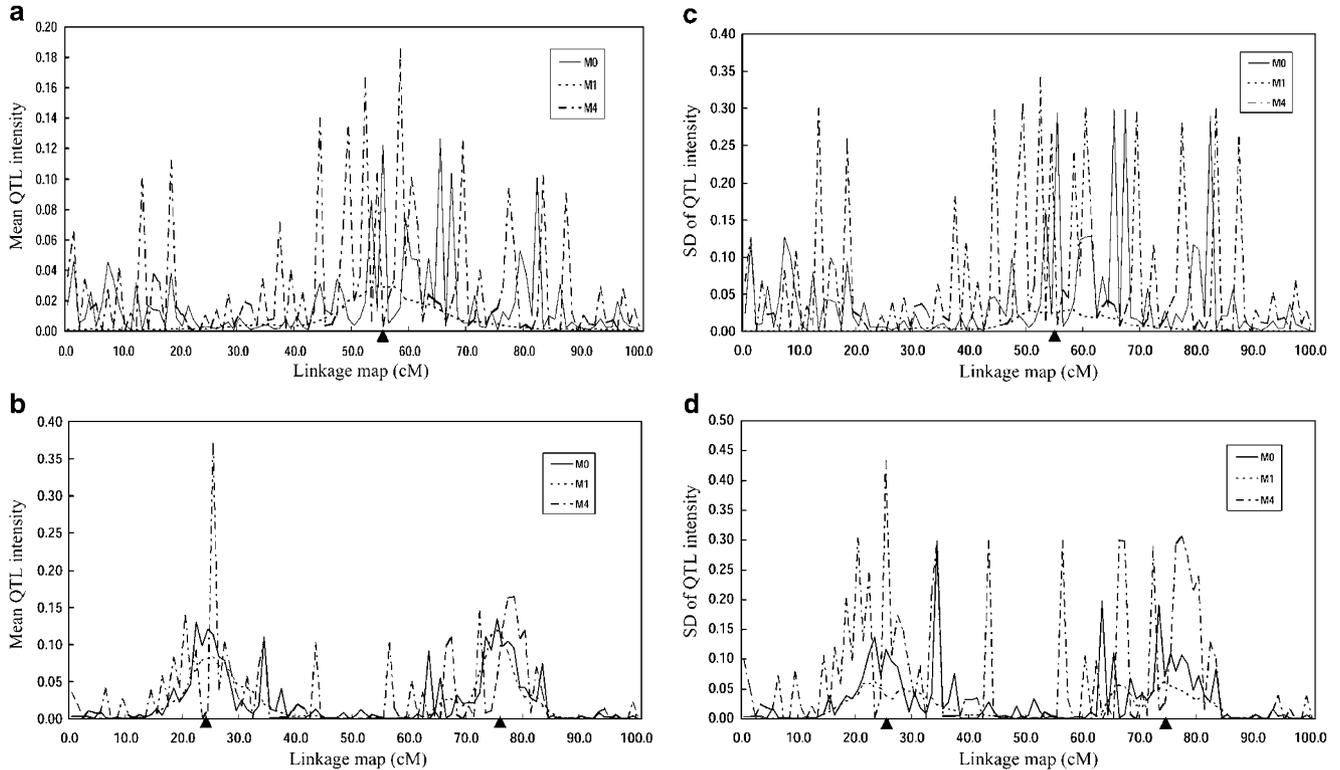


**Figure 2** Histograms of the means and the standard errors of posterior QTL intensities on two chromosomes with bin length of 1 cM over 100 data sets for Case I. There were three simulated QTL, which are indicated by closed triangles, located on 55 cM of chromosome 1, 25 cM of chromosome 2 and 75 cM of chromosome 2, respectively. (a) Mean posterior QTL intensities on chromosome 1, (b) mean posterior QTL intensities on chromosome 2, (c) standard errors of posterior QTL intensities on chromosome 1, (d) standard errors of posterior QTL intensities on chromosome 2.

The probable regions of QTL were determined with a Bayes factor (Sorensen and Gianola, 2002). The Bayes factor, BF, is defined to be the ratio of posterior odds ratio to prior odds ratio for two competing hypotheses:

$$BF = \{p(H_1|y)/p(H_0|y)\} / \{p(H_1)/p(H_0)\}$$

where  $p(H_i|y)$  and  $p(H_i)$  are posterior and prior distributions of  $H_i$  ( $i = 0,1$ ), respectively. For the presence of QTL on each small interval,  $H_1$  means that a QTL exists on the interval and  $H_0$  means that there is not. Considering that the total number of small intervals was 200, the prior mean of the QTL number was 2 and the



**Figure 3** Histograms of the means and the standard errors of posterior QTL intensities on two chromosomes with bin length of 1 cM over 10 data sets for Case II. There were three simulated QTL, which are indicated by closed triangles, located on 55 cM of chromosome 1, 25 cM of chromosome 2 and 75 cM of chromosome 2, respectively. (a) Mean posterior QTL intensities on chromosome 1, (b) mean posterior QTL intensities on chromosome 2, (c) standard errors of posterior QTL intensities on chromosome 1, (d) standard errors of posterior QTL intensities on chromosome 2.

prior distribution of QTL position was uniform, we obtained  $p(H_1) = 2/200 = 1/100$  and  $p(H_0) = 99/100$ . The posterior distribution  $p(H_1|y)$  was determined as QTL intensity on the interval and  $p(H_0|y)$  was given as  $1 - p(H_1|y)$ . The intervals on which we obtained  $\text{BF} > 2$  composed a probable region for QTL, which corresponded to the intervals with QTL intensity  $> 0.0196$ . The number of probable regions of QTL depended on the posterior probability of  $N_Q$ . Thus, in most of data sets of Case I, three probable regions of QTL were confirmed according to the true position of each of the three QTL for models M0 and M4. For some data sets, analyses in which M1 did not detect QTL1, only two probable QTL regions were observed corresponding to QTL2 and QTL3. When the posterior probabilities for  $N_Q > 3$  were not negligible (say, more than 0.3), as seen in the analyses of some data sets in Case II with M0 and M4, more than three probable QTL regions were obtained. For the evaluation of effectiveness in estimation of positions, QTL states, allelic effects and interaction effects of three simulated QTL, the probable QTL regions around the position of each QTL were combined and assigned to it. The intervals (30 cM, 80 cM) of chromosome 1, (0 cM, 50 cM) and [50 cM, 100 cM] of chromosome 2 were regarded as the probable regions of QTL1, QTL2 and QTL3, respectively. In the analyses with M0 and M4, all three QTL were detected in all data sets of Case I and Case II. In the analyses with M1, QTL2 and QTL3 were detected in all data sets, while QTL1 was detected in about half of

all data sets (51 data sets in Case I and six data sets in Case II).

For Case I, peaks of QTL intensities for QTL2 and QTL3 were observed around the true positions in most of data sets (Figure 2b). In Case I, the analyses with M1 provided relatively low peaks around QTL1 (Figure 2a). For Case II, QTL2 and QTL3 were confirmed around the true positions by the analyses with M0 and M4 (Figure 3b); however, other peaks of QTL intensity were often observed on chromosome 1 (Figure 3a). For Case II, QTL intensities were concentrated on some intervals, since the update of QTL locations was hampered due to the low marker informativeness. The variation in QTL intensity on each interval over all data sets appeared much greater in Case II than in Case I (Figures 2c, d, 3c, d).

The posterior probabilities of QTL states in grandparents are summarized in Table 4. In Case I, the means of posterior probabilities of  $S_1$ ,  $S_2$  and  $S_3$  took the highest values at the true QTL states of grandparents. However, the means of posterior probabilities of QTL states for Case II showed the highest value at  $S_i = 4$  ( $i = 1, 2, 3$ ) meaning that both of grandparents are generally estimated to be heterozygous at all QTL, thus, the QTL states of QTL2 and QTL3 were often incorrectly estimated. In Table 4, the means of the posterior probabilities that each of two grandparents was homozygous at each of QTL are also shown. The posterior probability that an F0 grandfather was homozygous at QTL was obtained by  $\text{Pr}(S = 1) + \text{Pr}(S = 3)$ , where  $\text{Pr}(S = i)$  is the probability of

**Table 4** Mean of posterior probabilities of the QTL states

		Mean of posterior probabilities of S				Probabilities of QTL homozygosity	
		S = 1	S = 2	S = 3	S = 4	Grandfather	Grandmother
Case I	QTL1	0.0019 (0.0043) <sup>a</sup>	0.0225 (0.0398)	0.0379 (0.0428)	0.9379 (0.0738)	0.0398 (0.0451)	0.0244 (0.0417)
	QTL2	0.3224 (0.1883)	0.2118 (0.1482)	0.2799 (0.1784)	0.1863 (0.1979)	0.6022 (0.2261)	0.5341 (0.2348)
	QTL3	0.0551 (0.0944)	0.5631 (0.2348)	0.0471 (0.0728)	0.3349 (0.2194)	0.1022 (0.1474)	0.6182 (0.2289)
Case II	QTL1	0.0056 (0.0095)	0.0523 (0.0599)	0.1120 (0.1060)	0.8304 (0.1472)	0.1176 (0.3524)	0.0579 (0.0621)
	QTL2	0.0272 (0.0252)	0.1550 (0.0607)	0.1903 (0.0759)	0.6276 (0.1161)	0.2176 (0.0901)	0.1822 (0.0646)
	QTL3	0.0533 (0.0454)	0.2961 (0.1299)	0.1037 (0.0700)	0.3349 (0.1066)	0.1570 (0.0837)	0.3493 (0.1488)

The QTL states of two F0 grandparents are indicated by S as follows: S = 1, both of F0 grandparents are homozygous; S = 2, F0 grandfather is heterozygous and F0 grandmother is homozygous; S = 3, F0 grandfather is homozygous and F0 grandmother is heterozygous; S = 4, both of F0 grandparents are heterozygous.

<sup>a</sup>Standard error calculated based on 100 data sets in Case I and 10 data sets in Case II.

**Table 5** Mean of posterior expectations for the QTL allelic effects estimated on the probable QTL region

		QTL1				QTL2				QTL3			
		$\alpha_1(0.2)^a$	$\alpha_2(-0.4)$	$\alpha_3(0.4)$	$\alpha_4(-0.2)$	$\alpha_1(0.3)$	$\alpha_2(0.3)$	$\alpha_3(-0.3)$	$\alpha_4(-0.3)$	$\alpha_1(0.1)$	$\alpha_2(0.5)$	$\alpha_3(-0.3)$	$\alpha_4(-0.3)$
Case I	M0 <sup>b</sup>	0.1857 (0.1238) <sup>c</sup>	-0.3758 (0.1576)	0.3304 (0.2133)	-0.1412 (0.1928)	0.2889 (0.0864)	0.2864 (0.0813)	-0.2889 (0.0881)	-0.2867 (0.0994)	0.1230 (0.1310)	0.4491 (0.1608)	-0.2832 (0.1100)	-0.2885 (0.1001)
	M1	-0.1397 (0.0279)	-0.1397 (0.0279)	0.1397 (0.0279)	0.1397 (0.0279)	0.2992 (0.0464)	0.2992 (0.0464)	-0.2992 (0.0464)	-0.2992 (0.0464)	0.2979 (0.0524)	0.2979 (0.0524)	-0.2979 (0.0524)	-0.2979 (0.0524)
	M4	0.2017 (0.1200)	-0.4000 (0.1374)	0.3529 (0.2102)	-0.1546 (0.1979)	0.3034 (0.1256)	0.2847 (0.1135)	-0.2989 (0.1271)	-0.2893 (0.1341)	0.1189 (0.1340)	0.4849 (0.1298)	-0.2927 (0.1383)	-0.3111 (0.1189)
Case II	M0	-0.0826 (0.0676)	-0.0638 (0.0692)	0.0643 (0.0901)	0.0813 (0.0735)	0.2773 (0.0903)	0.2855 (0.0817)	-0.2965 (0.0977)	-0.2658 (0.0812)	0.2786 (0.0882)	0.2668 (0.0800)	-0.2967 (0.0874)	-0.2607 (0.0802)
	M1	-0.1569 (0.0243)	-0.1569 (0.0243)	0.1569 (0.0243)	0.1569 (0.0243)	0.2947 (0.0430)	0.2947 (0.0430)	-0.2947 (0.0430)	-0.2947 (0.0430)	0.3120 (0.0370)	0.3120 (0.0370)	-0.3120 (0.0370)	-0.3120 (0.0370)
	M4	-0.0760 (0.0467)	-0.0627 (0.0397)	0.0488 (0.0488)	0.0578 (0.0430)	0.2256 (0.0817)	0.2049 (0.0798)	-0.2288 (0.0880)	-0.2017 (0.0720)	0.2345 (0.0826)	0.2477 (0.0990)	-0.2584 (0.0960)	-0.2237 (0.0816)

<sup>a</sup>Simulated values of allelic effects for each QTL.

<sup>b</sup>Estimates of allelic effects using model-averaging means listed for M0.

<sup>c</sup>Standard deviations obtained in 100 data sets in Case I and 10 data sets in Case II.

S = i for the QTL. Similarly, the posterior probability of an F0 grandmother being homozygous was given by Pr(S = 1) + Pr(S = 2). These showed good agreement with the true QTL genotypes of grandparents in Case I, while the posterior probabilities obtained in Case II were low for all of QTL meaning that their homozygosity was less accurately estimated (Table 4).

The allelic effects and interaction effects of each QTL were estimated as the posterior expectations given each of QTL states on the probable QTL regions. Table 5 shows the means of the estimates of QTL allelic effects using model averaging (Ball, 2001; Sillanpää and Corander, 2002), in which the posterior expectations of allelic effects were averaged with weight according to the posterior probabilities of each QTL state. In Case I, the estimates were slightly less than the true effects except for  $\alpha_1$  of QTL3 in the analyses with M0. The estimates of allelic effects and interaction effects were almost unbiased for all of three QTL in the analyses with model M4 for Case I. For the analyses with M0 and M4 in Case II,

since two alleles of an F0 grandparent heterozygous at QTL were not well discriminated due to the low marker informativeness,  $\alpha_1$  and  $\alpha_2$  of QTL1 and QTL3 and  $\alpha_3$  and  $\alpha_4$  of QTL1 were not separated and estimated incorrectly. As expected, the analyses with model M1 provided biased estimates for QTL1 and QTL3 with QTL state S ≠ 1. The estimates of all of QTL interaction effects ranged from -0.0515 to 0.0616 with standard deviations from 0.1089 to 0.2600 for Case I and from -0.0941 to 0.0772 with standard deviations from 0.0063 to 0.1532 for Case II.

The means and standard deviations of posterior expectations for general mean (intercept), sex effect and residual variance are shown in Table 6 for Case I and Case II. In both cases, the analyses with models M0, M1 and M4 provided similar estimates for general mean and sex effect. In the analyses with M1, residual variance tended to be overestimated, as it contained the variation of effects of undetected QTL (QTL1). In the analyses with M0 and M4, residual variance was correctly estimated in

**Table 6** Means of posterior expectations for the general mean, sex effect and residual variance

		General mean	Sex effect	Residual variance
Case I	M0	-0.0010 (0.1809)*	0.3100 (0.1081)	0.9971 (0.0719)
	M1	-0.0283 (0.1394)	0.3181 (0.1254)	1.2373 (0.0857)
	M4	-0.0490 (0.2027)	0.3087 (0.1104)	0.9876 (0.0727)
Case II	M0	0.1299 (0.0597)	0.2415 (0.0888)	0.8438 (0.1242)
	M1	0.1205 (0.1025)	0.2331 (0.0983)	1.2546 (0.0400)
	M4	0.1194 (0.0469)	0.2330 (0.0738)	0.6851 (0.1518)

\*Standard deviations calculated with 100 data sets in Case I and 10 data sets in Case II.

Case I, while it was underestimated in Case II since a part of residual variance was removed by the fitting of false QTL.

## Discussion

F2 families are suitable experimental populations for QTL mapping in outbred organisms as well as inbred organisms. The analyses of outbred F2 families allowed the detection of heterozygosity in the grandparents, as well as genetic differences between them. The Bayesian methods of Sillanpää and Arjas (1999) and Yi and Xu (2001) are applicable for such analyses: however, the number of founder alleles in F0 grandparents is fixed in the estimation procedure. Although a pedigree analysis method with Bayesian approach, proposed by Uimari and Sillanpää (2001) and Bink *et al* (2002), could be used for F2 families, only biallelic QTL were considered in their methods. In this paper, we proposed a more flexible Bayesian method for the QTL analyses of outbred F2 families, in which we estimate whether each of F0 grandparents is homozygous or heterozygous at QTL. The great benefit of our proposed method is that one does not need to assume biallelic QTL while still maintaining low number of parameters. The accurate estimates of the QTL states and QTL effects are valuable information for breeding programs targeting the QTL. Using a model that parameterizes separately all QTL alleles in F0 grandparents, such as the model of Yi and Xu (2001), it is possible to infer the QTL states in F0 grandparents by comparing the estimates of allelic effects between QTL alleles. However, the posterior probabilities derived here provide more useful and direct information.

In Case I, where the informativeness of grandparental markers was enough to discriminate haplotypes for each of F0 grandparents, analyses with the proposed method (M0) provided accurate estimates for QTL number, QTL states, positions and effects the QTL as well as general mean, sex effect and residual variance in most data sets. A simplified analysis, which assumed that F0 QTL were heterozygous (M4), showed similar effectiveness in estimation. In Case II, where marker informativeness was too low to discriminate haplotypes in the F0 grandparents, there was a tendency to fit excessive false QTL in order to explain phenotypic variation in M0 and especially M4. In the analyses with M1, where QTL states were assumed to be fixed at  $S=1$ , it is not necessary to distinguish the alleles in an F0 individual and similar effectiveness in estimation was observed in Case I and

Case II. However, QTL1 was often undetected with M1 due to the small mean allelic effect. QTL2 and QTL3 with the larger mean allelic effects were successfully detected in most of data sets in the analyses with M1. From the results of simulation experiments and taking the advantage of obtaining information of QTL states in F0 grandparents into account, it could be concluded that the analyses with M0 are preferable to those with M1 and M4 in practical analyses of outbred populations, in which some heterozygosity is anticipated.

The efficiencies of the proposed method depend on the number of F0 grandparents of the F2 family. In this paper, we considered the case of F2 families derived from two F0 grandparents; however, the proposed method can be extended for more grandparents. When the number of pairs of F0 grandparents is  $n$ , the number of all possible QTL states is  $4^n$ . Thus, as the number of pairs of F0 grandparents is much increased, the convergence of the estimates in the MCMC procedure would become slower and the efficiencies may be decreased due to the large number of possible QTL states. In the case of a large number of F0 founders, some restriction for the number of QTL alleles would be useful, as seen in the assumption of biallelic QTL adopted by Uimari and Sillanpää (2001) and Bink *et al* (2002). Updating QTL states individual by individual instead of updating combinations of QTL states in pair of founder individuals would be an alternative choice for treating multiple pairs of F0 founders.

In the QTL mapping of F2 families in outbred plants such as forest trees, it is usual that two genetically diverged individuals are used as F0 grandparents. For the analyses of such F2 families, the proposed method would be suitable. The proposed method would be likewise applicable to the pig F2 families where a small number of F0 grandparents are used.

The marker informativeness of F0 grandparents strongly influences the accuracy of estimation in the proposed method (M0). As shown in simulation experiments, we could obtain accurate estimates of the model parameters when all markers were fully informative in F0 grandparents (Case I), while the QTL number, the QTL states and QTL effects were poorly estimated in case of F0 grandparents homozygous at all markers (Case II). When the marker informativeness is insufficient to discriminate the haplotypes of F0 grandparents, the QTL states tend to be estimated as  $S_q=4$  (that is, both F0 grandparents being heterozygous). This is because the number of parameters for allelic effects and dominance effects is maximized for  $S_q=4$ , and then the better model-fitting is attained under the insufficient marker informativeness, unless  $S_q$  is constrained by the prior.

In the proposed Bayesian method, haplotypes were estimated in advance of MCMC cycles following the method of Knott *et al* (1996) to save computational time. The correct estimation of haplotypes was reliably obtained as there were a high number of progeny per parent individual. When the number of progeny is small, or marker informativeness is low, however, haplotypes might be poorly estimated in this method. Incorrectly estimated haplotypes lead to inaccurate sampling probability of QTL genotypes, and consequently inaccurate estimation of positions, QTL states and genetic effects of QTL. Nevertheless, in the proposed method, QTL states are variable in the estimation process, thus, for QTL in

which a mean effect of two alleles in an F0 grandparent is large and which can be detected assuming its QTL state to be  $S=1$ , relatively accurate mapping might be possible even with incorrect haplotype attribution.

When updating QTL position, QTL genotypes for all of the individuals were updated simultaneously. This is a 'block-update' (Uimari and Sillanpää, 2001; Yi and Xu, 2002). Since change in QTL position causes change in probable QTL genotypes in founder individuals and probable allele transmission to subsequent generations, such a block-update is required. This approach is particularly important since haplotype estimation is excluded from the MCMC cycles, so inference about a position depends on the QTL genotypes. When only a small number of F1 parents are available, some F0 QTL alleles may not be transmitted to F1 parents. It would then be impossible to accurately evaluate the QTL states of the F0 grandparent. A moderate number of F1 parents, say, five pairs are required for the method to work suitably.

One of the practical objectives of QTL mapping in F2 families is the utilization of the detected QTL in the marker-assisted breeding. The QTL states in founder individuals are then of great importance. The posterior probabilities for the QTL states obtained by the proposed method could therefore be used in the design of breeding programmes for outbred plants and animals.

## References

- Andersson L, Haley CS, Ellegren H, Knott SA, Johansson M, Andersson K *et al* (1994). Genetic mapping of quantitative trait loci for growth and fatness in pigs. *Science* **263**: 1771–1774.
- Ball RD (2001). Bayesian methods for quantitative trait loci mapping based on model selection: approximate analysis using the Bayesian information criterion. *Genetics* **159**: 1351–1364.
- Bidanel JP, Milan D, Iannuccelli N, Amigues Y, Boscher MY, Bourgeois F *et al* (2001). Detection of quantitative trait loci for growth and fatness in pigs. *Genet Sel Evol* **33**: 289–309.
- Bink MC, Uimari P, Sillanpää MJ, Janss LLG, Jansen RC (2002). Multiple QTL mapping in related plant populations via a pedigree-analysis approach. *Theor Appl Genet* **104**: 751–762.
- Bink MC, Van Arendonk JA (1999). Detection of quantitative trait loci in outbred populations with incomplete marker data. *Genetics* **151**: 409–420.
- Bliss FA, Arulsekhar S, Foolad MR, Becerra V, Gillen AM, Warburton ML *et al* (2002). An expanded genetic linkage map of *Prunus* based on an interspecific cross between almond and peach. *Genome* **2002**: 520–529.
- Bradshaw Jr HD, Stettler RF (1995). Molecular genetics of growth and development in *Populus*. IV. Mapping QTLs with large effects on growth, form, and phenology traits in a forest tree. *Genetics* **139**: 963–973.
- Costa P, Pot D, Dubos C, Frigerio JM, Pionneau C, Bodénès C *et al* (2000). A genetic map of Maritime pine based on AFLP, RAPD and protein markers. *Theor Appl Genet* **100**: 39–48.
- Dirlewanger E, Moing A, Rothan C, Svanella L, Pronier V, Guye A *et al* (1999). Mapping QTLs controlling fruit quality in peach (*Prunus persica* (L.) Batsch). *Theor Appl Genet* **98**: 18–31.
- Frewen BE, Chen THH, Glenn TH, Davis J, Rohde A, Boerjan W *et al* (2000). Quantitative trait loci and candidate gene mapping of bud set and bud flush in *Populus*. *Genetics* **154**: 837–845.
- Geman S, Geman D (1984). Stochastic relaxation, gibbs distribution, and the Bayesian restoration of images. *IEEE Trans Patt Anal Mach Intell* **6**: 721–741.
- Green PJ (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82**: 711–732.
- Haley CS, Knott SA, Elsen JM (1994). Mapping quantitative trait loci in crosses between outbred lines using least-squares. *Genetics* **136**: 1195–1207.
- Hastings WK (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**: 98–109.
- Heath SC (1997). Markov chain Monte Carlo segregation and linkage analysis for oligogenic models. *Am J Hum Genet* **61**: 748–760.
- Jannink J-L, Fernando RL (2004). On the Metropolis–Hastings acceptance probability to add or drop a quantitative trait locus in Markov chain Monte Carlo-based Bayesian analysis. *Genetics* **166**: 641–643.
- Jannink JL, Wu XL (2003). Estimating allelic number and identity in state of QTLs in interconnected families. *Genet Res* **81**: 133–144.
- Knott SA, Elsen JM, Haley CS (1996). Methods for multiple-marker mapping of quantitative trait loci in half-sib populations. *Theor Appl Genet* **93**: 71–80.
- Lander ES, Botstein D (1989). Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**: 185–199.
- Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953). Equation of state calculations by fast computing machines. *J Chem Phys* **21**: 1087–1092.
- Pérez-Enciso M, Varona L (2000). Quantitative trait loci mapping in F2 crosses between outbred lines. *Genetics* **155**: 391–405.
- Satagopan JM, Yandell BS, Newton MA, Osborn TC (1996). A Bayesian approach to detect quantitative trait loci using Markov chain Monte Carlo. *Genetics* **144**: 805–816.
- Sillanpää MJ, Arjas E (1998). Bayesian mapping of multiple quantitative trait loci from incomplete inbred line cross data. *Genetics* **148**: 1373–1388.
- Sillanpää MJ, Arjas E (1999). Bayesian mapping of multiple quantitative trait loci from incomplete outbred offspring data. *Genetics* **151**: 1605–1619.
- Sillanpää MJ, Corander J (2002). Model choice in gene mapping: what and why. *Trends Genet* **18**: 301–307.
- Sillanpää MJ, Gasbarra D, Arjas E (2004). Comment on 'On the Metropolis–Hastings acceptance probability to add or drop a quantitative trait locus in Markov Chain Monte Carlo-based Bayesian analyses'. *Genetics* **167**: 1037.
- Sorensen D, Gianola D (2002). *Likelihood, Bayesian, and MCMC Methods in Quantitative Genetics*. Springer: New York.
- Uimari P, Hoeschele I (1997). Mapping linked quantitative trait loci using Bayesian analysis and Markov chain Monte Carlo algorithms. *Genetics* **146**: 735–743.
- Uimari P, Sillanpää MJ (2001). Bayesian oligogenic analysis of quantitative and qualitative traits in general pedigrees. *Genet Epidemiol* **21**: 224–242.
- Walling GA, Visscher PM, Andersson L, Rothschild MF, Wang L, Moser G *et al* (2000). Combined analyses of data from quantitative trait loci mapping studies. Chromosome 4 effects on porcine growth and fatness. *Genetics* **155**: 1369–1378.
- Wu RL (1998). Genetic mapping of QTLs affecting tree growth and architecture in *Populus*: implication for ideotype breeding. *Theor Appl Genet* **96**: 447–457.
- Yi N, Xu S (2001). Bayesian mapping of quantitative trait loci under complicated mating designs. *Genetics* **157**: 1759–1771.
- Yi N, Xu S (2002). Linkage analysis of quantitative trait loci in multiple line crosses. *Genetica* **114**: 217–230.