BRIEFINGS IN BIOINFORMATICS. VOL 13. NO 5. 555–568
Advance Access published on 8 February 2012
doi:10.1093/bib/bbr079

Bayesian inference for genomic imprinting underlying developmental characteristics

Runqing Yang, Xin Wang and Yuehua Cui

Submitted: 3rd September 2011; Received (in revised form): 12th December 2011

Abstract

The identification of imprinted genes is becoming a standard procedure in searching for quantitative trait loci (QTL) underlying complex traits. When a developmental characteristic such as growth or drug response is observed at multiple time points, understanding the dynamics of gene function governing the underlying feature should provide more biological information regarding the genetic control of an organism. Recognizing that differential imprinting can be development-specific, mapping imprinted genes considering the dynamic imprinting effect can provide additional biological insights into the epigenetic control of a complex trait. In this study, we proposed a Bayesian imprinted QTL (iQTL) mapping framework considering the dynamics of imprinting effects and model multiple iQTLs with an efficient Bayesian model selection procedure. The method overcomes the limitation of likelihood-based mapping procedure, and can simultaneously identify multiple iQTLs with different gene action modes across the whole genome with high computational efficiency. An inference procedure using Bayes factors to distinguish different imprinting patterns of iQTL was proposed. Monte Carlo simulations were conducted to evaluate the performance of the method. The utility of the approach was illustrated through an analysis of a body weight growth data set in an F_2 family derived from LG/J and SM/J mouse stains. The proposed Bayesian mapping method provides an efficient and computationally feasible framework for genome-wide multiple iQTL inference with complex developmental traits.

Keywords: *Bayes factor; Bayesian model selection; developmental traits; Markov chain Monte Carlo; imprinting pattern; quantitative trait loci*

INTRODUCTION

Genomic imprinting is a genetic phenomenon in which the same genes express differently, depending on their parental origin [1]. On the molecular level, genomic imprinting may result from DNA methylation, histone modification, non-coding RNAs (ncRNA) and even long distance interchromosomal interactions [2]. As a ubiquitous phenomenon in nature, genomic imprinting has been broadly identified in plants [3], animals [4, 5] and humans [6, 7].

The role of genomic imprinting in shaping an organism's development has been unanimously recognized [8–10]. The imprinting effect on traits of interest can be characterized by different types. When the paternal allele at a gene is expressed and the maternal allele is inactivated, this feature of imprinting is referred to as paternal imprinting. Maternal imprinting is defined similarly. Genomic imprinting has been traditionally viewed as a mono-allelic expression with complete maternal or

Corresponding author. Yuehua Cui. Department of Statistics and Probability, Michigan State University, East Lansing, MI 48864, USA. Tel: +517-432-7098; Fax: +517-432-1405; E-mail: cui@stt.msu.edu

Runqing Yang is a Professor in the School of Agriculture and Biology, Shanghai Jiaotong University and also a Professor in the College of Animal Science and Veterinary Medicine, Heilongjiang Bayi Agricultural University. He is interested in developing Bayesian methods in QTL mapping and genetic association studies.

Xin Wang is currently a PhD student in the Bioinformatics Research Center at North Carolina State University.

Yuehua Cui is an Associate Professor in the Department of Statistics and Probability at Michigan State University. He obtained his PhD in Statistics at the University of Florida in 2005. His research focuses on developing novel statistical and computational methods to disentangle the genetic basis of complex traits by integrating various tools and sources in genetics, molecular biology, statistics, mathematics and computer sciences.

paternal silence. The definition has been revised by the inclusion of partial imprinting which signifies the different levels of expression for alleles inherited from different parents [11, 12]. Note that these classifications are all based on the additive effect of an imprinting locus. Often imprinting can cause change of interactions among alleles. Cheverud *et al.* [13] recently illustrated a scheme for characterizing the potential diversity of imprinting patterns, in which imprinting patterns are classified as either parental expression (paternal or maternal) or dominance (bipolar and polar). To date this is the most complete classification list for genomic imprinting.

Recent studies have shown the power of genetic mapping in the identification of epigenetic modification of imprinted genes or imprinted quantitative trait loci (iQTLs) on complex traits such as the variance component methods for family based pedigree data in human linkage analysis [14–16]; the variance component methods for experimental crosses [17, 18]; the regression-based approaches for controlled crosses between outbreed parents [19, 20] and between inbreed lines [21–23]. In a regular iQTL mapping study, two sex-specific reciprocal heterozygotes (e.g. $A_M a_F$ and $a_M A_F$) are not fully informative or distinguishable. However, as shown by Cui *et al.* [21], the information about sex-specific differences in recombination fraction can be used to infer the imprinting effect of an iQTL.

Most imprinted genes play important roles in controlling embryonic and post-natal growth and development in mammals [8–10]. As a highly complex process, genomic imprinting is involved in a number of growth axes operating coordinately at different development stages [24], and shows time-dependent effect during development [25]. The unbalanced expression of an imprinted gene that occurs during a development stage challenges the traditional paradigm of inheritance and mapping methods. We argue that traditional methods, by treating a trait measured at a certain developmental stage as mapping subject, without considering the correlation information at different developmental stages, are less powerful in dissecting dynamic iQTL effects. Cui *et al.* [26] recently proposed a functional iQTL mapping framework underlying developmental characteristics which incorporates a mathematical function that best describes a developmental feature into an iQTL mapping framework. Such an approach can estimate and test time-specific imprinting effect at specific developmental stages, and displays several merits over traditional iQTL mapping methods.

Current mapping procedures for iQTL inference are all based on single iQTL models, estimating and testing one locus at a time without considering the effects of other iQTLs. When multiple iQTLs are presented in the genome, such approaches are less efficient under the likelihood-based framework [27]. For a dynamic trait, the number of parameters being estimated is several folds larger than those for a univariate trait. In our previous QTL mapping model, we demonstrated that a Bayesian mapping method can handle this issue well with high computational efficiency [28]. In this study, we unify the two endeavors, Bayesian mapping of developmental traits and iQTL inference, into a unified framework called Bayesian functional multiple iQTL mapping (Bafmim). We propose an efficient Bayesian model selection strategy for multiple iQTL inference for developmental traits. The inference for the number, position and effect of multiple iQTLs as well as for different imprinting patterns is provided. The statistical behavior of the proposed method is illustrated by simulation studies. The utility of the method is shown by applying to a real data set. A total of six iQTLs are identified with Bafmim, among which two were missed by the likelihood-based method. The proposed approach has great implications in understanding the function of imprinted genes governing developmental characteristics.

STATISTICAL METHOD

The imprinting model

In a mapping population, assume that there are four distinguishable genotypes, denoted by $Q_M Q_P$, $Q_M q_P$, $q_M Q_P$ and $q_M q_P$, at each locus where the subscripts M and P refer to an allele inherited from the mother and father, respectively. A set of codominant molecular markers can be genotyped and phenotypes for a developmental trait which is measured at m time points on n individuals. In general, the additive effect a is defined as half of the phenotypic difference between two homozygotes; the dominance effect d, is defined as the difference between the joint mean of both heterozygotes and the mean of both homozygotes; and the imprinting effect i, is defined as the difference between both heterozygotes [19, 29]. Following the definitions for the genetic parameters, an imprinting model for a phenotype measured for

individual k at time t, denoted as $y_k(t)$, can be formulated as

$$y_k(t) = \mu(t) + \sum_{j=1}^{q} [c_{kj}a_j(t) + z_{kj}d_j(t) + s_{kj}i_j(t)] + \xi_k(t) + \varepsilon_k(t) \quad (1)$$

where q is the number of potential (i)QTLs in the genome; $\mu(t)$ is the population mean at time t; $\alpha_j(t)$, $d_j(t)$ and $i_j(t)$ for $j = 1, 2, \ldots, q$ are the additive, dominance and imprinting effects of the j-th iQTL at time point t; $\xi_k(t)$ $(k = 1, 2, \ldots, n)$ is an individual-specific time-dependent random environmental term modeled as a mean-zero Gaussian process, i.e. $N(0, \sigma^2_\xi(t))$ and $e_k(t)$ is a random environmental error assumed to be normally distributed with mean zero and variance σ^2. Note that c_{kj}, z_{kj} and s_{kj} are genotype-specific indicator variables related to genetic effects $\alpha_j(t)$, $d_j(t)$ and $i_j(t)$, which are defined by Mantey *et al.* [30] as

$$c_{kj} = \begin{cases} +1 \\ 0 \\ 0 \\ -1 \end{cases}, \; z_{kj} = \begin{cases} 0 \\ +1 \\ +1 \\ 0 \end{cases} \text{ and } \; s_{kj} = \begin{cases} 0 & \text{for} & Q_M Q_P \\ +1 & \text{for} & Q_M q_P \\ -1 & \text{for} & q_M Q_P \\ 0 & \text{for} & q_M q_P \end{cases}$$

Note that in practical mapping, unless QTLs are located exactly at the marker position (i.e. the markers and the QTLs overlap), these genotype-specific indicator variables are unknown and need to be inferred from flanking markers. We use Legendre polynomial of order r to fit changing trajectories of the population mean and the effects of each iQTL [31, 32]. Let $\boldsymbol{\psi}(t) = (\psi_0(t), \ldots, \psi_r(t))^T$ be the basis of the Legendre polynomial with order r and have that $\mu(t) = \boldsymbol{\psi}(t)\boldsymbol{\mu}$, $a_j(t) = \boldsymbol{\psi}(t)\mathbf{a}_j$, $d_j(t) = \boldsymbol{\psi}(t)\mathbf{d}_j$, $i_j(t) = \boldsymbol{\psi}(t)\mathbf{i}_j$ and $\xi_k(t) = \boldsymbol{\psi}(t)\boldsymbol{\xi}_k$, where each one of $\boldsymbol{\mu}$, $\mathbf{a}_j$, $\mathbf{d}_j$, $\mathbf{i}_j$ and $\boldsymbol{\xi}_k$ is a vector of $r+1$ dimensions. Model (1) can be then rewritten as

$$\mathbf{y}_k(t) = \boldsymbol{\psi}^{\mathrm{T}}(t)\boldsymbol{\mu} + \sum_{j=1}^{q} [z_{kj}\boldsymbol{\psi}^{\mathrm{T}}(t)\mathbf{a}_j + w_{kj}\boldsymbol{\psi}^{\mathrm{T}}(t)\mathbf{d}_j + s_{kj}\boldsymbol{\psi}^{\mathrm{T}}(t)\mathbf{i}_j] + \boldsymbol{\psi}^{\mathrm{T}}(t)\boldsymbol{\xi}_k + e_k(t) \quad (2)$$

Where $\boldsymbol{\xi}_k$ is a vector of regression coefficients for random effects, which is assumed to be multivariate normal with mean zero and an $(r+1) \times (r+1)$ positive definite covariance matrix $\boldsymbol{\Sigma}$.

For simplicity, we assume that each individual is measured at m time points and the time points are common for all individuals. Let $\mathbf{y}_k = [y_k(t_0) y_k(t_1) \ldots y_k(t_m)]^T$ be an $(m+1) \times 1$ column vector for the repeated measurements of a developmental trait and define $\boldsymbol{\psi} = [\psi^T(t_0)\psi^T(t_1)\ldots\psi^T(t_m)]$ as a $(r+1)\times(m+1)$ matrix. In matrix notation, Model (2) becomes

$$\mathbf{y}_k = \boldsymbol{\psi}^{\mathrm{T}}\boldsymbol{\mu} + \sum_{j=1}^{q} [z_{kj}\boldsymbol{\psi}^{\mathrm{T}}\mathbf{a}_j + w_{kj}\boldsymbol{\psi}^{\mathrm{T}}\mathbf{d}_j + s_{kj}\boldsymbol{\psi}^{\mathrm{T}}\mathbf{i}_j] + \boldsymbol{\psi}^{\mathrm{T}}\boldsymbol{\xi}_k + \mathbf{e}_k \quad (3)$$

where $\mathbf{e}_k = [e_k(t_0)\ e_k(t_1)\ \ldots\ e_k(t_m)]^T$ is an $(m+1) \times 1$ vector for the environmental errors, distributed as $\mathbf{e}_k \sim N(0, \mathbf{I}\sigma^2)$ with $\mathbf{I}$ being an $(m+1) \times (m+1)$ identity matrix.

Bayesian model selection for genetic parameters

Model (3) is a mixed-effect model where population mean and genetic effects are fixed and the time-dependent environmental effect is random. Also note that Model (3) is not a regular linear mixed-effect model, since the number of independent variables for the fixed effects and the associated indicator variables are unknown due to unknown number of QTLs. In principle, QTLs can be distributed anywhere in the genome, and hence any genomic positions can be potential QTL locations. Thus, we approximate them for all possible iQTLs by partitioning the entire genome into evenly spaced loci by 1 or 2 cM, covering all observed markers and additional loci between flanking markers. The expected values of the indicator variables at each locus can be calculated for an F_2 mapping population based on the conditional probabilities of a QTL genotype given on two flanking markers as follows:

$$E(z) = \pi_{QQ} - \pi_{qq},$$
$$E(w) = \pi_{Qq} + \pi_{qQ} \text{ and } E(s) = \pi_{Qq} - \pi_{qQ}$$

with π_{QQ}, π_{Qq}, π_{qQ} and π_{qq} being the conditional probabilities of a tested QTL genotype $Q_M Q_p$, $Q_M q_p$, $q_M Q_p$ or $q_M q_p$ given on two flanking markers (see Table 1 in Cui *et al.* [21] for details).

For a supersaturated model where each genomic location could contain a potential iQTL, it is almost impossible to estimate a huge number of genetic effects. So we preset an upper bound on the number of iQTLs in the model [52]. The upper bound should be larger than the potential number of detectable iQTLs in a given data set. Given an upper bound on the number of iQTLs (L), these iQTLs can be drawn from densely spaced loci over the genome. Even with a moderate number of upper bound, there are many genetic effects being

estimated in Model (3). In order to infer the existence of these effects, we introduce a random binary variable γ to indicate which genetic effects should be included in or excluded from the model, corresponding to $\gamma=1$ or $\gamma=0$ [33–35]. Model (3) then becomes

$$\mathbf{y}_k = \boldsymbol{\psi}^{\mathrm{T}}\boldsymbol{\mu} + \sum_{j=1}^{L}\left[\gamma_{aj}z_{kj}\boldsymbol{\psi}^{\mathrm{T}}\mathbf{a}_j + \gamma_{dj}w_{kj}\boldsymbol{\psi}^{\mathrm{T}}\mathbf{d}_j + \gamma_{ij}s_{kj}\boldsymbol{\psi}^{\mathrm{T}}\mathbf{i}_j\right] + \boldsymbol{\psi}^{\mathrm{T}}\boldsymbol{\xi}_k + \mathbf{e}_k, \quad (4)$$

where γ_{gj} ($g=\mathbf{a}$, $\mathbf{d}$ or $\mathbf{i}$) is the indicator variable for genetic effects $\mathbf{a}_j$, $\mathbf{d}_j$ or $\mathbf{i}_j$. Within the framework of Bayesian model selection, Bayesian sampling for unknown parameters including $\boldsymbol{\mu}$, γ_{gj}, $\mathbf{a}_j$, $\mathbf{d}_j$, $\mathbf{i}_j$, $\boldsymbol{\xi}_k$, $\boldsymbol{\Sigma}$ and QTL position λ_j, is implemented with the Markov chain Monte Carlo (MCMC) algorithm. Note that the released sampling value for binary variable γ at a previous round determines which genetic effects and position of an iQTL should be drawn or estimated at the next round. This will save computing time significantly.

Likelihood function

Given unknown parameters θ, the likelihood function of model (4) can be expressed as

$$L(\theta|\mathbf{y}) = \prod_{k=1}^{n} p(\mathbf{y}_k|\theta) \propto |\mathbf{V}|^{-n(m+1)/2} exp\left[\sum_{k=1}^{n}(\mathbf{y}_k-\mathbf{U}_k)^T\mathbf{V}^{-1}(\mathbf{y}_k-\mathbf{U}_k)\right],$$

where $\mathbf{U}_k = \boldsymbol{\psi}^{\mathrm{T}}\boldsymbol{\mu} + \sum_{j=1}^{L}(\gamma_{aj}z_{kj}\boldsymbol{\psi}^{\mathrm{T}}\mathbf{a}_j + \gamma_{dj}w_{kj}\boldsymbol{\psi}^{\mathrm{T}}\mathbf{d}_j + \gamma_{ij}s_{kj}\boldsymbol{\psi}^{\mathrm{T}}\mathbf{i}_j)$ and $\mathbf{V} = \boldsymbol{\psi}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{\psi} + \mathbf{I}\sigma^2$.

Prior specification

Following Yi *et al.*, we take upper bound of iQTL number L as $l_0 + 3\sqrt{l_0}$, where l_0 is the prior expected number of iQTL loci and is determined according to initial investigations with traditional methods [26]. The binary indicator γ is assumed to have an independent prior $p(\gamma_g) = \prod w_g^{\gamma_g}(1-w_g)^{(1-\gamma_g)}$, where $w_g = 1-\left[1-\frac{l_0}{L}\right]^{1/3}$ is the prior inclusion probability for certain QTL effect [52]. Priors on iQTL positions are assumed to be independent and uniformly distributed over the entire genome.

The prior for the population mean $\boldsymbol{\mu}$ is assumed to be $N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$. We can empirically set $\boldsymbol{\mu}_0 = \bar{\mathbf{b}} = \frac{1}{n}\sum_{k=1}^{n}\mathbf{b}_k$ and $\boldsymbol{\Sigma}_0 = \frac{1}{n-1}\sum_{k=1}^{n}(\mathbf{b}_k-\bar{\mathbf{b}})(\mathbf{b}_k-\bar{\mathbf{b}})^T$, where $\mathbf{b}_k = (\boldsymbol{\psi}^T\boldsymbol{\psi})^{-1}\boldsymbol{\psi}^T\mathbf{y}_k$ is a vector of regression coefficients obtained by fitting individual dynamic trajectory.

We propose the following hierarchical mixture prior for each additive genetic effect, $\boldsymbol{\beta}_j \sim N_{p+1}(0, \boldsymbol{\Sigma}_j)$ with $\boldsymbol{\Sigma}_j = \gamma_j c(\boldsymbol{\psi}^T\mathbf{V}^{-1}\boldsymbol{\psi}\sum_{k=1}^{n}x_{kj}^2)^{-1}$ and c being taken to n such that the prior variance of each fixed effect stays approximately the same as n increases.

The random effects $\boldsymbol{\xi}_k$ are assumed to have an independent multivariate normal distribution. That is, $\boldsymbol{\xi}_k \sim N_{r+1}(0,\mathbf{S}_a)$ with the hyperparameter $\mathbf{S}_a$ being an $(r+1)\times(r+1)$ matrix. An inverse Wishart prior is chosen for the covariance matrix of regression coefficients for random environmental effect, denoted by $\boldsymbol{\Sigma} \sim IW(\nu_a, \nu_a\mathbf{S}_a)$ with ν_a being hyperparameter. The residual variance is assumed to have a scaled inverse χ^2 distribution, i.e. $\sigma^2 \sim IC(\nu_e, \frac{1}{\nu_e s_e})$ where ν_e and s_e are hyperparameters. $p(\lambda_j) = \frac{1}{l_j}$ for $j=1,2,\ldots,L$, where l_j is the distance between the two neighboring QTLs [36–38].

MCMC sampling

A joint posterior density can be formed by multiplying priors of all unknown parameters and the likelihood function above. The joint posterior distribution is analytically intractable, hence we need to derive the 'conditional posterior' for each unknown parameter from the joint posterior density. The details about the derivation can be found in the Appendix 1. A MCMC methods such as Gibbs sampler [39] and Metropolis–Hastings algorithm [40, 41] are applied to sample each parameter conditional on all other parameters. Prior to implementing MCMC sampling, weset an upper bound for the number of QTLs and distribute these QTLs on the genome evenly and initialize all variables with some initial values or values sampled from their prior distributions (see the Appendix 1 for details about prior specification). The MCMC sampling procedure for unknown parameters is summarized as follows:

(i) Update population mean $\boldsymbol{\mu}$ by sampling from a normal distribution.
(ii) Update the genetic effects $\boldsymbol{\beta}_j$ corresponding to $\gamma_j=1$ by drawing from a normal distribution.
(iii) Update the binary indicators γ_j by adopting an efficient Metropolis–Hastings algorithm [52].

(iv) Update individual-specific regression coefficients for random environmental effects ξ_k by sampling from a normal distribution.
(v) Update the covariance matrix $\mathbf{\Sigma}$ of ξ_k by drawing from an inverse Wishart distribution.
(vi) Update the residual variance σ^2 by sampling from an inverse χ^2 distribution.
(vii) Update joint QTL positions λ by adopting the Metropolis–Hastings algorithm.
(viii) Repeat steps (i) –(viii) until the Markov chain reaches a desirable length.

Post-MCMC analysis

Post-MCMC analysis includes the monitoring of the mixing behavior and convergence rates of the MCMC algorithm, and the assessment of characteristics of the imprinting genetic architecture. The former can be checked by visually inspecting trace plots of the sample values of scalar quantities of interest or formal diagnostic methods provided in the package R/coda [42]. The latter can use model averaging which accounts for model uncertainty and average over possible models weighted by their posterior probabilities [43–45]. The posterior inclusion probability for each locus is estimated as its frequency in the posterior samples. Bayes factor (BF) is used as a measurement for inclusion against exclusion at each iQTL locus [46]. Generally, a threshold of BF is empirically determined as 3, or $2\ln BF = 2.1$, for declaring statistical significance of an (i)QTL.

Bayesian inference for imprinting mode of action

Generally speaking, (i)QTLs detected with the above Bayesian algorithm cannot be declared as iQTLs until we do further imprinting inference. After an (i)QTL is detected, we can adopt the idea of a Bayes factor to infer statistical significance for its imprinting effect with the form:

$$BF = \frac{p_g}{1-p_g} \cdot \frac{1-p}{p}$$

where p is a prior probability and p_g is a posterior probability for a certain genetic effect, which is calculated as the proportion of samples in which $\gamma_g = 1$ in MCMC sampling rounds. If the BF is >3 (or $2\ln BF > 2.1$) for the imprinting effect i, then the detected iQTL can be claimed as a true iQTL, otherwise as a Mendelian QTL.

Following the definition of imprinting types and the corresponding null hypothesis [13, 47, 48], we classify imprinting patterns as parental imprinting, i.e. $\mathbf{a} = \pm\mathbf{i}$ and $\mathbf{d} = 0$ which includes paternal ($\mathbf{a} = \mathbf{i}$) and maternal ($\mathbf{a} = -\mathbf{i}$) imprinting subtypes; and dominance imprinting with $\mathbf{a} = 0$ but $\mathbf{i} \neq 0$, which can be further distinguished as bipolar imprinting in which $\mathbf{d} = 0$ and $\mathbf{i} \neq 0$ and polar imprinting in which $\mathbf{d} = \pm\mathbf{i}$. Since the imprinting pattern for the detected QTL depends on whether genetic effect $\mathbf{a}$ or $\mathbf{d}$ equals imprinting effect i, it can also be statistically inferred through the Bayes factor statistic. The Bayes factor can be simplified as the ratio of posterior probabilities for the genetic effects being compared, due to same prior probability for each genetic effect. Imprinting patterns, hypotheses and corresponding statistical criteria for the iQTLs are detailed in [48].

SIMULATION STUDIES

We conducted simulation studies to evaluate the performance of the proposed Bayesian functional multiple iQTL mapping approach. A genome consisting of one single large chromosome of 600 cM was simulated covering 61 evenly spaced markers. The growth pattern of a dynamic trait was assumed to be controlled by one QTL inherited in a Mendelian fashion and four iQTLs with their imprinting patterns, positions and effects listed in Table 1. The order of the Legendre polynomial that generates the developmental trajectory was assumed to be $r = 3$. The trajectory for each (i)QTL genetic effect is plotted in Figure 1. We simulated a dynamic trait measured at eleven time points assuming different sample sizes ($n = 250$, 500) with inbreed F_2 individuals. The marker and QTL genotypes in the F_2 family were generated by mimicking sex-specific recombination fractions (see Cui *et al.* [21] for more details). The population mean and the individual-specific environmental error covariance matrix were set the same as described in Yang and Xu [31] and the residual variance was set as four.

In Bayesian analyses for simulated data, we set the prior number of main-effect QTLs as 3. The upper bound of the number of QTLs was then equal to $L = 3 + 3\sqrt{3} = 8$. The actual values for the hyperparameters used here mimic the results obtained in real data analyses (see the real data analysis section). The initial values of all variables were sampled from their prior distributions. The MCMC was run for

Table 1: The imprinting type and parameters (regression coefficients of Legendre polynomials) of iQTLs used in simulation studies

QTL no.	Imprinting type	Position	a_0	a_1	a_2	a_3	d_0	d_1	d_2	d_3	i_0	i_1	i_2	i_3
1	Mendelian	23	1.82	−0.80	−1.20	−0.80								
2	Paternal	148	0.00	1.65	2.52	1.20					0.00	1.65	2.52	1.20
3	Maternal	256	2.55	1.36	−2.02	−1.27					−2.55	−1.36	2.02	1.27
4	Bipolar	332									2.94	0.00	1.08	1.72
5	Polar	522					2.00	−1.25	0.00	−1.28	2.00	−1.25	0.00	−1.28

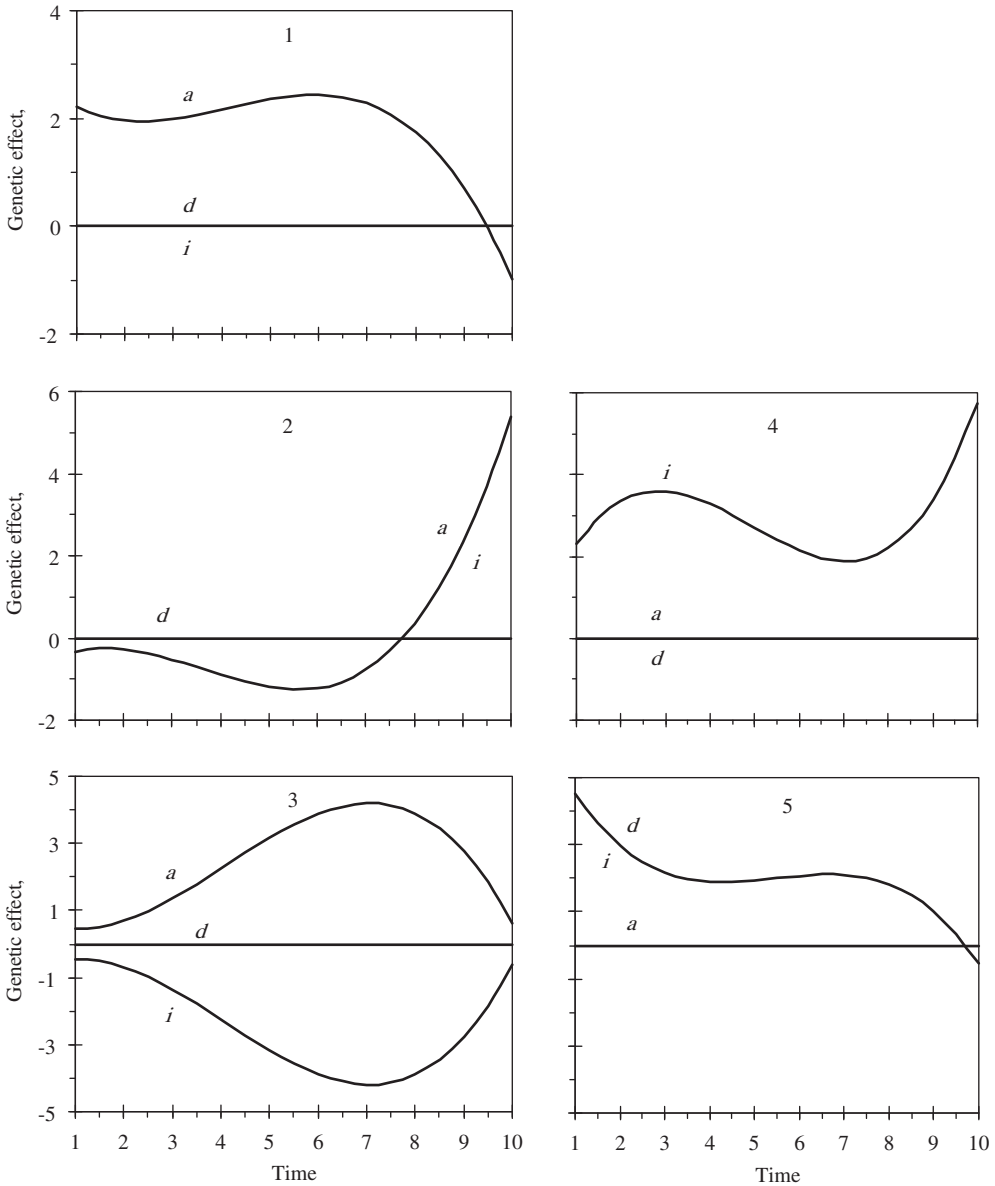

Figure 1: The trajectories for genetic effects of the simulated (i)QTLs: (1) $d = i = 0$, (2) $a = i$, $d = 0$, (3) $a = -i$, $d = 0$, (4) $a = d = 0$ and (5) $d = i$, $a = 0$. In each plot, time is the original time measurement ranging from 1 to 10 weeks as in real data.

Table 2: Mean estimates and SDs (in parentheses) of iQTL regression effects with polynomial residuals

Sample size	QTL no.	a_0	a_1	a_2	a_3	d_0	d_1	d_2	d_3	i_0	i_1	i_2	i_3
250	1	1.77 (0.23)	−0.95 (0.35)	−1.17 (0.40)	−0.92 (0.33)								
	2	−0.07 (0.24)	1.68 (0.36)	2.47 (0.45)	1.16 (0.39)					−0.03 (0.31)	1.63 (0.28)	2.42 (0.49)	1.17 (0.33)
	3	2.58 (0.43)	1.42 (0.30)	−1.97 (0.41)	−1.30 (0.43)					−2.53 (0.44)	−1.39 (0.35)	1.99 (0.48)	1.32 (0.51)
	4									3.01 (0.49)	0.04 (0.18)	0.98 (0.33)	1.79 (0.31)
	5					1.96 (0.32)	−1.31 (0.31)	0.05 (0.25)	−1.19 (0.40)	1.91 (0.31)	−1.36 (0.40)	0.02 (0.21)	−1.15 (0.47)
500	1	1.86 (0.21)	−0.85 (0.33)	−1.18 (0.34)	−0.84 (0.25)								
	2	0.02 (0.19)	1.67 (0.22)	2.49 (0.38)	1.18 (0.30)					0.01 (0.22)	1.64 (0.23)	2.45 (0.34)	1.15 (0.28)
	3	2.56 (0.25)	1.40 (0.24)	−2.04 (0.32)	−1.30 (0.29)					−2.50 (0.23)	−1.44 (0.28)	2.00 (0.35)	1.34 (0.26)
	4									2.96 (0.28)	0.01 (0.11)	1.07 (0.24)	1.74 (0.25)
	5					2.03 (0.19)	−1.27 (0.24)	0.03 (0.18)	−1.25 (0.24)	1.97 (0.21)	−1.16 (0.20)	0.01 (0.15)	−1.30 (0.20)

10 000 cycles as a burn-in period (deleted) and then for an additional 150 000 cycles after the burn-in. The chain was then thinned to reduce serial correlation by saving one observation in every 50 cycles. The posterior sample contained 3000 observations for post-MCMC analysis. Note that the length of the burn-in was judged by visually inspecting the plots of some posterior samples across rounds and was set to enough cycles to ensure MCMC convergence. The simulation experiment was replicated 100 times in order to evaluate the statistical power of our method.

Table 2 shows the mean estimates as well as their SD (in parenthesis) for the parameters given in Table 1. The relative statistical power to detect each QTL is also listed. Overall, the Bayesian mapping approach is able to estimate the regression effects of the iQTLs with reasonable precision. All four iQTL positions can be accurately estimated with high precision. As we expected, increasing sample size always leads to small bias, increased precision of parameter estimation, and high mapping power. For example, the mapping power for QTL 1 increased from 70 to 85% when sample size was increased from 250 to 500. Even with small sample size ($n = 250$), the QTL position can also be estimated with high precision. In addition, we can also accurately infer the imprinting pattern of the detected locus using Bayes factor (data not shown). These indicate the power of Bayesian mapping for multiple iQTL inference. The simulated data sets were also analyzed by likelihood-based functional iQTL mapping [26]. We found that the likelihood-based functional mapping provides less accurate iQTL parameter estimates, cannot infer the given imprinting pattern well, and performs poor with relatively low statistical power of iQTL detection than the proposed Bayesian method (Table 3). Moreover, the likelihood-based functional mapping has a high false positive rate (FPR) in iQTLs detection with 12% for sample size of 250 and 9% for sample size of 500, while the Bayesian method gives lower FPR. As the maximum likelihood-based method is not efficient in handling multiple iQTLs, this might explain the performance difference of the two approaches. The simulation results indicate the robustness of the proposed method in multiple iQTL detection for dynamic traits with moderate sample size.

Under the same simulation scenarios, replacing both time-dependent permanent environmental

Table 3: Mean estimates and SDs (in parentheses) of iQTL positions and statistical power of iQTL detection

Sample size	Method		QTL no.				
			1	2	3	4	5
250	Bayesian	Position	21.3 (3.6)	147.3 (4.2)	258.7 (5.4)	334.1 (5.8)	524.7 (6.1)
		Power (%)	84	72	80	90	84
	Likelihood	Position	24.3 (5.3)	152.1 (8.1)	262.3 (9.2)	338.0 (11.8)	526.9 (13.5)
		Power (%)	71	60	68	77	70
500	Bayesian	Position	22.9 (3.3)	148.5 (3.7)	257.9 (4.1)	333.7 (4.6)	520.5 (5.0)
		Power (%)	100	96	100	100	98
	Likelihood	Position	23.8 (4.9)	150.9 (7.1)	261.1 (8.6)	336.6 (10.4)	526.0 (11.9)
		Power (%)	92	82	98	100	85

effects and residual errors in model (1) with AR(1) structure where autoregressive coefficient was taken as 0.6, we simulated phenotypes and fit model (1) to the simulated data sets with the Bayesian method. Table 4 lists mean estimates and SDs (in parentheses) of the given (i)QTL regression effects. As expected, the Bayesian method can powerfully identify the given (i)QTLs and estimate the given (i)QTL regression effects reasonably well based on model (1), indicating the flexibility of our model in fitting different time-dependent residuals.

We also simulated data assuming the true dynamics of the repeated measurement following a logistic regression curve for genotypic effects of the five simulated QTLs. The simulated parameters for the logistic function are obtained by fitting changes in genotypic means generated from the Legendre polynomials (Supplementary Table 1). Residuals are modeled by AR(1) structure with autoregressive coefficient of 0.6. Estimates of QTL positions and powers to detect QTLs are listed in Supplementary Table 2. The results indicate that the proposed method performs better than the likelihood-based method.

REAL DATA ANALYSIS

We illustrated the application of our proposed approach by reanalyzing a mouse body weight growth data set with an F_2 mating population derived from two inbreed strains, the Large (LG/J) and the Small (SM/J). A total of 502 F_2 mice were genotyped for 96 microsatellite markers located on 19 autosomal chromosomes. A linkage map of a total length 1780 cM has been constructed [49]. The body mass was measured on each mouse at 10 weekly intervals starting at Day 7. The raw weights were adjusted for the effects of each covariate due to dam, litter size at birth and parity, and sex [49]. The data set has been analyzed by Cui *et al.* [26] with likelihood-based functional mapping.

By fitting the mean change of weight growth over age, we chose the Legendre polynomial of order 3 as the base model to describe the changing trajectory for each component except for residuals, described in Model (1). The female-to-male recombination rate of 1.25:1 is used to estimate conditional probabilities for the four QTL genotypes [21]. The expected number of main-effect QTLs was set as $l_0 = 4$ according to the results by Cui *et al.* [26] and the upper bound of the number of QTLs was then calculated as $L = 4 + 3\sqrt{4} = 10$. Thus, the prior inclusion probability for QTL effects was determined to be $1 - \left[1 - \frac{4}{10}\right]^{\frac{1}{3}} = 0.156$. The actual values for the hyperparameters were set as $S_h = S_e = 0.5I$, $\nu_h = r + 1$ and $\nu_e = 0$. The initial values of all variables were sampled from their prior distributions. The MCMC was run 200 000 cycles after the burn-in period of 10 000 cycles. The analysis of the real data set with our MATLAB code took ~3 h on an Intel Core 2 PC with a 2 GHz processor and 4 GB RAM.

Figure 2 plots the profiles of 2logBF obtained with the Bayesian model selection. The top panel shows that there are six peaks for 2logBFs exceeding the horizontal reference line with an empirical critical value 2.1, indicating that six iQTLs are detected on chromosomes 2, 4, 6, 7, 10 and 15. All iQTLs detected show significant imprinting effect, as their relative 2logBFs are >2.1 for imprinting effects (the bottom panel in Figure 2). Table 5 tabulates the position on each chromosome and the estimated effects (additive, dominance and imprinting) for the six detected iQTLs. The estimated regression coefficients for each iQTL have no biological meaning, but they can be used to predict the effects of an iQTL at a given time point by substituting time

Table 4: Mean estimates and SDs (in parentheses) of iQTL regression effects with AR(1) residuals

Sample size	QTL no.	a_0	a_1	a_2	a_3	d_0	d_1	d_2	d_3	i_0	i_1	i_2	i_3
250	1	1.68 (0.43)	−1.05 (0.55)	−1.07 (0.61)	−1.03 (0.53)								
	2	−0.18 (0.44)	1.88 (0.66)	2.30 (0.75)	1.07 (0.61)					−0.10 (0.54)	1.50 (0.47)	2.30 (0.70)	1.07 (0.55)
	3	2.42 (0.73)	1.53 (0.55)	−1.89 (0.73)	−1.45 (0.71)					−2.39 (0.70)	−1.25 (0.63)	1.89 (0.80)	1.44 (0.87)
	4									3.09 (0.68)	0.09 (0.28)	0.88 (0.55)	1.88 (0.54)
	5					1.86 (0.53)	−1.42 (0.53)	0.09 (0.46)	−1.07 (0.71)	1.80 (0.53)	−1.44 (0.72)	−0.86 (0.40)	−1.05 (0.83)
500	1	1.74 (0.38)	−0.95 (0.43)	−1.12 (0.53)	−0.93 (0.41)					−0.04 (0.42)	1.54 (0.41)	2.37 (0.60)	1.15 (0.46)
	2	0.06 (0.33)	1.76 (0.42)	2.42 (0.50)	1.12 (0.52)					−2.44 (0.56)	−1.30 (0.39)	1.93 (0.58)	1.36 (0.49)
	3	2.47 (0.45)	1.48 (0.37)	−1.96 (0.44)	−1.37 (0.43)					3.02 (0.46)	0.06 (0.19)	0.97 (0.43)	1.62 (0.43)
	4					2.07 (0.33)	−1.30 (0.44)	0.05 (0.32)	−1.17 (0.43)	1.90 (0.39)	−1.08 (0.35)	0.04 (0.26)	−1.18 (0.33)
	5												

into the Legendre polynomial with the regression coefficients. Figure 3 shows the dynamic trajectory of the genetic effects over time. Among the six iQTLs, the main effect (i.e. *a* or *d*) shows a stronger effect on mouse growth than the imprinting effect, which makes biological sense as imprinting effect generally plays a modifying role.

We further evaluated the imprinting pattern of the six detected iQTLs. Two of them (on chromosomes 4 and 10) show polar imprinting, two (on chromosomes 2 and 6) show maternal imprinting and two (on chromosomes 7 and 15) show paternal imprinting, based on the results obtained from the significance analysis given in Table 6. Compared to the likelihood-based method [26], the Bayesian method identified two more QTLs (on chromosomes 2 and 4).

DISCUSSION

The epigenetic phenomenon in genomic imprinting has been constantly challenging and revising the traditional paradigm of inheritance. The inheritable property of imprinting provides clues for complicated genetic disorders [6]. In the meantime, it also brings challenges for statistical modeling and mapping. We developed a Bayesian model selection method to identifying multiple iQTLs for developmental traits illustrated in an F_2 mating population. Extensions to other mating designs such as a reciprocal backcross design [50], is straightforward. The Bayesian method has shown its relative merits in handling multiple QTLs partly due to its flexibility to handle a large parameter space [28, 31, 51, 52]. Both simulation and real data analysis indicate the power and relative advantage of the proposed Bayesian multiple iQTL mapping for dynamics traits compared to the likelihood-based mapping method. In addition, we proposed several inference procedures to infer different imprinting types, which have not been discussed in the likelihood-based iQTL mapping study. The identified iQTLs as well as their imprinting property provide valuable information for further experimental verification.

The current method is developed specifically for longitudinal or functional traits. Recently, Hayashi and Awata [15] proposed a Bayesian mapping approach that can simultaneously map multiple QTLs, and further discriminating Mendelian and imprinting expressions of a QTL. Although the approach shows improvement in iQTL detection,

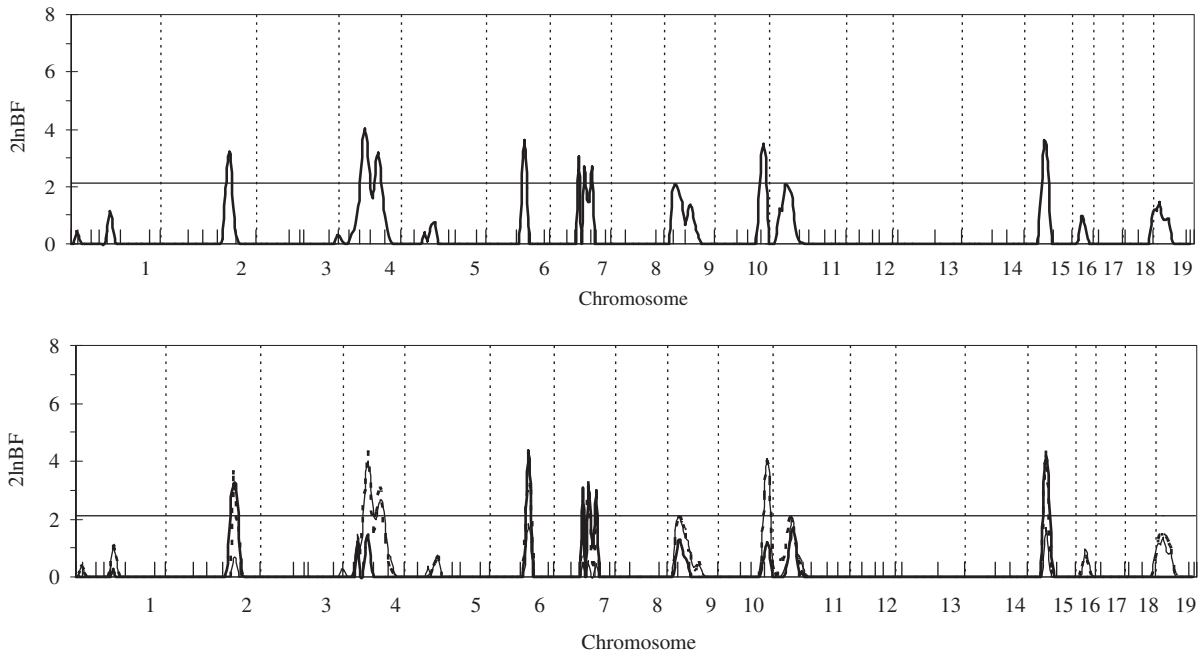

Figure 2: The profile plot of 2lnBF for genome-wide iQTL scan (top) and for various genetic effects inference of iQTLs (bottom) obtained with the Bafmim algorithm in mouse weight growth. In both figures, linkage groups are separated by the vertical dotted lines and marker positions are indicated by the ticks on the horizontal axis. The horizontal reference line is the empirical critical value 2.1 for 2lnBF. In the bottom panel, the thick solid, thin solid and dashed curves represent additive, dominance and imprinting effects, respectively.

Table 5: Estimates of iQTL positions and regression coefficients for body weight growth in mice

QTL	Position (Chr.-cM)	a_0	a_1	a_2	a_3	d_0	d_1	d_2	d_3	i_0	i_1	i_2	i_3
1	2–104.3	0.254	0.633	0.166	0.033					0.074	0.180	0.061	0.013
2	4–38.1					−0.070	−0.360	−0.132	0.036	−0.050	−0.118	−0.028	0.028
3	6–73.2	0.425	0.525	0.075	0.055					0.162	0.210	0.060	0.032
4	7–63.1	0.787	0.851	−0.044	0.103					0.185	0.210	0.016	0.141
5	10–72.7					0.027	−0.165	0.018	−0.018	0.023	−0.061	0.004	−0.006
6	15–12.7	0.896	0.311	−0.314	0.062					0.117	0.071	−0.035	0.034

it is limited by a number of facts. For example, drawing number of QTLs with a reversible-jump MCMC procedure may have low convergence efficiency. Moreover, the method is developed for univariate traits and ignores the dynamics of gene effects. In an earlier article, Yang *et al.* [53] developed a Bayesian multiple iQTL mapping method for univariate traits. The current work is an extension of our previous work, but is more challenging in modeling the dynamics of genetic effects.

Our method assumes a maximum number of detectable iQTLs and introduces latent binary variables to indicate which main effects for a putative iQTL should be included or excluded from the model. Compared to the likelihood-based method [26], it allows MCMC sampling for iQTL parameters to carry out in the reduced model space, and thus enhances the computational efficiency of Bayesian multiple iQTL mapping with many parameters. Computationally, it also outperforms the likelihood-based method. In addition, it facilitates statistical inference for imprinting patterns of the detected iQTLs with appropriately defined Bayes factors. In the real data analysis, we took the dynamic course of a developmental trait as the mapping subject, and identified more iQTLs altering the developmental trajectory than separately performing iQTL mapping at each time point [21].

The key issue for iQTL mapping of developmental traits is to choose appropriate submodels for imprinting inference. In the likelihood-based mapping framework, Cui *et al.* [26] proposed to describe

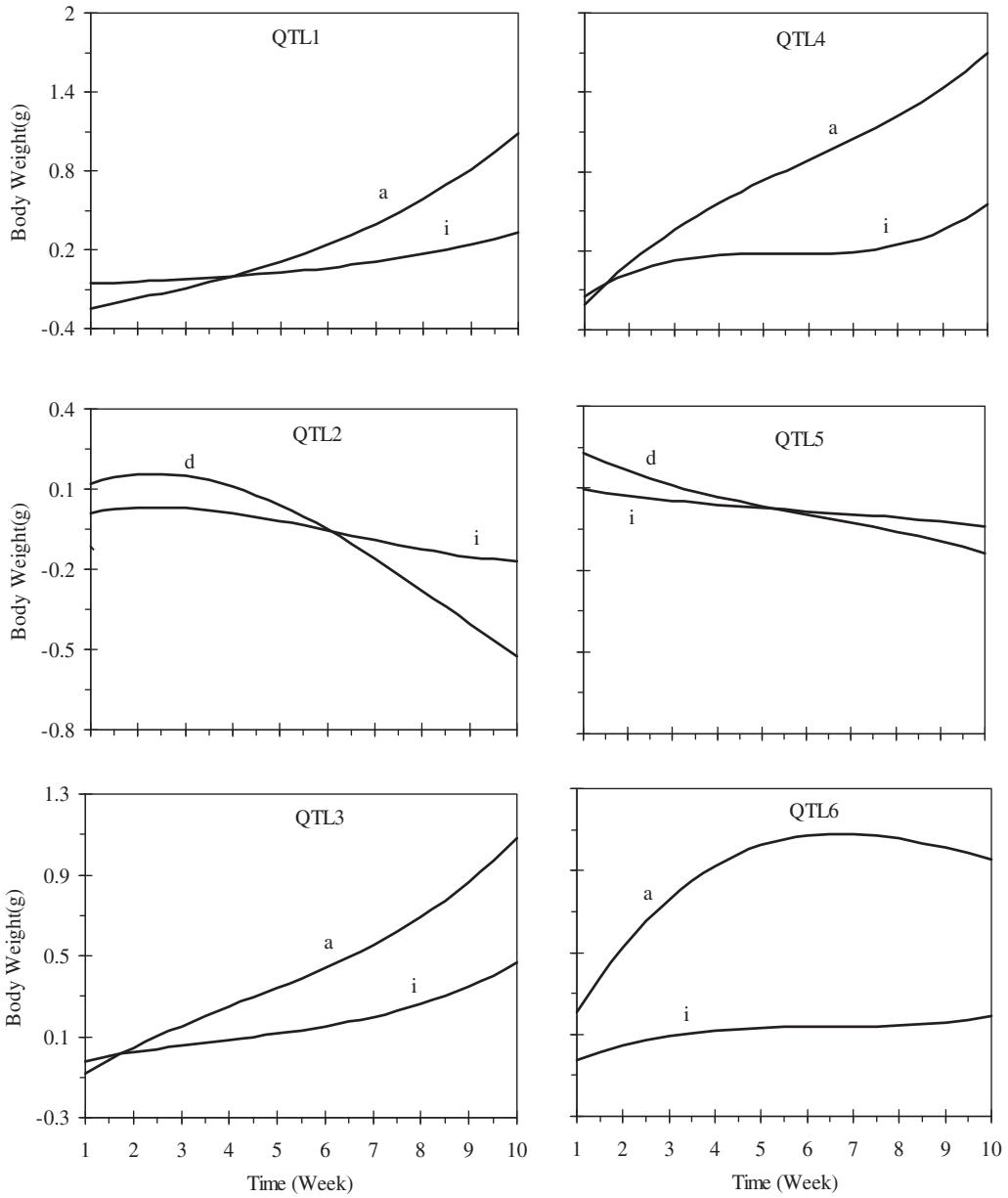

Figure 3: The trajectories for different genetic effects (*a*, *i* and *d*) of the detected iQTLs for body weight growth in mice.

Table 6: List of 2lnBFs for genetic effects and imprinting types of the detected iQTLs for body weight growth in mice

QTL	*a*	*d*	*i*	Imprinting type
1	3.29	0.64	2.20	Additive-Paternal
2	1.01	3.34	4.47	Dominance-Over
3	4.72	1.83	3.24	Additive-Paternal
4	2.95	0.54	2.97	Additive-Paternal
5	0.87	4.32	3.68	Dominance-Over
6	4.46	1.70	4.23	Additive-Paternal

the changes of iQTL genotypic effects by a logistic growth function. This parametric assumption is, however, difficult to implement in a multiple iQTL model for Bayesian mapping due to the non-linearity of different submodels. With the orthogonal polynomial function, it is flexible enough to capture the underlying trajectory for dynamic iQTL inference. One issue when fitting data with a polynomial function is to appropriately select optimal polynomial order. Cui *et al.* [32] proposed two methods to choose an optimal order: (i) assume the same order for different phenotypic

trajectories which can be chosen under the null hypothesis or (ii) assume different orders for different phenotypic trajectories. The second method, however, is computationally intensive. Even though one can select the polynomial order for different QTL effects, the technical challenge lies on the selection of the order for the time-dependent permanent environment effects [54]. Thus, we adopted the first method and assumed the same order to fit changes in QTL genetic effects and time-dependent environmental effects. The polynomial regression coefficients determine the shape of the population mean and each genetic effect. Certainly, an optimal mapping strategy is to choose polynomials of different orders to best model each component in the imprinting model, which we leave it for future investigation.

In functional mapping, how to model the correlation structure for repeated measurements is also a challenging problem. In most functional mapping studies, a parametric residual covariance structure such as the autoregressive model with order 1 [AR(1)] is often assumed [55]. However, it is difficult to implement a Bayesian method for this covariance function due to the difficulty to choose an appropriate prior for the autoregressive coefficient. In contrast, the covariance structure described by $\boldsymbol{\psi}^T \sum \boldsymbol{\psi} + I\sigma^2$ is more flexible than the parametric structure because we can actually choose different degrees of polynomial order to fit a covariance structure with a large degree of complexity [31]. Moreover, we can easily sample the covariance matrix $\sum$ from a closed form of marginal posterior distribution. Yap *et al.* [56] recently proposed a nonparametric method for covariance structure modeling in functional mapping. More work is needed to integrate these techniques into a Bayesian mapping framework to improve mapping power.

When unbalanced data are recorded, we can change time covariate $\boldsymbol{\psi}(t)$ or $\boldsymbol{\psi}$ to $\boldsymbol{\psi}_k(t)$ or $\boldsymbol{\psi}_k$ depending on the kth individual, to make the method more general. Also, the multiple iQTL model for developmental traits proposed herein could be treated as a general form of the model for analyzing genomic imprinting of a quantitative trait. For instance, let $\boldsymbol{\psi} = \mathbf{I}$ and $\boldsymbol{\xi}_k = 0$ in scale, that is, only one measurement at a fixed time point is taken for each individual, leading to a multiple interacting iQTL model for a univariate quantitative trait; take $\boldsymbol{\psi}$ to be an identity matrix of order m, and $\boldsymbol{\xi}_k$ to be a zero vector, resulting in a multiple interacting iQTL model for a multiple quantitative trait. If $\boldsymbol{\xi}_k$ is assigned to be non-zero in the two cases above, this leads to a multiple iQTL model for a single developmental trait or multiple developmental traits. Corresponding Bayesian model selection approaches can be likewise obtained by taking different values for $\boldsymbol{\psi}$ and $\boldsymbol{\xi}_k$. The computational code to implement the proposed method (termed Bafmim) is available upon request.

SUPPLEMENTARY DATA

Supplementary Data are available online at http://bib.oxfordjournals.org/.

Key Points

- We developed a Bayesian method for mapping iQTL underlying dynamic/developmental traits.
- A Bayesian model selection strategy for longitudinal traits was proposed to infer iQTLs with different imprinting patterns and mechanisms as discussed in Cheverud *et al.* [13].
- The method can do multiple iQTL inference and outperforms maximum likelihood-based single iQTL mapping methods for developmental traits in terms of power to detect (i)QTLs.
- The method was applied to identify iQTLs responsible for growth trajectory of body weight in mice. The pattern of imprinting for the identified iQTLs was further dissected.

Acknowledgements

We thank the four anonymous referees for their insightful comments that greatly improved the article. We also thank Prof. J. Cheverud for providing the mouse data set and Prof. J Stapleton for careful reading of the article.

FUNDING

Chinese National Natural Science Foundation (grants 30972077 and 31110103065 to R.Y.); US National Science Foundation (grant DMS-0707031 to Y.C.).

References

1. Reik W, Walter J. Genomic imprinting: parental influence on the genome. *Nat Rev Genet* 2001;**2**:21–32.
2. Wood HM, Grahame JW, Humphray S, *et al*. Sequence differentiation in regions identified by a genome scan for local adaptation. *Mol Ecol* 2008;**17**:3123–35.
3. Alleman M, Doctor J. Genomic imprinting in plants: observations and evolutionary implications. *Plant Mol Biol* 2000; **43**:147–61.
4. Nezer C, Moreau L, Brouwers B, *et al*. An imprinted QTL with major effect on muscle mass and fat deposition maps to the IGF2 locus in pigs. *Nat Genet* 1999;**21**:155–6.

5. Van Laere AS, Nguyen M, Braunschweig M, *et al*. A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. *Nature* 2003;**425**:832–6.
6. Falls JG, Pulford DJ, Wylie AA, *et al*. Genomic imprinting: implications for human disease. *Am J Pathol* 1999;**154**: 635–47.
7. McInnis MG, Lan TH, Willour VL, *et al*. Genome-wide scan of bipolar disorder in 65 pedigrees: supportive evidence for linkage at 8q24, 18q22, 4q32, 2p12, and 13q12. *Mol Psychiatry* 2003;**8**:288–98.
8. Isles AR, Holland AJ. Imprinted genes and mother-offspring interactions. *Early Hum Dev* 2005;**81**:73–7.
9. Tycko B, Morison IM. Physiological functions of imprinted genes. *J Cell Physiol* 2002;**192**:245–58.
10. Constancia M, Kelsey G, Reik W. Resourceful imprinting. *Nature* 2004;**432**:53–7.
11. Naumova AK, Croteau S. Mechanisms of epigenetic variation: polymorphic imprinting. *Curr Genomics* 2004;**5**: 417–29.
12. Sandovici I, Kassovska-Bratinova S, Loredo-Osti JC, *et al*. Interindividual variability and parent of origin DNA methylation differences at specific human Alu elements. *Hum Mol Genet* 2005;**14**:2135–43.
13. Cheverud JM, Hager R, Roseman C, *et al*. Genomic imprinting effects on adult body composition in mice. *Proc Natl Acad Sci USA* 2008;**105**:4253–8.
14. Hanson RL, Kobes S, Lindsay RS, *et al*. Assessment of parent-of-origin effects in linkage analysis of quantitative traits. *Am J Hum Genet* 2001;**68**:951–62.
15. Haghighi F, Hodge SE. Likelihood formulation of parent-of-origin effects on segregation analysis, including ascertainment. *Am J Hum Genet* 2002;**70**:142–56.
16. Shete S, Amos CI. Testing for genetic linkage in families by a variance-components approach in the presence of genomic imprinting. *Am J Hum Genet* 2002;**70**:751–7.
17. Li GX, Cui YH. A statistical variance components framework for mapping imprinted quantitative trait loci in experimental crosses. *J Probab Statist* 2009; Article ID 689489.
18. Li GX, Cui YH. A general statistical framework for dissecting parent-of-origin effects underlying triploid endosperm traits in flowering plants. *Ann Appl Statist* 2010;**4**: 1214–33.
19. Knott SA, Marklund L, Haley CS, *et al*. Multiple marker mapping of quantitative trait loci in a cross between outbred wild boar and large white pigs. *Genetics* 1998;**149**: 1069–80.
20. de Koning DJ, Bovenhuis H, van Arendonk JA. On the detection of imprinted quantitative trait loci in experimental crosses of outbred species. *Genetics* 2002;**161**:931–8.
21. Cui Y, Lu Q, Cheverud JM, *et al*. Model for mapping imprinted quantitative trait loci in an inbred F2 design. *Genomics* 2006;**87**:543–51.
22. Cui Y. A statistical framework for genome-wide scanning and testing of imprinted quantitative trait loci. *J Theor Biol* 2007;**244**:115–26.
23. Wang C, Wang Z, Prows DR, *et al*. A computational model for the inheritance pattern of genomic imprinting for complex traits. *Brief Bioinfor* 2011;**13**(1):34–45.
24. Bartolomei MS, Tilghman SM. Genomic imprinting in mammals. *Annu Rev Genet* 1997;**31**:493–525.
25. Villar AJ, Eddy EM, Pedersen RA. Developmental regulation of genomic imprinting during gametogenesis. *Dev Biol* 1995;**172**:264–71.
26. Cui Y, Li S, Li G. Functional mapping imprinted quantitative trait loci underlying developmental characteristics. *Theor Biol Med Model* 2008;**5**:6.
27. Kao CH, Zeng ZB, Teasdale RD. Multiple interval mapping for quantitative trait loci. *Genetics* 1999;**152**:1203–16.
28. Yang R, Tian Q, Xu S. Mapping quantitative trait loci for longitudinal traits in line crosses. *Genetics* 2006;**173**:2339–56.
29. Falconer DS, Mackay TFC. *Introduction to Quantitative Genetics*. 4th edn. London: Longman, 1996.
30. Mantey C, Brockmann GA, Kalm E, *et al*. Mapping and exclusion mapping of genomic imprinting effects in mouse F2 families. *J Hered* 2005;**96**:329–38.
31. Yang R, Xu S. Bayesian shrinkage analysis of quantitative trait Loci for dynamic traits. *Genetics* 2007;**176**:1169–85.
32. Cui Y, Wu R, Casella G, *et al*. Nonparametric functional mapping of quantitative trait loci underlying programmed cell death. *Stat Appl Genet Mol Biol* 2008;**7**. Article4.
33. George EI, Mccullochr RE. Approaches for Bayesian variable selection. *Stat Sin* 1997;**7**:339–73.
34. Kuo L, Mallick B. Variable selection for regression models. *Sankhya B* 1998;**60**:65–81.
35. Chipman H, Edwards EI, Mcculloch RE. *The Practical Implementation of Bayesian Model Selection,* Vol. 13. Institute of Mathematical Statistics, 2001, 65–134.
36. Wang H, Zhang YM, Li X, *et al*. Bayesian shrinkage estimation of quantitative trait loci parameters. *Genetics* 2005;**170**: 465–80.
37. Sillanpää MJ, Arjas E. Bayesian mapping of multiple quantitative trait loci from incomplete inbred line cross data. *Genetics* 1998;**148**:1373–88.
38. Sillanpää MJ, Arjas E. Bayesian mapping of multiple quantitative trait loci from incomplete outbred offspring data. *Genetics* 1999;**151**:1605–19.
39. Gelman A, Roberts G, Gilks W. Efficient metropolis jumping rules. In: Bernardo JM, Berger JO, Dawid AP, *et al*, (eds). *Bayesian Statistics 5*. Oxford University Press, 1995, 599–607.
40. Metropolis N, Rosenbluth AW, Rosenbluth MN, *et al*. Equations of state calculations by fast computing machines. *J Chem Phys* 1953;**21**:1087–91.
41. Hastings WK. Monte Carlo sampling methods using Markov Chains and their applications. *Biometrika* 1970;**57**: 97–109.
42. Plummer M, Best N, Cowles K, *et al*. *Output Analysis and Diagnostics for MCMC* 2004. http://cran.r-project.org/web/packages/coda/coda.pdf.
43. Raftery AE, Madigan D, Hoteing JA. Bayesian model averaging for linear regression models. *J Am Stat Assoc* 1997;**92**: 179–91.
44. Ball RD. Bayesian methods for quantitative trait loci mapping based on model selection: approximate analysis using the Bayesian information criterion. *Genetics* 2001;**159**: 1351–64.
45. Sillanpää MJ, Corander J. Model choice in gene mapping: what and why. *Trend Genet* 2002;**18**:301–7.
46. Kass RE, Raftery AE. Bayes factors. *J Am Stat Assoc* 1995;**90**: 773–95.

47. Wolf JB, Cheverud JM, Roseman C, *et al*. Genome-wide analysis reveals a complex pattern of genomic imprinting in mice. *PLoS Genet* 2008;**4**:1–12.
48. Yang R, Wang X, Wu Z, *et al*. Bayesian model selection for characterizing genomic imprinting effects and patterns. *Bioinformatics* 2010;**26**:235–41.
49. Vaughn TT, Pletscher LS, Peripato A, *et al*. Mapping quantitative trait loci for murine growth: a closer look at genetic architecture. *Genet Res* 1999;**74**:313–22.
50. Cui Y, Cheverud J, Wu R. A statistical model for dissecting genomic imprinting through genetic mapping. *Genetica* 2007;**130**:227–39.
51. Yi N, Banerjee S, Pomp D, *et al*. Bayesian mapping of genomewide interacting quantitative trait loci for ordinal traits. *Genetics* 2007;**176**:1855–64.
52. Yi N, Yandell BS, Churchill GA, *et al*. Bayesian model selection for genome-wide epistatic quantitative trait loci analysis. *Genetics* 2005;**170**:1333–44.
53. Yang R, Wang X, Wu Z, *et al*. Bayesian model selection for characterizing genomic imprinting effects and patterns. *Bioinformatics* 2010;**26**:235–41.
54. Min L, Yang R, Wang X, Wang B. Bayesian analysis for genetic architecture of dynamics traits. *Heredity* 2011;**106**: 124–33.
55. Ma CX, Casella G, Wu R. Functional mapping of quantitative trait loci underlying the character process: a theoretical framework. *Genetics* 2002;**161**:1751–62.
56. Yap JS, Fan J, Wu R. Nonparametric modeling of longitudinal covariance structure in functional mapping of quantitative trait loci. *Biometrics* 2009;**65**:1068–77.
57. Zhang YM, Xu S. Advanced statistical methods for detecting multiple quantitative trait loci. *Recent Res Devel Genet Breeding* 2005;1–23.

APPENDIX 1
The derivation of conditional posteriors

The joint posterior density can be expressed as

$$p(\theta|\mathbf{y}) = p(\mathbf{y}|\theta)p(\theta) = p(\mathbf{y}|\theta)p(\boldsymbol{\mu})p(\boldsymbol{\beta})p(\gamma)p(\boldsymbol{\xi})p(\boldsymbol{\Sigma})p(\lambda),$$

where $\theta = (\boldsymbol{\mu} \quad \boldsymbol{\beta} \quad \gamma \quad \boldsymbol{\xi} \quad \boldsymbol{\Sigma} \quad \lambda)$ with $\boldsymbol{\beta} = \{\boldsymbol{\beta}_j\}$, $\gamma = \{\gamma_j, \lambda\} = \{\lambda_\mathbf{j}\}$ for $\mathbf{j} = 1, 2, \ldots, \mathbf{L}$ and $\boldsymbol{\xi} = \{\boldsymbol{\xi}_k\}$ for $k = 1, 2, \ldots, n$.

The full conditional posterior distributions of all parameters can be derived from the joint posterior density by fixing other parameters as constants.

Given the other parameters, the full conditional posterior distribution of $\boldsymbol{\mu}$ is derived as a multivariate normal with mean $(n\boldsymbol{\psi}V^{-1}\boldsymbol{\psi}^{\mathrm{T}})^{-1}\boldsymbol{\psi}V^{-1} \sum_{k=1}^{n}(\mathbf{y}_k - \mathbf{U}_i + \boldsymbol{\psi}^{\mathrm{T}}\boldsymbol{\mu})$ and covariance matrix $(n\boldsymbol{\psi}V^{-1}\boldsymbol{\psi}^{\mathrm{T}})^{-1}$.

The full conditional posterior distribution of $\boldsymbol{\beta}_j$ $(j = 1, 2, \ldots, L)$ is also normal, whose mean is $\hat{\beta}_j = \left[\left(1 + \frac{1}{c}\right)\sum_{k=1}^{n} x_{kj}^2 \boldsymbol{\psi}V^{-1}\boldsymbol{\psi}^{\mathrm{T}}\right]^{-1} \boldsymbol{\psi}V^{-1} \sum_{k=1}^{n} x_{kj}(\mathbf{y}_k - \mathbf{U}_k + \gamma_j x_{kj}\boldsymbol{\psi}^{\mathrm{T}}\boldsymbol{\beta}_j)$ and covariance matrix is $\hat{\Sigma}_j = \left[\left(1 + \frac{1}{c}\right)\sum_{k=1}^{n} x_{kj}^2 \boldsymbol{\psi}V^{-1}\boldsymbol{\psi}^{\mathrm{T}}\right]^{-1}$, where $x_{kj} = z_{kj}$, w_{kj} or s_{kj} and $\boldsymbol{\beta}_j = \mathbf{a}_j, \mathbf{d}_j$ or $\mathbf{i}_j$.

The full conditional posterior distribution of $\boldsymbol{\xi}_k$ follows a normal distribution with mean $\boldsymbol{\Sigma}\boldsymbol{\psi}V^{-1}(\mathbf{y}_k - \mathbf{U}_k)$ and covariance matrix $\boldsymbol{\Sigma} - \boldsymbol{\Sigma}\boldsymbol{\psi}V^{-1}\boldsymbol{\psi}^{\mathrm{T}}\boldsymbol{\Sigma}$.

The full conditional posterior distribution of $\boldsymbol{\Sigma}$ follows an inverse Wishart distribution $\mathrm{IW}\left(v_a + n, \sum_{k=1}^{n} \boldsymbol{\xi}_k\boldsymbol{\xi}_k^{\mathrm{T}} + \mathbf{S}_a\right)$.

The full conditional posterior distribution for residual variance σ^2 corresponds to a scaled inverse χ^2 with parameters $v_e + n$ and $(v_e + n)s_e + (\sum_{k=1}^{n} \mathbf{e}_k^T\mathbf{e}_k)^{-1}$, where $\mathbf{e}_k = \mathbf{y}_k - \mathbf{U}_k - \boldsymbol{\psi}^T\boldsymbol{\xi}_k$.

The full conditional posterior for γ_j $(j = 1, 2, \ldots, L)$ can be constructed as a Bernoulli distribution. To improve sampling efficiency, Metropolis–Hastings algorithm is applied with the acceptance rate given by

$$\begin{aligned}\alpha &= \frac{p(\mathbf{y}\left|\gamma_j = 1, \gamma_{-j}, \boldsymbol{\beta}_{-j}\right.)}{p(\mathbf{y}\left|\gamma_j = 0, \gamma_{-j}, \boldsymbol{\beta}_{-j}\right.)} \cdot \frac{1-w}{w} \\ &= \frac{p(\mathbf{y}\left|\gamma_j = 1, \gamma_{-j}, \boldsymbol{\beta}_{-j}, \boldsymbol{\beta}_j\right.)p(\boldsymbol{\beta}_j\left|\gamma_j = 1, \gamma_{-j}, \boldsymbol{\beta}_{-j}\right.)}{p(\mathbf{y}\left|\gamma_j = 0, \gamma_{-j}, \boldsymbol{\beta}_{-j}\right.)p(\boldsymbol{\beta}_j\left|\gamma_j = 1, \gamma_{-j}, \boldsymbol{\beta}_{-j}, \mathbf{y}\right.)} \cdot \frac{1-w}{w} \\ &= \sqrt{\frac{c}{c+1}} \exp\left(-\frac{1}{2}\hat{\beta}_j^{\,T}\hat{\Sigma}_j^{-1}\hat{\beta}_j\right) \cdot \frac{1-w}{w}\end{aligned}$$

where γ_{-j} means all the elements of γ except for γ_j and $\boldsymbol{\beta}_{-j}$ represents all the elements of $\boldsymbol{\beta}$ except for $\boldsymbol{\beta}_j$. $p(\boldsymbol{\beta}_j|\gamma_j = 1, \gamma_{-j}, \boldsymbol{\beta}_{-j})$. and $p(\boldsymbol{\beta}_j|\gamma_j = 1, \gamma_{-j}, \boldsymbol{\beta}_{-j}, \mathbf{y})$. are prior and posterior probabilities, respectively. By generating a probability from a uniform distribution, we accept $\gamma_j = 1$ with a probability of min(1, α).

The full conditional posterior distribution for QTL positions does not have a closed form, Metropolis–Hastings algorithm will be used to sample QTL positions. Each locus is sampled from a variable interval size whose boundaries are the positions of adjoining QTLs [36, 57]. We restricted the minimal distance between two QTLs to be 10cM at least. A new position λ_j^* $(j = 1, 2, \ldots, L)$ is sampled from two placed loci on the left and right of existing position λ_j^0 with a uniform distribution, then the new position is accepted with a probability of min(1, α) with

$$\alpha = \frac{\prod_{k=1}^{n} p(\mathbf{y}_k|\lambda_j^*, \theta_{-\lambda_j})}{\prod_{k=1}^{n} p(\mathbf{y}_k|\lambda_j^0, \theta_{-\lambda_j})}$$