Data-Driven Reversible Jump for QTL Mapping

Daiane Aparecida Zuanetti¹ and Luis Aparecido Milan

Departamento de Estatística, Universidade Federal de São Carlos, São Carlos, SP, 13565905, Brazil

ABSTRACT We propose a birth–death–merge data-driven reversible jump (DDRJ) for multiple-QTL mapping where the phenotypic trait is modeled as a linear function of the additive and dominance effects of the unknown QTL genotypes. We compare the performance of the proposed methodology, usual reversible jump (RJ) and multiple-interval mapping (MIM), using simulated and real data sets. Compared with RJ, DDRJ shows a better performance to estimate the number of QTLs and their locations on the genome mainly when the QTLs effect is moderate, basically as a result of better mixing for transdimensional moves. The inclusion of a merge step of consecutive QTLs in DDRJ is efficient, under tested conditions, to avoid the split of true QTL's effects between false QTLs and, consequently, selection of the wrong model. DDRJ is also more precise to estimate the QTLs location than MIM in which the number of QTLs need to be specified in advance. As DDRJ is more efficient to identify and characterize QTLs with smaller effect, this method also appears to be useful and brings contributions to identifying single-nucleotide polymorphisms (SNPs) that usually have a small effect on phenotype.

KEYWORDS QTL mapping; model selection; data-driven reversible jump; birth–death–merge movements; mixing of MCMC

GENETICISTS and molecular biologists have aimed at locating regions associated with quantitative traits in a chromosome. These chromosomal regions are known as quantitative trait *loci* (QTL) and their location and effects on the phenotypic traits are estimated by genetic markers. The most popular genetic markers are simple sequence repeats (SSR) and single-nucleotide polymorphism (SNP); their location is specified by the linkage map and their genotype is known.

A phenotype is usually modeled as a linear function of the additive and dominance effects of the QTL genotypes and several methods have been developed for the localization and characterization of QTLs. The standard estimation method in experimental crosses is the interval mapping (IM) presented by Lander and Botstein (1989) and Haley and Knott (1992). Lander and Botstein (1989) propose using the EM algorithm (Dempster *et al.* 1977), assuming a single putative QTL at each location on the genome and comparing the hypothesis of a single QTL to the null hypothesis of no segregation QTLs by the logarithm of the odds ratio (LOD score). However, the estimate of the QTL effects can be influenced by the effect of other possible QTLs in adjacent regions since this effect is not

controlled in the model and nonexistent or ghost QTLs can be identified. A ghost QTL appears when two or more QTLs are linked in coupling (meaning that their effects have the same sign) and interval mapping gives a maximum LOD score at a location between the two QTLs (Broman and Speed 1999).

Jansen (1993), Jansen and Stam (1994), and Zeng (1994) propose composite-interval mapping (CIM) to control the effect of QTLs located in adjacent regions and avoid the identification of ghost QTLs. They propose to include in the single putative QTL regression model a subset of markers as cofactors. Kao *et al.* (1999) propose multiple-interval mapping (MIM), which considers the effect of all possible QTLs and the epistatic effect between them in a single model. This model, with a fixed number of QTLs, is estimated by the EM algorithm and the number of QTLs is selected by model selection methods such as the Akaike information criterion (AIC) and the Bayesian information criterion (BIC), among others.

Bayesian methods for QTL mapping are interesting tools since they allow us to select and estimate the model jointly. Earlier Bayesian approaches were proposed by Stephens and Smith (1993) and Satagopan *et al.* (1996). The authors estimate the locations and effect of a prespecified number of QTLs. In practice, however, the number of QTLs is unknown and must be estimated. Satagopan and Yandell (1996) and Stephens and Fisch (1998) propose variants of reversible-jump (RJ) Markov chain Monte Carlo to estimate it and the remaining parameters of the model jointly. An important characteristic in the chain generated in MCMC is that it mixes

Copyright © 2016 by the Genetics Society of America

doi: 10.1534/genetics.115.180802

Manuscript received July 16, 2015; accepted for publication October 30, 2015; published Early Online November 4, 2015.

Supporting information is available online at www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.180802/-DC1.

¹Departamento de Estatística, Universidade Federal de São Carlos, Rodovia Washington Luís, KM 235, São Carlos, SP 13565905, Brazil. E-mail: dzuanetti@yahoo.com.br

well; i.e., it moves around the parameter space rather easily and quickly finds its stationary distribution. Forming a good Markov chain and monitoring its behavior is delicate and sophisticated work (Broman and Speed 1999).

Over the past decade, different ways to generate proposal parameters in MCMC have been suggested to facilitate the moves between models and accelerate the convergence of the original RJ algorithm. Green and Mira (2001) propose an algorithm that, on rejection, makes a second attempt to move. Regarding the inclusion of a new QTL, Yi and Xu (2002) suggest generating its effects (additive and dominance) from the conditional *a posteriori* distribution. Yi *et al.* (2005) propose updating the location of a specific QTL and its genotypes together. As a QTLs location and genotype are correlated, the acceptance probability of a new QTLs location is higher if its genotype is updated jointly.

To accelerate the search procedure of the correct number of QTLs, K , more suitable and efficient dimensional change candidates must be generated. For this purpose, we propose a birth–death–merge data-driven reversible jump (DDRJ) for multiple-QTL mapping. It simulates a more likely location for a new QTL using the available data, chooses a QTL to be excluded according to its importance in the current model, or merges the effects of two consecutive QTLs if their genotypes are correlated. Consequently, candidates are more likely to be accepted and the space of possible models is more easily explored. Jain and Neal (2004, 2007) and Saraiva and Milan (2012) show that data-driven methods are effective in simplifying the methodology and improving the chain mixing.

The merge movement of consecutive QTLs is efficient under tested conditions to avoid identification of false QTLs. Usually, as both QTLs have similar estimated genotypes, the effects of the true QTL are split between the two QTLs and bias the estimate of the number of QTLs and their effects. Split QTLs can be seen as the opposite problem to that of ghost QTLs.

The proposed method has also the advantage of providing interval estimates that can be used to analyze the uncertainty of estimates. The usual methods generally provide only point estimates or asymptotic confidence intervals for big samples.

This article is organized as follows: *Model for Quantitative Traits* presents such a model and discusses the likelihood function; *Bayesian Approach* addresses the Bayesian approach for the model, including the DDRJ procedure to estimate the number of QTLs; *Applications* analyzes the performance of DDRJ and compares it with RJ and MIM performance in simulated and real data sets. Finally, *Discussion* provides a discussion of the methods.

Model for Quantitative Traits

Let $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)$ be a quantitative trait of n individuals from an F_2 population. Assume this phenotype has been affected by K QTLs located at positions $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_K)$, $\lambda_k < \lambda_{k+1}$ for $k = 1, \dots, K-1$, between m different genotyped markers with a known linkage map.

Phenotype Y_i for the i th individual can be modeled by

$$Y_i = \mu + \sum_{k=1}^K \alpha_k Q_{ik} + \sum_{k=1}^K \delta_k (1 - |Q_{ik}|) + \epsilon_i, \quad (1)$$

where μ is the average of expected values of genotypes AA and aa ; α_k is the additive effect of the k th QTL; δ_k is the dominance effect of the k th QTL; Q_{ik} represents the genotype of the k th QTL of the i th individual coded as $-1, 0$ or 1 for aa , Aa , or AA , respectively, $k = 1, \dots, K$ and $i = 1, 2, \dots, n$; $\epsilon_i \sim \text{Normal}(0, \sigma^2)$ is the random error; and ϵ_i and $\epsilon_{i'}$ are supposed to be independent for $i \neq i'$.

The phenotype can also be affected by environmental covariates and interactions among QTLs or between covariates and QTLs. The model defined by Equation 1 does not consider these effects, but extensions (modeling environmental covariates as fixed effects, for example) are straightforward.

The data set consists of $\mathbf{y} = (y_1, y_2, \dots, y_n)$; the observations regarding the quantitative trait of n individuals; $\mathbf{M}_{(n \times m)}$, the markers' genotype are coded as $-1, 0$ or 1 for aa , Aa , or AA , respectively; and $\mathbf{D} = \{D_1, D_2, \dots, D_m\}$, the distances (in centimorgans) between each marker and the first marker, where $D_1 = 0$.

We assume there is at most one QTL between two consecutive markers, therefore $K < m$, and the QTLs genotype is explained only by flanking markers; i.e., $Q_{ik}|M_{ir_k}, M_{il_k}$ and $Q_{ik'}|M_{ir_{k'}}, M_{il_{k'}}$ are independent for $k \neq k'$, where M_{ir_k} is the genotype of the marker to the right of the k th QTL for the i th individual and M_{il_k} is the genotype of the marker to the left of the k th QTL for the i th individual.

The joint probability distribution of $\mathbf{Y}$ and $\mathbf{Q}$, where $\mathbf{Q} = \{Q_{ik}\}$ is the matrix of the K QTLs genotypes for the n individuals, is

$$f_{\mathbf{Y}, \mathbf{Q} | \mathbf{M}, \mathbf{D}}(\mathbf{y}, \mathbf{q}) = \prod_{i=1}^n f_{Y_i | \mathbf{q}_i}(y_i) \Pr(\mathbf{Q}_i = \mathbf{q}_i | \mathbf{M}_i, \mathbf{D}), \quad (2)$$

where $\Pr(Q_{i1} = q_{i1}, \dots, Q_{iK} = q_{iK} | \mathbf{M}_i, \mathbf{D}) = \prod_{k=1}^K \Pr(Q_{ik} = q_{ik} | M_{ir_k}, M_{il_k}, \mathbf{D})$, for $i = 1, \dots, n$; $\sum_{q_{ik}} \Pr(Q_{ik} = q_{ik} | M_{ir_k}, M_{il_k}, \mathbf{D}) = 1$, for $q_{ik} = -1, 0, 1$; and f is the conditional normal density for Y_i .

In practice, the number of QTLs K is unknown and the parameters of the model are $\boldsymbol{\theta} = (K, \boldsymbol{\lambda}, \mu, \boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_K), \boldsymbol{\delta} = (\delta_1, \dots, \delta_K), \sigma^2)$. The likelihood function of $\boldsymbol{\theta}$ given $\mathbf{Y} = \mathbf{y}$ and $\mathbf{Q} = \mathbf{q}$ is

$$L(\boldsymbol{\theta} | \mathbf{y}, \mathbf{q}) = (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n \epsilon_i^2\right\} \times \prod_{i=1}^n \prod_{k=1}^K \Pr(Q_{ik} = q_{ik} | M_{ir_k}, M_{il_k}, \mathbf{D}), \quad (3)$$

where $\epsilon_i = y_i - \mu - \sum_{k=1}^K \alpha_k q_{ik} - \sum_{k=1}^K \delta_k (1 - |q_{ik}|)$ is the residual of the i th observation and $\Pr(Q_{ik} = q_{ik} | M_{ir_k}, M_{il_k}, \mathbf{D})$ is the conditional probability of the QTL genotype given the flanking markers genotypes as defined by Stephens and Fisch (1998). Such a probability is a function of recombination

fractions between the k th QTL and its flanking markers calculated by the Haldane distance function. Note that Q_{ik} , $i = 1, \dots, n$ and $k = 1, \dots, K$, is nonobservable and must be estimated.

Without losing the generality and for simplicity, consider the models with one and two QTLs defined, respectively, as

$$Y_i = \mu + \alpha_1 Q_{i1} + \delta_1(1 - |Q_{i1}|) + \epsilon_i(M_1)$$

and

$$Y_i = \mu + \alpha'_1 Q'_{i1} + \alpha'_2 Q'_{i2} + \delta'_1(1 - |Q'_{i1}|) + \delta'_2(1 - |Q'_{i2}|) + \epsilon_i(M_2),$$

for $i = 1, \dots, n$. Observe if $Q'_{i1} = Q'_{i2} = Q_{i1}$ for all or almost all individuals, $\alpha'_1 + \alpha'_2 = \alpha_1$ and $\delta'_1 + \delta'_2 = \delta_1$, the models M_1 and M_2 are equally or almost equally likely and it can be hard to select the correct model in this situation. The genotype of two loci has a high probability of being equal when they are close on the same chromosome and the model is wrongly estimated if the effect of two or more true close QTLs are merged in only one QTL or if the effect of one true QTL is split with one or more false close QTLs. We note in our simulated data sets, some of them shown in *Applications*, using multiple QTLs methods to estimate the model, that often methods split the effect of one true QTL with one or more false QTLs. Conventional methodologies for QTL mapping often do not deal well with this problem.

Data availability

File S1 contains the conditional *a posteriori* distributions of parameters. File S2 contains DDRJ and RJ effective sample size of the parameters of the models. R codes of DDRJ methods are provided in File S3 and File S4.

Bayesian Approach

The usual Bayesian methodology for models with unknown K is the RJ proposed by Green (1995). This method consists of running Metropolis–Hastings steps that either accept or reject different moves, like “birth” or “death” of a QTL. These steps enable transitions from the current model to models of higher or lower dimensions.

Parameters $\lambda|K$, $\alpha|K$, $\delta|K$, μ , σ^2 , and elements of α and δ are supposed to be independent and the joint *a priori* density for θ is written as

$$\pi(\theta) = \pi(K)\pi(\lambda|K)\left(\prod_{k=1}^K \pi(\alpha_k)\pi(\delta_k)\right)\pi(\mu)\pi(\sigma^2). \quad (4)$$

Particularly, we consider

1. $K \sim \text{Uniform}(0, 1, \dots, m-1)$.
2. $\alpha_k \sim \text{Normal}(\nu_\alpha, \sigma_\alpha^2)$, $k = 1, \dots, K$, where ν_α and $\sigma_\alpha^2 > 0$ are known hyperparameters.
3. $\delta_k \sim \text{Normal}(\nu_\delta, \sigma_\delta^2)$, $k = 1, \dots, K$, where ν_δ and $\sigma_\delta^2 > 0$ are known hyperparameters.

4. $\mu \sim \text{Normal}(\nu_\mu, \sigma_\mu^2)$, where ν_μ and $\sigma_\mu^2 > 0$ are known hyperparameters.
5. $\sigma^2 \sim \text{Inverse-gamma}(\eta_a, \eta_b)$, where $\eta_a > 0$ and $\eta_b > 0$ are known hyperparameters.
6. $\pi(\lambda|K) = \pi(\lambda_1, \dots, \lambda_K|K) = \pi(\lambda_1|K)\pi(\lambda_2|\lambda_1, K) \dots \pi(\lambda_K|\lambda_{K-1}, K)$. If there is no *a priori* information about the QTLs location, each location is assumed uniformly distributed over the possible loci.

Combining the likelihood function in Equation 3 with the *a priori* distributions, we obtain the conditional *a posteriori* distributions of $\mu|(\mathbf{y}, \mathbf{q}, \theta_{-\mu})$, $\sigma^2|(\mathbf{y}, \mathbf{q}, \theta_{-\sigma^2})$, $\alpha_k|(\mathbf{y}, \mathbf{q}, \theta_{-\alpha_k})$, $\delta_k|(\mathbf{y}, \mathbf{q}, \theta_{-\delta_k})$, and $\lambda_k|(\mathbf{y}, \mathbf{q}, \theta_{-\lambda_k})$, $k = 1, \dots, K$, provided in Supporting Information, File S1.

The nonobservable genotype Q_{ik} , $i = 1, \dots, n$ and $k = 1, \dots, K$, is simulated and updated by its conditional *a posteriori* distribution given by

$$\begin{aligned} \Pr(Q_{ik} = q_{ik} | \mathbf{y}, \mathbf{q}_{-q_{ik}}, \mathbf{M}, \mathbf{D}) &\propto \\ \Pr(Q_{ik} = q_{ik}, Y_i = y_i | \mathbf{q}_{-q_{ik}}, M_{ir_k}, M_{il_k}, \mathbf{D}) & \\ = f_{Y_i|q_i}(y_i) \Pr(Q_{ik} = q_{ik} | M_{ir_k}, M_{il_k}, \mathbf{D}), & \end{aligned} \quad (5)$$

for $q_{ik} \in \{-1, 0, 1\}$ and where $f_{Y_i|q_i}(y_i)$ is the $\text{Normal}(\mu + \sum_{k=1}^K \alpha_k q_{ik} + \sum_{k=1}^K \delta_k(1 - |q_{ik}|), \sigma^2)$ density function.

From Equation 5, $Q_{ik} | (\mathbf{y}, \mathbf{q}_{-q_{ik}}, \mathbf{M}, \mathbf{D}) \sim \text{Multinomial}(1, (p_{ik-1}, p_{ik0}, p_{ik1}))$, where

$$p_{ikj} = \frac{f_{Y_i|q_i}(y_i) \Pr(Q_{ik} = j | M_{ir_k}, M_{il_k}, \mathbf{D})}{\sum_j f_{Y_i|q_i}(y_i) \Pr(Q_{ik} = j | M_{ir_k}, M_{il_k}, \mathbf{D})},$$

$$j = -1, 0, 1.$$

Parameters μ , σ^2 , α_k , and δ_k and nonobservable variables Q_{ik} , $i = 1, \dots, n$ and $k = 1, \dots, K$, are updated by Gibbs sampling steps and λ_k is updated jointly with $\mathbf{Q}_k$ by Metropolis–Hastings steps, in which λ'_k is sampled from a $\text{Uniform}(D_{l_k}, D_{r_k})$ distribution and the block $(\lambda'_k, \mathbf{Q}'_k)$ is accepted according to probability $\Psi((\lambda'_k, \mathbf{q}'_k) | (\lambda_k, \mathbf{q}_k)) = \min(1, A)$, where

$$\begin{aligned} A = \frac{\exp\{-(1/2\sigma^2)\sum_{i=1}^n \epsilon_i'^2\}}{\exp\{-(1/2\sigma^2)\sum_{i=1}^n \epsilon_i^2\}} \frac{\prod_{i=1}^n \Pr(Q_{ik} = q'_{ik} | M_{ir_k}, M_{il_k}, \mathbf{D})}{\prod_{i=1}^n \Pr(Q_{ik} = q_{ik} | M_{ir_k}, M_{il_k}, \mathbf{D})} \\ \times \frac{\prod_{i=1}^n \Pr(Q_{ik} = q_{ik} | \mathbf{y}, \mathbf{q}_{-q_{ik}}, \mathbf{M}, \mathbf{D})}{\prod_{i=1}^n \Pr(Q_{ik} = q'_{ik} | \mathbf{y}, \mathbf{q}_{-q_{ik}}, \mathbf{M}, \mathbf{D})}, \end{aligned} \quad (6)$$

$\epsilon_i = y_i - \mu - \sum_{k=1}^K \alpha_k q_{ik} - \sum_{k=1}^K \delta_k(1 - |q_{ik}|)$ is the residual of the i th individual, $i = 1, \dots, n$, and ϵ_i' is calculated using $\mathbf{q}_{-q_k}$ and $\mathbf{q}'_k$.

DDRJ

The movements that change K are called birth (b), death (d), or merge (mg) moves when a new QTL is included in the model or, conversely, one QTL is excluded from the current model or the effects of two QTLs are summed into a single QTL. The birth, death, and merge moves are implemented by

Figure 1 Trace plot of K for $\sigma = 0.5$. (A) RJ sequence; (B) DDRJ sequence.

Metropolis–Hastings steps and either increase or reduce the number of QTLs by one at each step.

Consider $x = (\mathbf{q}, \boldsymbol{\theta})$ the current state of the MCMC procedure with K QTLs and $x' = (\mathbf{q}', \boldsymbol{\theta}')$ the proposed movement, where $'$ means a birth (b), a death (d), or a merge (mg) of QTLs. Therefore, $K' = K + 1$ if a birth movement is proposed or $K' = K - 1$ if a death or a merge movement is proposed. This move is accepted according to Metropolis–Hastings probability $\Psi(x'|x) = \min(1, A')$, where

$$A' = \frac{L(\boldsymbol{\theta}'|\mathbf{y}, \mathbf{q}')}{L(\boldsymbol{\theta}|\mathbf{y}, \mathbf{q})} \frac{\pi(\boldsymbol{\theta}')}{\pi(\boldsymbol{\theta})} \frac{q(x|x')}{q(x'|x)}, \quad (7)$$

and $q(\cdot)$ is the transition function, described below.

At each step, we choose a movement to increase or reduce the number of QTLs as follows:

1. If $0 < K < m - 1$, a birth or a death is randomly chosen, according to its probability. Here, we assume $\Pr(b|K) = 1/2$ and $\Pr(d|K) = 1/2$.
2. If $K = 0$, a birth is chosen; i.e., $\Pr(b|K) = 1$.
3. If $K = m - 1$, a death is chosen; i.e., $\Pr(d|K) = 1$.

Birth proposal: When a birth movement is chosen, a location is selected for the new QTL in a marker interval that has no QTL and its genotype and effect parameters must be defined. The selection of a location through a Uniform distribution can be inefficient, mainly if we have a large number of marker intervals.

If there is a strong association between a marker and a trait, it is reasonable to suppose there is a QTL nearby that marker. Therefore, the association between markers and trait can be used to guide the search for new QTLs in the estimation process. As each marker can be seen as a factor with three levels affecting differently the phenotype mean or the residual mean of the current model, we use the Kruskal–Wallis test statistic to measure this association. The F statistics in a one-way analysis of variance could also be used. Higher values indicate

the residual mean is different for the distinct levels of the marker and there is a higher chance of a QTL close to it whose effect is not considered in the current model. Values close to zero indicate the residual mean is the same for all levels of the marker and its contribution to explain the quantitative trait is not relevant or its effect is already considered in the model.

The complete birth step is built as follows:

1. Select a marker to allocate the new QTL from a Multinomial($1, (pb_1, \dots, pb_m)$), where $pb_j = KW_j / \sum_{j=1}^m KW_j$, $j = 1, \dots, m$, and KW_j is the statistics of the Kruskal–Wallis test from residuals of the current model and the j th marker genotype, defined as

$$KW_j = (n - 1) \frac{\sum_{l=1}^3 n_l (\bar{r}_l - \bar{r})^2}{\sum_{l=1}^3 \sum_{i=1}^{n_l} r_{li} - \bar{r}^2},$$

where n_l is the number of individuals in the l th group and the three groups are specified by the genotype of the j th marker, r_{li} is the rank (among all individuals) of the i th individual from the l th group, $\bar{r}_l = \sum_{i=1}^{n_l} r_{li} / n_l$, and $\bar{r} = 0.5(n + 1)$ is the average of all the r_{li} . Note that markers which most affect the residual mean are more likely to be chosen;

2. Assume the j^* th marker has been chosen, $j^* \neq 1$ and $j^* \neq m$, and suppose there is no QTL between the $(j^* - 1)$ and $(j^* + 1)$ th markers. The new QTL can be located in $[D_{j^*-1}, D_{j^*+1}]$ and λ_{K+1} is defined as $D_{j^*-1} + (D_{j^*+1} - D_{j^*-1})^*Z$, where $Z \sim \text{Beta}(a, 1)$ and a is calculated according to

$$E[Z] = \frac{\sum_{j=(j^*-1)}^{(j^*+1)} ((D_j - D_{j^*-1}) / (D_{j^*+1} - D_{j^*-1})) KW_j}{\sum_{j=(j^*-1)}^{(j^*+1)} KW_j};$$

$$\text{i.e., } a = \frac{E[Z]}{1 - E[Z]}.$$

Consequently, the expected value of λ_{K+1} is the average of the j^* th marker and its flanking markers' position weighted

Figure 2 Trace plot of K for $\sigma = 1.0$. (A) RJ sequence; (B) DDRJ sequence.

by their effect on the residual mean of the current model and the new QTL is more likely to be close to the marker that has the most relevant effect on the residual mean. Note the $\text{Beta}(a, 1)$ distribution is $\text{Uniform}(0, 1)$ when M_{j^*-1} , M_{j^*} , and M_{j^*+1} have the same effect on the residual mean and the j^* th marker is in the center of $[D_{j^*-1}, D_{j^*+1}]$.

If $j^* = 1$, $j^* = m$, $[D_{j^*-1}, D_{j^*}]$ or $[D_{j^*}, D_{j^*+1}]$ already contains a QTL then the new QTL will be located in $[D_1, D_2]$, $[D_{m-1}, D_m]$, $[D_{j^*}, D_{j^*+1}]$, or $[D_{j^*-1}, D_{j^*}]$, respectively, and its position is simulated as in step 2, considering only two markers and not three.

3. Sample the genotype of the new QTL for all individuals, $\mathbf{q}_{K+1}$, from

$$\Pr(Q_{iK+1} = q_{iK+1} | M_{iK+1}, M_{iL_{K+1}}, \mathbf{D}).$$

4. Sample α_{K+1} from its conditional *a posteriori* distribution considering $\mathbf{q}^b = (\mathbf{q}, \mathbf{q}_{K+1})$ and $\delta_{K+1} = 0$.
5. Sample δ_{K+1} from its conditional *a posteriori* distribution considering $\mathbf{q}^b$ and $\alpha^b = (\alpha, \alpha_{K+1})$.
6. Sample μ^b from its conditional *a posteriori* distribution, considering $\mathbf{q}^b$, α^b , and $\delta^b = (\delta, \delta_{K+1})$.
7. Sample σ^{2b} from its conditional *a posteriori* distribution, considering $\mathbf{q}^b$, α^b , δ^b , and μ^b .

Therefore, we have a new set of QTL genotypes and parameters $\mathbf{x}^b = (\mathbf{q}^b, \boldsymbol{\theta}^b)$. This transition proposal is denoted by $\mathbf{x}^b | \mathbf{x}$ and its probability is

$$\begin{aligned} q(\mathbf{x}^b | \mathbf{x}) &= \Pr(b|K) p_{b,j^*} f_Z(z) \\ &\times \prod_{i=1}^n (\Pr(Q_{iK+1} = q_{iK+1} | M_{iK+1}, M_{iL_{K+1}}, \mathbf{D})) \\ &\times \pi(\alpha_{K+1} | \mathbf{y}, \mathbf{q}^b, \boldsymbol{\theta}_{-K}, K+1, \lambda_{K+1}, \delta_{K+1}) \\ &\times \pi(\delta_{K+1} | \mathbf{y}, \mathbf{q}^b, \boldsymbol{\theta}_{-K}, K+1, \lambda_{K+1}, \alpha_{K+1}) \\ &\times \pi(\mu^b | \mathbf{y}, \mathbf{q}^b, \boldsymbol{\theta}_{-(\mu^b, \sigma^{2b})}^b, \sigma^2) \pi(\sigma^{2b} | \mathbf{y}, \mathbf{q}^b, \boldsymbol{\theta}_{-\sigma^{2b}}^b), \end{aligned} \quad (8)$$

where $\pi(\cdot | \cdot)$ is the conditional *a posteriori* distribution for each parameter used to sample the candidate values. The acceptance probability for the birth move is $\Psi(\mathbf{x}^b | \mathbf{x}) = \min(1, A^b)$, where A^b is given by Equation 7. The probability of the transition proposal denoted by $\mathbf{x} | \mathbf{x}^b$ is

$$\begin{aligned} q(\mathbf{x} | \mathbf{x}^b) &= \Pr(d|K+1) p_{d,K+1} \pi(\mu | \mathbf{y}, \mathbf{q}, \boldsymbol{\theta}_{-(\mu, \sigma^2)}, \sigma^{2b}) \\ &\times \pi(\sigma^2 | \mathbf{y}, \mathbf{q}, \boldsymbol{\theta}_{-\sigma^2}). \end{aligned} \quad (9)$$

Death proposal: Since a death move has been selected, we choose a QTL from the current model to be deleted.

As Q_{ik} assumes only values $-1, 0$, and 1 and $(1 - |Q_{ik}|)$ assumes only 0 and 1 , for $i = 1, \dots, n$ and $k = 1, \dots, K$, the current absolute value of α_k and δ_k shows the importance and significance of the k th QTL, *i.e.*, higher absolute values of α_k or δ_k indicate the k th QTL is more relevant to explain the phenotype. The current values of these parameters are useful for the choice of the QTL to be excluded without changing significantly the predictive power of the model.

Instead of selecting a QTL to be excluded from a $\text{Uniform}(1, \dots, K)$, we select it from a $\text{Multinomial}(1, (pd_1, \dots, pd_K))$, where $pd_k = (1/(|\alpha_k| + |\delta_k|)) / \sum_{k=1}^K (1/(|\alpha_k| + |\delta_k|))$, for $k = 1, \dots, K$; *i.e.*, QTLs that exert the strongest effects and are the most relevant to the model are less likely to be selected and deleted. Therefore, the acceptance probability of the death movement is improved.

The complete death step is as follows:

1. Select the QTL to be excluded from $\text{Multinomial}(1, (pd_1, \dots, pd_K))$, the k^* th QTL.
2. Delete $\mathbf{q}_k^*$, λ_k^* , α_k^* , and δ_k^* from $\mathbf{q}$, λ , α , and δ , respectively.
3. Sample μ^d from its conditional *a posteriori* distribution, considering only $K - 1$ QTLs.
4. Sample σ^{2d} from its conditional *a posteriori* distribution, considering the reduced model.

Figure 3 Trace plot of K for $\sigma = 1.5$. (A) RJ sequence; (B) DDRJ sequence.

We have a new set of QTL's genotypes and parameters $x^d = (\mathbf{q}^d, \boldsymbol{\theta}^d = (K-1, \boldsymbol{\lambda}^d, \boldsymbol{\alpha}^d, \boldsymbol{\delta}^d, \boldsymbol{\mu}^d, \sigma^{2d}))$. This transition proposal is denoted by $x^d|x$ and its probability is

$$q(x^d|x) = \Pr(d|K)pd_{k^*}\pi(\boldsymbol{\mu}^d|\mathbf{y}, \mathbf{q}^d, \boldsymbol{\theta}^d_{-(\boldsymbol{\mu}^d, \sigma^{2d})}, \sigma^2) \times \pi(\sigma^{2d}|\mathbf{y}, \mathbf{q}^d, \boldsymbol{\theta}^d_{-\sigma^{2d}}), \quad (10)$$

where $\pi(\cdot|\cdot)$ is the conditional *a posteriori* distribution of each parameter used to generate the candidate values.

The acceptance probability for the death movement is $\Psi(x^d|x) = \min(1, A^d)$, where $A^d = 1/A^b$ with some suitable substitutions. The probability of transition proposal denoted by $x|x^d$ is defined as

$$q(x|x^d) = \Pr(b|K-1) \left(pb_{l_{k^*}fz} \left(\frac{\lambda_{k^*} - D_{l_{k^*}-1}}{D_{l_{k^*}+1} - D_{l_{k^*}-1}} \right) + pb_{r_{k^*}fz} \left(\frac{\lambda_{k^*} - D_{r_{k^*}-1}}{D_{r_{k^*}+1} - D_{r_{k^*}-1}} \right) \right) \times \prod_{i=1}^n \left(\Pr(Q_{ik^*} = q_{ik^*} | M_{i_{l_{k^*}}}, M_{i_{r_{k^*}}}, \mathbf{D}, \lambda_{k^*}) \right) \times \pi(\alpha_{k^*}|\mathbf{y}, \mathbf{q}, \boldsymbol{\theta}^d_{-(K-1)}, K, \lambda_{k^*}, \delta_{k^*}) \times \pi(\delta_{k^*}|\mathbf{y}, \mathbf{q}, \boldsymbol{\theta}^d_{-(K-1)}, K, \lambda_{k^*}, \alpha_{k^*}) \times \pi(\boldsymbol{\mu}|\mathbf{y}, \mathbf{q}, \boldsymbol{\theta}_{-(\boldsymbol{\mu}, \sigma^2)}, \sigma^{2d}) \pi(\sigma^2|\mathbf{y}, \mathbf{q}, \boldsymbol{\theta}_{-\sigma^2}), \quad (11)$$

where l_{k^*} is the marker on the left of the k^* th QTL and r_{k^*} is the marker on the right of the k^* th QTL.

Note that if we first choose a birth movement in state x , giving x^b , and then choose the death of the $(K+1)$ th QTL, we can recover x and state x is likely to be recovered after a birth process of x^d . If the candidate movement is not accepted, the chain remains in the current model, the value of K does not change, and the remaining parameters

of the model are updated by Metropolis–Hastings or Gibbs steps.

Merge proposal: Instead of proposing data driven with only birth and death steps, we also include a merge movement in the procedure since the model can be wrongly estimated if the effect of a true QTL is split between two or more false QTLs. The split of a QTL may happen if a QTL appears very close to an existent QTL and, as their genotypes are very similar, both are in the model and split the additive and dominance effect that would be of only one QTL. The death of one of these QTLs is not generally accepted since the effects of both QTLs are relevant to explain the phenotype variability. The merge moves of two consecutive QTLs are usually accepted and effective to avoid split QTLs since the effect of the QTL that is removed from the model is added to the effect of an adjacent QTL and the predictive power of the model does not change significantly.

For merging two QTLs we must choose a pair of consecutive QTLs to be merged and choose one QTL to be removed from the model. Its effects are added to the effect of the other QTL. We propose to build a data-driven merge candidate as follows:

1. Select a pair of consecutive QTLs to be merged from $\text{Multinomial}(1, (\text{pmg}_{12}, \text{pmg}_{23}, \dots, \text{pmg}_{(K-1)K}))$, where $\text{pmg}_{kj} = V_{kj} / \sum_{k=1}^{K-1} \sum_{j=k+1}^K V_{kj}$, $k = 1, \dots, K-1$ and $j = k+1, \dots, K$, and V_{kj} is Cramér's V measure of association between the genotypes of the k th and the j th QTLs. Note that pairs of successive QTLs with more associated genotypes have higher probability to be merged since the split happens between QTLs with similar genotype. Suppose the pair of QTLs k^* and k^*+1 has been selected.
2. Choose the k^* th or (k^*+1) th to be excluded from the current model, according to $pd_k = (1/(|\alpha_k| + |\delta_k|)) / \sum_{k=k^*}^{k^*+1} (1/(|\alpha_k| + |\delta_k|))$, $k = k^*, k^*+1$. Consider that the (k^*+1) th has been chosen to be excluded.

Table 1 ESS of K sequences

Error variability	RJ	DDRJ
$\sigma = 0.5$	3	357
$\sigma = 1.0$	159	330
$\sigma = 1.5$	445	894

3. Delete $\mathbf{q}_{k^*+1}$, λ_{k^*+1} , α_{k^*+1} , and δ_{k^*+1} from $\mathbf{q}$, λ , α , and δ , respectively.
4. Update α_{k^*} , δ_{k^*} , μ , and σ^2 , successively, from their conditional *a posteriori* distribution considering $\mathbf{q}^{\text{mg}}$, α^{mg} , and δ^{mg} with $k - 1$ QTLs.

Instead of adding the value of α_{k^*+1} and δ_{k^*+1} to α_{k^*} and δ_{k^*} , respectively, we propose to update α_{k^*} and δ_{k^*} from their conditional *a posteriori* probability, using the reduced model. It is equivalent since we remove the effects of the $(k^* + 1)$ th QTL from the current model to update α_{k^*} and δ_{k^*} and simplify the calculation of merge acceptance probability since is not necessary to define deterministic transformations to reduce the dimension of the model.

We have a new set of QTL's genotypes and parameters $x^{\text{mg}} = (\mathbf{q}^{\text{mg}}, \boldsymbol{\theta}^{\text{mg}} = (K - 1, \lambda^{\text{mg}}, \alpha^{\text{mg}}, \delta^{\text{mg}}, \mu^{\text{mg}}, \sigma^{2\text{mg}}))$. This transition proposal is denoted by $x^{\text{mg}}|x$ and its probability is

$$\begin{aligned}
 q(x^{\text{mg}}|x) &= \text{pmg}_{k^*(k^*+1)} p d_{k^*+1} \\
 &\times \pi(\alpha_{k^*+1} | \mathbf{y}, \mathbf{q}^{\text{mg}}, K - 1, \lambda^{\text{mg}}, \alpha^{\text{mg}}_{-\alpha_{k^*}}, \delta^{\text{mg}}, \mu, \sigma^2) \\
 &\times \pi(\delta_{k^*+1} | \mathbf{y}, \mathbf{q}^{\text{mg}}, K - 1, \lambda^{\text{mg}}, \alpha^{\text{mg}}, \delta^{\text{mg}}_{-\delta_{k^*}}, \mu, \sigma^2) \\
 &\times \pi(\mu | \mathbf{y}, \mathbf{q}^{\text{mg}}, \boldsymbol{\theta}^{\text{mg}}_{-(\mu^{\text{mg}}, \sigma^{2\text{mg}})}, \sigma^2) \\
 &\times \pi(\sigma^2 | \mathbf{y}, \mathbf{q}^{\text{mg}}, \boldsymbol{\theta}^{\text{mg}}_{-\sigma^{2\text{mg}}}), \quad (12)
 \end{aligned}$$

where $\pi(\cdot|\cdot)$ is the conditional *a posteriori* distribution of each parameter used to sample the candidate values.

The acceptance probability for the merge movement is $\Psi(x^{\text{mg}}|x) = \min(1, A^{\text{mg}})$, where A^{mg} is defined by Equation 7. The probability of a transition proposal denoted by $x|x^{\text{mg}}$ that represents a split of the k^* th QTL is defined as

$$\begin{aligned}
 q(x|x^{\text{mg}}) &= \left(p b_{l_{k^*+1}} f_Z \left(\frac{\lambda_{k^*+1} - D_{l_{k^*+1}-1}}{D_{l_{k^*+1}+1} - D_{l_{k^*+1}-1}} \right) \right. \\
 &\quad \left. + p b_{r_{k^*+1}} f_Z \left(\frac{\lambda_{k^*+1} - D_{r_{k^*+1}-1}}{D_{r_{k^*+1}+1} - D_{r_{k^*+1}-1}} \right) \right) \\
 &\times \prod_{i=1}^n (\Pr(Q_{ik^*+1} = q_{ik^*+1} | M_{ir_{k^*+1}}, M_{il_{k^*+1}}, \mathbf{D}, \lambda_{k^*+1}))
 \end{aligned}$$

Table 2 *A posteriori* probability for K

K	σ					
	0.5		1.0		1.5	
	RJ	DDRJ	RJ	DDRJ	RJ	DDRJ
1	0.000	0.000	0.000	0.000	0.308	0.016
2	0.000	0.000	0.268	0.007	0.490	0.191
3	0.443	0.000	0.724	0.914	0.201	0.706
4	0.002	0.000	0.007	0.075	0.001	0.081
5	0.554	0.971	0.001	0.004	0.000	0.005
6	0.001	0.028	0.000	0.000	0.000	<0.001
7	0.000	0.001	0.000	0.000	0.000	0.000

The highest *a posteriori* probability estimate for each case is in boldface type.

$$\begin{aligned}
 &\times \pi(\alpha_{k^*+1} | \mathbf{y}, \mathbf{q}, \boldsymbol{\theta}^{\text{mg}}_{-(K-1)}, K, \lambda_{k^*+1}, \delta_{k^*+1} = 0) \\
 &\times \pi(\delta_{k^*+1} | \mathbf{y}, \mathbf{q}, \boldsymbol{\theta}^{\text{mg}}_{-(K-1)}, K, \lambda_{k^*+1}, \alpha_{k^*+1}) \\
 &\times \pi(\alpha_{k^*} | \mathbf{y}, \mathbf{q}, \boldsymbol{\theta}_{-\alpha_{k^*}}) \pi(\delta_{k^*} | \mathbf{y}, \mathbf{q}, \boldsymbol{\theta}_{-\delta_{k^*}}) \\
 &\times \pi(\mu | \mathbf{y}, \mathbf{q}, \boldsymbol{\theta}_{-(\mu, \sigma^2)}, \sigma^{2\text{mg}}) \pi(\sigma^2 | \mathbf{y}, \mathbf{q}, \boldsymbol{\theta}_{-\sigma^2}), \quad (13)
 \end{aligned}$$

where l_{k^*+1} is the marker on the left of the $(k^* + 1)$ th QTL and r_{k^*+1} is the marker on the right of the $(k^* + 1)$ th QTL.

Since we include the QTL merge move only to avoid split QTLs, we do not include a QTL split step in this procedure. However, a split step could be easily included in the algorithm, using the transition function of a split movement $q(x^{\text{sp}}|x) = q(x|x^{\text{mg}})$ defined in Equation 13.

Algorithm: The birth–death–merge DDRJ is specified as follows:

1. Initialize a configuration for $\boldsymbol{\theta}$ and $\mathbf{q}$.
2. For the l th iteration, $l = 1, \dots, L$,
 - a. Choose a death or birth movement.
 - b. Generate the candidate values of x' .
 - c. Accept the proposal with probability $\Psi(x'|x)$, where ' $\cdot$ ' means either b or d .
 - i. If a birth movement has been accepted, do $K^{(l)} = K^{(l-1)} + 1$ and consider x^b .
 - ii. If a death movement has been accepted, do $K^{(l)} = K^{(l-1)} - 1$ and consider x^d .
 - iii. If no movement has been accepted, do $K^{(l)} = K^{(l-1)}$ and consider x .
 - d. If $K^{(l)} \geq 2$, generate and evaluate the acceptance of a merge of a QTLs pair. If a merge movement has been accepted, do $K^{(l)} = K^{(l)} - 1$ and consider x^{mg} .
 - e. Update λ_k , $k = 1, \dots, K^{(l)}$.
 - f. Update Q_{ik} , $i = 1, \dots, n$ and $k = 1, \dots, K^{(l)}$, from its conditional *a posteriori* distribution.
 - g. Update α_k and δ_k , $k = 1, \dots, K^{(l)}$, from their conditional *a posteriori* distributions.

- h. Update μ from its conditional *a posteriori* distribution.
- i. Update σ^2 from its conditional *a posteriori* distribution.

This algorithm is implemented in R language and the codes are available in [File S3](#) and [File S4](#). R is a free software environment for statistical computing and graphics and more details are found in its homepage, "<https://www.r-project.org>".

Applications

We apply the proposed method to simulated and real data sets and compare the performance of the RJ, DDRJ, and MIM methodologies. Although the computational efficiency is an important feature of the methods, we focus on analyzing and comparing their performance in selecting and estimating the correct model. We set hyperparameters $\nu_\alpha = \nu_\delta = \nu_\mu = 0$, $\sigma_\alpha^2 = \sigma_\delta^2 = \sigma_\mu^2 = 100$, and $\eta_a = \eta_b = 0.1$. This setup provides *a priori* distributions with large variability and weak information about the parameters.

Simulated data sets

We simulate a high-dimension linkage map with 450 loci that are allocated on a large chromosome of 450 cM (average distance between the loci is 1 cM) and their genotype for an F_2 population of 300 individuals by QTL Cartographer 2.5 software available at "<http://statgen.ncsu.edu/qtlcart/WQTLCart.htm>" (Basten *et al.* 1997). We choose $K = 5$ loci located at $\lambda = \{15.0, 82.4, 299.8, 363.1, 391.1\}$ to be the QTLs and simulate the phenotype using $\alpha = (-0.60, 0.90, 0.25, -0.40, 0.40)$, $\delta = (0.30, 0.05, -0.25, 0.15, -0.15)$, $\mu = 20$, and three values of σ (0.5, 1.0, 1.5). The effects of the first and the second QTLs are stronger and are easily identified, the fourth and fifth QTLs have opposite effects, and the effect of the third QTL is the weakest.

We run RJ and DDRJ chains $L = 55,000$ iterations, discard the first 5000 iterations, and take one for every 10 iterations. The chains are initialized with $K = 0$. Convergence is verified using trace plots.

Figure 1, Figure 2, and Figure 3 show the RJ and DDRJ trace plots of K for $\sigma = 0.5, 1.0$, and 1.5 , respectively. We observe DDRJ chains show better mixing since they easily move around the models space throughout the chain as a consequence of better proposal candidates. The RJ chain moves with greater difficulty between the possible models and it can get stuck in a specific model for longer periods even if it is a wrong model. When $\sigma = 0.5$, we observe a very poor mixing of the RJ chain since it gets stuck for long periods (at the beginning and end of the chain) in the model with $K = 3$ (wrong model). When $\sigma = 1.0$, the RJ chain moves easily around the models space in the beginning of the chain but not in its end.

We also analyze the mixing of the chains by their effective sample size (ESS) (Kass *et al.* 1998), which is the number of effectively independent draws from the *a posteriori* distribution.

Table 3 A *posteriori* estimates of the model parameters

Real value	$\sigma = 0.5$			$\sigma = 1.0$			$\sigma = 1.5$		
	RJ	DDRJ		RJ	DDRJ		RJ	DDRJ	
$\lambda_1 = 15.0$	16.3 (16.0; 16.8)	13.8 (13.4; 14.0)		16.4 (16.0; 17.0)	16.5 (13.4; 17.9)		45.3 (19.7; 80.4)	15.4 (13.0; 23.7)	
$\lambda_2 = 82.4$	82.6 (81.9; 83.2)	83.5 (83.4; 83.9)		80.9 (77.4; 87.2)	82.8 (81.8; 83.4)		178.5 (81.4; 302.6)	82.7 (77.7; 87.8)	
$\lambda_3 = 299.8$	295.4 (294.8; 295.8)	299.1 (295.9; 302.7)		296.4 (291.2; 302.9)	296.0 (292.6; 302.4)			294.9 (289.4; 303.1)	
$\lambda_4 = 363.1$	363.1 (362.2; 364.0)	361.6 (361.1; 362.1)							
$\lambda_5 = 391.1$	387.8 (386.2; 402.6)	389.7 (389.2; 390.1.6)							
$\mu = 20.0$	19.92 (19.79; 20.05)	19.99 (18.86; 20.12)		19.94 (19.70; 20.18)	19.93 (19.71; 20.16)		20.00 (19.70; 20.29)	19.92 (19.65; 20.20)	
$\alpha_1 = -0.60$	-0.60 (-0.68; -0.52)	-0.59 (-0.67; -0.51)		-0.62 (-0.79; -0.44)	-0.59 (-0.75; -0.41)		0.03 (-0.82; 0.98)	-0.59 (-0.80; -0.28)	
$\alpha_2 = -0.90$	0.93 (0.84; 1.01)	0.91 (0.83; 0.99)		0.91 (0.74; 1.09)	0.97 (0.80; 1.13)		0.78 (0.39; 1.14)	0.96 (0.75; 1.18)	
$\alpha_3 = 0.25$	0.36 (0.27; 0.45)	0.37 (0.29; 0.46)		0.44 (0.29; 0.62)	0.45 (0.28; 0.61)			0.57 (0.27; 0.78)	
$\alpha_4 = -0.40$	-0.37 (-0.49; -0.25)	-0.41 (-0.51; -0.30)							
$\alpha_5 = 0.40$	0.33 (0.21; 0.44)	0.38 (0.28; 0.48)							
$\delta_1 = 0.30$	0.24 (0.13; 0.35)	0.24 (0.13; 0.36)		0.14 (-0.08; 0.38)	0.13 (-0.10; 0.37)		0.07 (-0.32; 0.44)	0.10 (-0.19; 0.39)	
$\delta_2 = 0.05$	0.02 (-0.10; 0.13)	0.01 (-0.10; 0.12)		-0.02 (-0.32; 0.25)	0.03 (-0.20; 0.26)		-0.15 (-0.50; 0.21)	-0.04 (-0.37; 0.26)	
$\delta_3 = -0.25$	-0.15 (-0.27; -0.02)	-0.16 (-0.28; -0.03)		-0.05 (-0.31; 0.23)	-0.05 (-0.32; 0.19)			0.04 (-0.29; 0.37)	
$\delta_4 = 0.15$	0.17 (0.04; 0.31)	0.15 (0.03; 0.28)							
$\delta_5 = -0.15$	-0.11 (-0.24; 0.03)	-0.15 (-0.27; -0.02)							
σ	0.50 (0.46; 0.54)	0.49 (0.45; 0.53)		1.02 (0.93; 1.10)	0.98 (0.91; 1.06)		1.49 (1.38; 1.62)	1.44 (1.33; 1.57)	

Table 4 MIM estimates of the parameters

Parameter	Real value	$\sigma = 0.5$	$\sigma = 1.0$	$\sigma = 1.5$
λ	(15.0, 82.4, 299.8, 363.1, 391.1)	(9.0, 14.0, 83.4, 298.8, 363.1, 390.1)	(14.0, 83.4, 293.8)	(14.0, 83.4, 293.8)
α	(-0.60, 0.90, 0.25, -0.40, 0.40)	(0.24, -0.80, 0.89, 0.40, -0.43, 0.40)	(-0.58, 0.96, 0.47)	(-0.59, 0.98, 0.61)
δ	(0.30, 0.05, -0.25, 0.15, -0.15)	(-0.21, 0.42, -0.01, -0.19, 0.18, -0.13)	(0.18, 0.01, -0.001)	(0.15, -0.02, 0.07)

A large discrepancy between the ESS and the simulation sample size indicates poor mixing. Table 1 shows the ESS for the RJ and DDRJ K sequences and we observe the DDRJ ESS is larger than the RJ ESS, which confirms a better mixing of DDRJ chains. We observe a very poor mixing of the RJ chain mainly for $\sigma = 0.5$. DDRJ and RJ ESSs of the remaining parameters of the models are shown in Table A of File S2 and DDRJ ESSs are in most cases larger than RJ ESSs.

Table 2 shows *a posteriori* probabilities for K calculated as the relative frequency of each value of K in the sequence. The highest *a posteriori* probability estimate for each situation is in boldface type and the argument that maximizes this probability is the estimate of K . In situations where the genetic effects of QTLs are strong compared with the size of the error variability ($\sigma = 0.5$) both methodologies estimate correctly $K = 5$. However, as a result of weak mixing, the RJ chain gets stuck in $K = 3$ for long periods and tends to underestimate the *a posteriori* probability of K . Since $\sigma = 0.5$ represents a small variability of the random error and, consequently, the effect of QTLs is more evident, the choice of the correct model should be precise. When $\sigma \geq 1.0$, the opposite fourth and fifth QTLs, although they have higher additive effect than the third QTL, are not identified by RJ and DDRJ since their effects cancel each other. For $\sigma = 1.5$, the RJ procedure estimates only $K = 2$ and shows greater difficulties in locating the QTLs.

Table 3 shows the estimates (*a posteriori* average) of parameters and their 95% credibility interval. The estimates of both methodologies are similar when $\sigma = 0.5$ and close to the true values. The DDRJ point estimates of the additive and the dominance effect of the fourth and fifth QTLs are closer to the true simulated parameters than the RJ estimates. Zero belongs to the RJ credibility interval of δ_5 . The additive and dominance effects of the third QTL are the worst estimate in both methods. When $\sigma = 1.0$, RJ and DDRJ estimates for the model with $K = 3$ QTLs are similar and the additive and dominance effects estimates of the third QTL are also the worst estimate in both methods. For $\sigma = 1.5$, RJ shows a low performance to estimate the number of QTLs and the parameters associated with them. The RJ point estimates are different from the parameters and interval estimates are large.

We also analyze the simulated data sets, using the MIM method available in QTL Cartographer. The main model selection criterion available in QTL Cartographer to select the number of QTLs is $BIC = -2\log(L(\hat{\theta}|\mathbf{y}, \mathbf{q})) + pc(n)$, where $\hat{\theta}$ is the maximum-likelihood estimator of θ , p is the number of free parameters to be estimated, and $c(n) = \log(n)$.

Other definitions of $c(n)$ are used and available in QTL Cartographer such as $c(n) = 2$ (AIC), $c(n) = 2\log(\log(n))$, $c(n) = 2\log(n)$, $c(n) = 3\log(n)$, and $c(n) = 10X\log(n)$, where we define $X = 0.01$. We choose the MIM forward search method to estimate the initial model and test the six model selection criteria to optimize QTLs positions, search for new QTLs, and test existing QTLs. We report the results of $c(n) = \log(n)$, which shows the best results for the simulated data sets.

The MIM method combined with BIC model selection methodologies and optimization procedures of QTL location and effect estimates $K = 6, 3, 3$ for $\sigma = 0.5, 1.0$, and 1.5 , respectively. Table 4 shows the MIM estimates of the remaining parameters of the models. The method identifies one non-existing QTL at 9.0 cM when $\sigma = 0.5$ and the additive and dominance effects of the second QTL are biased. We observe that if we sum the estimates of additive and dominance effects of first and second QTLs, we have estimates closer to additive and dominance effects of the QTL located at 15.0 cM; that is, the effects of the true QTL estimated at 14 cM are split with a false QTL identified at 9 cM. When $\sigma = 1.0$ and 1.5 , the opposite fourth and fifth QTLs are not identified and the DDRJ estimates of the remaining parameters, especially estimates associated with the third QTL that has weaker effects, are better than MIM estimates. We do not have a confidence interval to analyze the uncertainty of the parameters.

Unlike BIC ($c(n) = \log(n)$), we stop the AIC, BIC-like criteria with $c(n) = 2\log(\log(n))$ and $c(n) = 0.1\log(n)$ estimation when they wrongly identify $K = 12, 9$, and 9 significant QTLs for $\sigma = 0.5, 1.0$, and 1.5 located at $\hat{\lambda} = \{9.0, 14.0, 83.4, 86.4, 91.5, 298.8, 339.8, 351.2, 360.2, 363.1, 388.2, 390.1\}$, $\hat{\lambda} = \{10.0, 14.0, 83.4, 293.8, 301.8, 309.8, 337.8, 388.1, 390.1\}$, and $\hat{\lambda} = \{3.0, 9.0, 14.0, 83.4, 86.4, 91.5, 293.8, 338.8, 410.1, 390.1\}$, respectively. The BIC-like criterion with $c(n) = 2\log(n)$ estimates $K = 3$ significant QTLs located at $\hat{\lambda} = \{14.0, 83.4, 293.8\}$ for all values of σ and the BIC-like criterion with $c(n) = 3\log(n)$ estimates $K = 3$ QTLs located at $\hat{\lambda} = \{14.0, 83.4, 296.8\}$ for $\sigma = 0.5$, $K = 2$ QTLs located at $\hat{\lambda} = \{15.0, 83.4\}$ for $\sigma = 1.0$, and $K = 1$ significant QTL located at $\hat{\lambda} = 83.5$ for $\sigma = 1.5$. Therefore, we observe the MIM method combined with BIC model selection is sensitive to $c(n)$ choice for these simulated data sets; that is, depending on the $c(n)$ choice, the method overestimates or underestimates the number of QTLs. If the data were not simulated and we did not know the correct model, we could estimate the model by the six MIM model selection criteria and select the estimated model that was the most frequent

Table 5 DDRJ *A posteriori* DDRJ probability for K in each chromosome

K	Chromosome									
	1	2	3	4	5	6	7	8	9	10
0	0.60	0.93	0.90	0.86	0.95	0.92	0.30	0.85	0.63	0.76
1	0.25	0.06	0.09	0.11	0.04	0.06	0.67	0.12	0.33	0.17
2	0.13	0.01	0.01	0.03	0.01	0.02	0.03	0.03	0.03	0.06
≥ 3	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01
K	Chromosome									
	11	12	13	14	15	16	17	18	19	
0	0.47	0.79	0.88	0.91	0.92	0.94	0.76	0.82	0.59	
1	0.42	0.18	0.11	0.08	0.07	0.05	0.21	0.16	0.38	
2	0.10	0.03	0.01	0.01	0.01	0.01	0.02	0.02	0.03	
≥ 3	0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	

between all criteria. In this case, we would choose, for all values of σ , the model estimated by the AIC, BIC-like criterion with $c(n) = 2\log(\log(n))$ and $c(n) = 0.1\log(n)$, which is the worst estimated model.

Real data set

We apply RJ and DDRJ to the bone mineral density data set. It consists of 661 female F_2 mice derived from matings of F_1 individuals from NZB/B1NJ $\times$ RF/J parents. This cross is designed to identify the genetic loci regulating femur mechanical properties, geometric properties, and bone mineral density (BMD). The data have 94 genetic markers located in 19 chromosomes. NZB, RF, and heterozygous markers are coded as 1, -1 , and 0, respectively. The data were downloaded from the site “<http://qtlarchive.org/db/q?pg=projlist>”.

Twenty-three phenotypes were measured in all individuals; however, we analyze only the total femur volumetric BMD in milligrams per cubic centimeter. The trait was log-transformed before analysis to be comparable with Wergedal *et al.* (2006) and Cox *et al.* (2009) results.

We run $L = 110,000$ RJ iterations, discard the first 10,000 and take one for every 10 iterations. We run $L = 55,000$ DDRJ iterations, discard the first 5000, and take one for every 10 iterations. The sequences are initialized with $K = 0$ and, in DDRJ, we update the birth candidate 10 times before evaluating its acceptance, as proposed by Green and Mira (2001). We analyze the convergence and conclude the number of iterations is sufficient for reliable results.

Table 5 shows the *a posteriori* DDRJ probability (relative frequency) for K in each chromosome whose value is evidence of a QTL presence. The *a posteriori* probability of the model with one QTL is 0.67 in chromosome 7, 0.42 in chromosome 11, 0.38 in chromosome 19, 0.33 in chromosome 9, and 0.25 in chromosome 1, which represents strong evidence of a QTL in chromosome 7 since $K = 1$ is the argument that maximizes the *a posteriori* probability of K and moderate probability in chromosomes 1, 9, 11, and 19 since, despite that the maximum *a posteriori* probability is not for $K = 1$, it is >0.25 . In chromosomes 10, 12, 17, and 18, the probability of a QTL is not negligible. Depending on the cost and researcher interest, these loci can be studied in more detail. Therefore, we identify at least $K = 5$ QTLs regulating bone mineral density.

Table 6 shows estimates and 95% credibility intervals for QTLs' locations (centimorgans) and additive and dominance effects in chromosomes 1, 7, 9, 10, 11, 12, 17, 18, and 19. Additive and dominance effects explain how QTLs genotypes are associated to bone mineral density and their estimates are small (close to zero) because of the scale of the $\log(\text{BMD})$. Although the chance of a QTL in chromosomes 10, 17, and 19 is not negligible, zero belongs to their additive and dominance effects 95% credibility interval. Therefore, DDRJ identifies relevant QTLs at chromosomes 1, 7, 9, 11, 12, and 18.

We also analyze these data by a RJ and MIM forward search method combined with BIC model selection ($c(n) = \log(n)$), which shows better results in the simulated data sets. We observe only RJ low *a posteriori* probabilities 0.0006, 0.0009, and 0.027 for one QTL in chromosomes 7, 9, and 11, respectively. MIM identifies one QTL in chromosomes 1, 7, 9, 11, and 12 located at 88, 65, 70, 34, and 28 cM, respectively. The MIM point estimates of additive and dominance effects are $\hat{\alpha} = (0.009, 0.009, 0.012, -0.014, 0.009)$ and $\hat{\delta} = (0.008, 0.016, -0.005, -0.004, 0.004)$. The MIM effect estimates are close to DDRJ estimates; however, we do not have information about MIM estimates uncertainty. Wergedal *et al.* (2006) use a three-stage strategy and LOD score to identify $K = 5$ QTLs located in chromosomes 3, 7, 10, 11, and 18 at 10-, 65-, 65-, 40-, and 50-cM positions, respectively.

If we use the DDRJ *a posteriori* probability of K as evidence of QTL presence, we observe DDRJ, MIM, and Wergedal methodologies identify QTLs in chromosomes 7 and 11; DDRJ and MIM identify three more QTLs in chromosomes 1, 9, and 12; and DDRJ and Wergedal methods identify another QTL in chromosome 18. The Wergedal method also identifies one QTL in chromosomes 3 and 10 whose credibility interval of additive effect and dominance effect contains zero. Therefore, for this data set, DDRJ methodology identifies QTLs with strong and weak effects in BMD that are not identified by other QTL mapping methods.

Discussion

We propose a birth–death–merge DDRJ for QTL mapping in an F_2 population with an unknown number of QTLs. We compare the performance of the proposed method with

Table 6 DDRJ estimates and 95% credibility intervals of parameters

Chromosome	λ	α	δ
1	84.1 (52.8; 99.9)	0.008 (0.001; 0.013)	0.009 (−0.001; 0.002)
7	63.5 (48.3; 68.6)	0.009 (0.003; 0.014)	0.015 (0.006; 0.023)
9	64.4 (45.4; 70.8)	0.011 (0.006; 0.017)	−0.006 (−0.015; 0.005)
10	60.4 (47.0; 64.7)	0.003 (−0.003; 0.009)	0.003 (−0.006; 0.010)
11	32.5 (21.9; 43.1)	−0.013 (−0.019; −0.008)	−0.002 (−0.011; 0.007)
12	30.7 (5.8; 57.6)	0.007 (0.001; 0.013)	0.001 (−0.012; 0.015)
17	35.2 (18.0; 54.2)	0.002 (−0.004; 0.008)	−0.009 (−0.017; 0.001)
18	44.9 (30.7; 55.7)	−0.008 (−0.014; −0.003)	0.005 (−0.009; 0.015)
19	43.5 (28.5; 51.4)	−0.0002 (−0.005; 0.005)	−0.005 (−0.014; 0.005)

traditional RJ and MIM combined with a model selection method and optimization procedures that are the most popular methodologies for QTL mapping in experimental crosses. Although computational efficiency is an important feature of the methods, we focus on analyzing and comparing their performance in identifying significant QTL regions.

DDRJ shows a better performance to identify and estimate QTLs mainly when their effects are moderate and RJ does not identify them. The better performance of DDRJ occurs because it facilitates the moves around the models space and improves the chain mixing as a consequence of better proposals in transdimensional moves. Unlike DDRJ, the RJ method moves with greater difficulty between the possible models and it can get stuck in a specific model for longer periods even if it is a wrong model. Compared with MIM combined with model selection methods, DDRJ also shows better performance in identifying QTL regions and provides uncertainty information for all parameters through credibility intervals. For simulated data sets, MIM shows sensitivity to the choice of model selection criterion and, depending on the criterion choice, the method overestimates or underestimates the number of QTLs. As QTLs single effects are not so high in practice, mainly the effect of SNP QTLs (Yang *et al.* (2010)), the proposed methodology appears to be useful and brings contributions to identification and characterization of QTLs.

The DDRJ *a posteriori* probability of K is evidence of QTL presence and, even when this value is not maximum for $K > 0$, it allows us to specify regions that can be further explored by genetic researchers. The application in a real data set illustrates an example where DDRJ identifies QTLs with strong, moderate, and weak effects on the phenotype that are not identified by RJ, MIM, or other QTL mapping methods.

The inclusion of merge moves in DDRJ is efficient under analyzed data sets to avoid the split of a true QTL effect with one or more false QTLs. The conventional methodologies usually deal with a ghost QTL that appears between two or more QTLs linked in coupling and is generally more significant than the true QTLs. The problem presented here is the opposite of that of a ghost QTL since the true QTLs share their importance with one or more false QTLs. Ghost QTLs are usually avoided by multiple-QTL mapping methods and merge moves included in DDRJ reduce the chance of split QTLs. Since we include the QTLs merge move only to avoid split QTLs, we do not include a QTL split step in this procedure.

The R codes of birth–death–merge data-driven reversible jump are available in [File S3](#) and [File S4](#) and we are improving them to be more efficient and user friendly.

The amplitude of the DDRJ credibility interval of QTLs' location is large when error variability is higher. To improve the DDRJ performance, we can estimate the genotype of a QTL using more than the two flanking markers or using nonconjugate samplers, and analyze the results in future work. The proposed data-driven method can be extended to generalized linear models and identifies QTLs that affect binary or discrete phenotypes or for QTL mapping in pedigree data in which the individuals' genotype is correlated if they are relatives and improves SNP mapping methods that have a smaller single effect on the phenotype.

Acknowledgment

The authors thank two referees for useful comments and suggestions that improved the manuscript.

Literature Cited

- Basten, C. J., B. S. Weir, and Z.-B. Zeng, 1997 *QTL Cartographer: A Reference Manual and Tutorial for QTL Mapping*. Department of Statistics, North Carolina State University, Raleigh, NC.
- Broman, K. W., and T. Speed, 1999 *A Review of Methods for Identifying QTLs in Experimental Crosses* (Lecture Notes-Monograph Series), pp. 114–142. Hayward, California.
- Cox, A., C. L. Ackert-Bicknell, B. L. Dumont, Y. Ding, J. T. Bell *et al.*, 2009 A new standard genetic map for the laboratory mouse. *Genetics* 182: 1335–1344.
- Dempster, A. P., N. M. Laird, and D. B. Rubin, 1977 Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. B Methodol.* 39: 1–38.
- Green, P. J., 1995 Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* 82(4): 711–732.
- Green, P. J., and A. Mira, 2001 Delayed rejection in reversible jump Metropolis–Hastings. *Biometrika* 88(4): 1035–1053.
- Haley, C. S., and S. A. Knott, 1992 A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* 69(4): 315–324.
- Jain, S., and R. M. Neal, 2004 A split-merge Markov chain Monte Carlo procedure for the Dirichlet process mixture model. *J. Comput. Graph. Stat.* 13: 158–182.
- Jain, S., and R. M. Neal, 2007 Splitting and merging components of a nonconjugate Dirichlet process mixture model. *Bayesian Anal.* 2(3): 445–472.

- Jansen, R. C., 1993 Interval mapping of multiple quantitative trait loci. *Genetics* 135: 205–211.
- Jansen, R. C., and P. Stam, 1994 High resolution of quantitative traits into multiple loci via interval mapping. *Genetics* 136: 1447–1455.
- Kao, C.-H., Z.-B. Zeng, and R. D. Teasdale, 1999 Multiple interval mapping for quantitative trait loci. *Genetics* 152: 1203–1216.
- Kass, R. E., B. P. Carlin, A. Gelman, and R. M. Neal, 1998 Markov chain Monte Carlo in practice: a roundtable discussion. *Am. Stat.* 52(2): 93–100.
- Lander, E. S., and D. Botstein, 1989 Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121: 185–199.
- Saraiva, E. F., and L. A. Milan, 2012 Clustering gene expression data using a posterior split-merge-birth procedure. *Scand. J. Stat.* 39(3): 399–415.
- Satagopan, J. M., and B. S. Yandell, 1996 Estimating the number of quantitative trait loci via Bayesian model determination. *Proceedings of the Joint Statistical Meetings*.
- Satagopan, J. M., B. S. Yandell, M. A. Newton, and T. C. Osborn, 1996 A Bayesian approach to detect quantitative trait loci using Markov chain Monte Carlo. *Genetics* 144: 805–816.
- Stephens, D., and R. Fisch, 1998 Bayesian analysis of quantitative trait locus data using reversible jump Markov chain Monte Carlo. *Biometrics* 54: 1334–1347.
- Stephens, D., and A. Smith, 1993 Bayesian inference in multi-point gene mapping. *Ann. Hum. Genet.* 57(1): 65–82.
- Wergedal, J. E., C. L. Ackert-Bicknell, S.-W. Tsaih, M. H.-C. Sheng, R. Li *et al.*, 2006 Femur mechanical properties in the F2 progeny of an NZB/B1NJ $\times$ RF/J cross are regulated predominantly by genetic loci that regulate bone geometry. *J. Bone Miner. Res.* 21(8): 1256–1266.
- Yang, J., B. Benyamin, B. P. McEvoy, S. Gordon, A. K. Henders *et al.*, 2010 Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* 42(7): 565–569.
- Yi, N., and S. Xu, 2002 Mapping quantitative trait loci with epistatic effects. *Genet. Res.* 79(02): 185–198.
- Yi, N., B. S. Yandell, G. A. Churchill, D. B. Allison, E. J. Eisen *et al.*, 2005 Bayesian model selection for genome-wide epistatic quantitative trait loci analysis. *Genetics* 170: 1333–1344.
- Zeng, Z.-B., 1994 Precision mapping of quantitative trait loci. *Genetics* 136: 1457–1468.

Communicating editor: N. Yi

GENETICS

Supporting Information

www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.180802/-/DC1

Data-Driven Reversible Jump for QTL Mapping

Daiane Aparecida Zuanetti and Luis Aparecido Milan

1 Conditional *a posteriori* distribution of parameters

Combining the likelihood function with the *a priori* distributions, we obtain the conditional *a posteriori* distribution of $\mu | (\mathbf{y}, \mathbf{q}, \boldsymbol{\theta}_{-\mu})$, $\sigma^2 | (\mathbf{y}, \mathbf{q}, \boldsymbol{\theta}_{-\sigma^2})$, $\alpha_k | (\mathbf{y}, \mathbf{q}, \boldsymbol{\theta}_{-\alpha_k})$, $\delta_k | (\mathbf{y}, \mathbf{q}, \boldsymbol{\theta}_{-\delta_k})$ and $\lambda_k | (\mathbf{y}, \mathbf{q}, \boldsymbol{\theta}_{-\lambda_k})$, $k = 1, \dots, K$.

Specifically,

$$\begin{aligned} \mu | (\mathbf{y}, \mathbf{q}, \boldsymbol{\theta}_{-\mu}) &\sim \text{Normal} \left(\frac{\frac{\sum_{i=1}^n (y_i - \sum_{k=1}^K \alpha_k q_{ik} - \sum_{k=1}^K \delta_k (1 - |q_{ik}|))}{\sigma^2} + \frac{\nu_\mu}{\sigma_\mu^2}}{\frac{n}{\sigma^2} + \frac{1}{\sigma_\mu^2}}, \frac{1}{\frac{n}{\sigma^2} + \frac{1}{\sigma_\mu^2}} \right), \\ \sigma^2 | (\mathbf{y}, \mathbf{q}, \boldsymbol{\theta}_{-\sigma^2}) &\sim \text{Inverse-gamma} \left(\frac{n}{2} + \eta_a, \frac{\sum_{i=1}^n (y_i - \sum_{k=1}^K \alpha_k q_{ik} - \sum_{k=1}^K \delta_k (1 - |q_{ik}|))^2}{2} + \eta_b \right), \\ \alpha_{k^*} | (\mathbf{y}, \mathbf{q}, \boldsymbol{\theta}_{-\alpha_{k^*}}) &\sim \text{Normal} \left(\frac{\frac{\sum_{i=1}^n q_{ik^*} (y_i - \mu - \sum_{k \neq k^*} \alpha_k q_{ik} - \sum_{k=1}^K \delta_k (1 - |q_{ik}|))}{\sigma^2} + \frac{\nu_\alpha}{\sigma_\alpha^2}}{\frac{\sum_{i=1}^n q_{ik^*}^2}{\sigma^2} + \frac{1}{\sigma_\alpha^2}}, \frac{1}{\frac{\sum_{i=1}^n q_{ik^*}^2}{\sigma^2} + \frac{1}{\sigma_\alpha^2}} \right), \\ \delta_{k^*} | (\mathbf{y}, \mathbf{q}, \boldsymbol{\theta}_{-\delta_{k^*}}) &\sim \text{Normal} \left(\frac{\frac{\sum_{i=1}^n (1 - |q_{ik^*}|) (y_i - \mu - \sum_{k=1}^K \alpha_k q_{ik} - \sum_{k \neq k^*} \delta_k (1 - |q_{ik}|))}{\sigma^2} + \frac{\nu_\delta}{\sigma_\delta^2}}{\frac{\sum_{i=1}^n (1 - |q_{ik^*}|)^2}{\sigma^2} + \frac{1}{\sigma_\delta^2}}, \frac{1}{\frac{\sum_{i=1}^n (1 - |q_{ik^*}|)^2}{\sigma^2} + \frac{1}{\sigma_\delta^2}} \right) \text{ and} \\ \pi(\lambda_{k^*} | (\mathbf{y}, \mathbf{q}, \boldsymbol{\theta}_{-\lambda_{k^*}})) &\propto \frac{\prod_{i=1}^n \text{Pr}(Q_{ik^*} = q_{ik^*} | M_{ir_{k^*}}, M_{il_{k^*}}, \mathbf{D})}{D_{m-(K-k^*)-D_{r_{k^*}-1}}, \text{ for } k^* = 1, \dots, K \text{ and } k = 1, \dots, K. \end{aligned}$$

2 Additional information about RJ and DDRJ chains of high dimension simulated data sets

Table A shows the effective sample size (ESS) of RJ and DDRJ chains of all parameters of the estimated models for simulated data sets. We observe DDRJ ESS is higher than RJ ESS in most cases which confirms a better mixing of DDRJ chains.

Table A: Effective sample size of RJ and DDRJ chains.

Parameter	$\sigma = 0.5$		$\sigma = 1.0$		$\sigma = 1.5$	
	RJ	DDRJ	RJ	DDRJ	RJ	DDRJ
λ_1	2774	4395	2727	1853	216	417
λ_2	2195	4037	192	4569	226	388
λ_3	2186	3	94	61		1762
λ_4	1621	4752				
λ_5	16	4752				
μ	2774	897	155	1252	1605	2399
α_1	2603	4752	272	4569	230	512
δ_1	2774	4752	3621	607	686	1959
α_2	2561	5004	3621	4569	416	1382
δ_2	2774	4752	27	4569	1528	286
α_3	2483	4752	592	2746		369
δ_3	2774	62	262	548		1168
α_4	301	4396				
δ_4	794	898				
α_5	284	5782				
δ_5	217	333				
σ	1591	5166	3807	3459	2266	3364

FileS3 Functions codes to run DDRJ in R. (.txt, 15 KB)

Available for download as a .txt file at
www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.180802/-/DC1/FileS3.txt

FileS4 An example of running DDRJ in R. (.txt, 17 KB)

Available for download as a .txt file at
www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.180802/-/DC1/FileS4.txt