

Joint Analysis of Strain and Parent-of-Origin Effects for Recombinant Inbred Intercrosses Generated from Multiparent Populations with the Collaborative Cross as an Example

Yanyan Liu,* Sican Xiong,* Wei Sun,[†] and Fei Zou*¹

*School of Mathematics and Statistics, Wuhan University, Hubei 430072, China, [†]Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, and [‡]Department of Biostatistics, University of North Carolina at Chapel Hill, North Carolina 27599

ORCID IDs: 0000-0002-8304-5663 (S.X.); 0000-0002-6350-1107 (W.S.)

ABSTRACT Multiparent populations (MPP) have become popular resources for complex trait mapping because of their wider allelic diversity and larger population size compared with traditional two-way recombinant inbred (RI) strains. In mice, the collaborative cross (CC) is one of the most popular MPP and is derived from eight genetically diverse inbred founder strains. The strategy of generating RI intercrosses (RIX) from MPP in general and from the CC in particular can produce a large number of completely reproducible heterozygote genomes that better represent the (outbred) human population. Since both maternal and paternal haplotypes of each RIX are readily available, RIX is a powerful resource for studying both standing genetic and epigenetic variations of complex traits, in particular, the parent-of-origin (PoO) effects, which are important contributors to many complex traits. Furthermore, most complex traits are affected by >1 genes, where multiple quantitative trait locus mapping could be more advantageous. In this paper, for MPP-RIX data but taking CC-RIX as a working example, we propose a general Bayesian variable selection procedure to simultaneously search for multiple genes with founder allelic effects and PoO effects. The proposed model respects the complex relationship among RIX samples, and the performance of the proposed method is examined by extensive simulations.

KEYWORDS

imprinting
Bayesian variable selection
parameter-expanded
Gaussian priors
Multiparental Populations
MPP

Most human genes have functional mouse counterparts, and genomes of both organisms are organized similarly. Thus, mouse serves as a good model organism for complex human diseases. Recombinant inbred (RI) mice are among the major mouse resources in biomedical and genetic research. However, traditional mouse RI lines are derived from only two inbred parental strains, with limited a number of lines available, resulting in a low percentage (15%) of genetic variation across all mouse inbred

strains (Yuan *et al.* 2011) and extensive blind spots where a sizable proportion of the genome is identical by descent. These limitations make traditional RI insufficient for studying complex traits. With the need for more powerful resources, multiparent populations (MPP) (de Koning and McIntyre 2017), a set of inbred lines using multiple lines as founders, can overcome the limitations of traditional RI lines and have become an innovative tool for fine quantitative trait locus (QTL) mapping. In the past 15 yr, different kinds of MPP have been established in plants and animals, such as nested association mapping design (Yu *et al.* 2008) and multiparent advanced generation intercross (MAGIC) populations (Cavanagh *et al.* 2008; Kover *et al.* 2009; Huang *et al.* 2015; Ladejobi *et al.* 2016) in plants, and the *Drosophila* Synthetic Population Resource (King *et al.* 2012) in animals. The mouse MPP include the collaborative cross (CC) (Complex Trait Consortium 2004), in which a genetically diverse set of eight inbred strains were selected as breeding founders (Iraqi *et al.* 2008) to generate a large number of RI lines. The eight founder strains were predicted to represent ~90% of the genetic variation presented in laboratory mice (Threadgill and Churchill 2012).

Copyright © 2018 Liu *et al.*

doi: <https://doi.org/10.1534/g3.117.300483>

Manuscript received November 7, 2017; accepted for publication December 3, 2017; published Early Online December 18, 2017.

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Supplemental material is available online at www.g3journal.org/lookup/suppl/doi:10.1534/g3.117.300483/-/DC1.

¹Corresponding author: Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599. E-mail: feizou@email.unc.edu

The CC thus greatly overcomes the limitations of the traditional mouse RI lines, and is the only mammalian resource with high genome-wide genetic variation that is uniformly distributed across a large, heterogeneous, and infinitely reproducible population (Threadgill *et al.* 2011).

Through the generation of RI intercrosses (RIX) of RI lines, a large number of potential “outbred” RIX samples can be generated. That is, given L RI lines, $L(L - 1)$ or $L(L - 1)/2$ genetically distinct reciprocal or nonreciprocal F_1 individuals, or RIX, can be produced. An example of using CC-RIX to illustrate a distinct response to Ebola virus infection, by Rasmussen *et al.* (2014), was summarized by the editor: “the CC-RIX mice could prove valuable for preliminary screens of candidate therapeutics and vaccines.” Since all parental RI lines are isogenic at each locus, genotypes of RIX can be imputed in advance from those of their parental RI lines. Additional advantages of RIX can be found in Threadgill and Churchill (2012), in particular, its power in support of analysis of parent-of-origin (PoO) effects, where effects of certain alleles are different depending on whether those alleles are inherited maternally or paternally. PoO makes a significant contribution to the heritability of most complex traits (Mott *et al.* 2014). In addition, genomic PoO effects provide a great model to study epigenetic regulation of gene expression (Barlow 2011).

During the past 20 yr, QTL mapping methods, including analysis of variance (ANOVA), interval mapping (Lander and Botstein 1989), composite interval mapping (Jansen and Stam 1994; Zeng 1994), and multiple interval mapping (Kao and Zeng 1997; Kao *et al.* 1999), have been well developed using experimental crosses, such as backcross, F_2 , and RI (see Broman (2001) for reviews), and many excellent open software packages, such as QTLCart (Basten *et al.* 1999), MapManager (Manly and Olson 1999), and R/qtl (Broman *et al.* 2003), are freely available online. For diallel data obtained from the CC founder strains, Lenarcic *et al.* (2012) employed a general Bayesian model for decomposing phenotypic variance into biologically intuitive components. Crowley *et al.* (2014) also applied a Bayesian method to a diallel cross of the eight founder strains to estimate genome-wide genetic and PoO effects (not QTL mapping) on responses to haloperidol, an antipsychotic drug. Zhang *et al.* (2014) proposed a general single QTL Bayesian framework for MAGIC data to coherently estimate haplotype and diplotype effects of founder alleles. Similarly, for MAGIC data, Wei and Xu (2016) developed a single QTL mapping method with a random-effects model by treating the founder allelic effects of each locus as random, and scanning the entire genome one locus at a time using a likelihood ratio test. For RIX mice, a mixed-effects model was developed for modeling the unbalanced relatedness among them (Tsaih *et al.* 2005; Zou *et al.* 2005). Gong and Zou (2012) proposed a more flexible, nonparametric single QTL mapping method for detecting QTL with time-varying coefficients. Hallin *et al.* (2016) applied a random-effects model to map single QTL by partitioning the variance of growth traits across different environments of yeast strains into additive, dominance, and pairwise epistatic components. However, many complex traits are affected by >1 gene, and multiple QTL mapping may be more powerful than locus-by-locus analysis. For CC-RIX data, Yuan *et al.* (2011) constructed a mixed-effects model to simultaneously map multiple QTL, but treated the eight CC founder alleles as standard biallelic single-nucleotide polymorphisms (SNPs). For many complex traits, it is arguable that modeling the effects of the eight founder alleles could lead to improved mapping power (Collaborative Cross Consortium 2012; Vered *et al.* 2014). In addition, for complex traits where PoO effects are suspected, modeling such effects may further improve QTL mapping power (Lawson *et al.* 2013) and understanding of etiologies of complex traits (Threadgill and Churchill 2012).

In this paper, for MPP-RIX data, we develop a Bayesian variable selection procedure to simultaneously map multiple genes with founder allelic effects and PoO effects. We demonstrate our method with CC-RIX

data, but the method is general enough for other MPP-RIX populations. We place parameter-expanded Gaussian (PeG) priors on both the random founder allelic effects and PoO effects for variable selection.

The paper is organized as follows. In the *Statistical Method* section, we first introduce the CC-RIX experiment, then propose a random-effects model and describe a Bayesian variable selection procedure. In the *Simulation Study* section, we perform extensive simulation to examine the proposed method. Summary comments are given in the *Discussion* section.

STATISTICAL METHOD

The CC-RIX panel is the RI intercross of CC lines. For L CC RI lines, a total of $L(L - 1)$ reciprocal CC-RIX and $L(L - 1)/2$ nonreciprocal CC-RIX can be generated [see Figure 1 of Zou *et al.* (2005), Yuan *et al.* (2011)]. Let n be the total number of CC-RIX samples and p be the total number of genetic markers. Further, let the phenotype of individual i (the dependent variable) be y_i ($i = 1, \dots, n$).

Model

In order to account for the unbalanced relatedness between the CC-RIX mice and model the eight founder alleles and PoO effects, we extend the mixed additive random-effects model of Gong and Zou (2012) and Yuan *et al.* (2011) as follows:

$$y_i = \mu + \sum_{j=1}^p \gamma_{Q_j} \mathbf{x}_{ij}^T \boldsymbol{\beta}_j + \sum_{j=1}^p \gamma_{P_j} \mathbf{z}_{ij}^T \boldsymbol{\xi}_j + \sum_{l=1}^L a_{il} \alpha_l + e_i \quad (1)$$

where γ_{Q_j} and γ_{P_j} are binary {0, 1} variables used to decide whether the j th QTL and PoO should be included in or excluded from the model and μ is the overall mean; $\boldsymbol{\beta}_j = (\beta_{j1}, \dots, \beta_{j8})^T$ and the k th element of \mathbf{x}_{ij}^T equals 2, 1, or 0, depending on whether CC-RIX $_i$ inherits 2, 1, or 0 copies of the k th founder allele ($k = 1, \dots, 8$) at the j th candidate locus; $a_{il} = \#$ of parents of CC-RIX $_i$ that are equal to CC-RI $_l$. Let \mathbf{A} be an $n \times L$ matrix whose (i, l) th element equals a_{il} . Clearly, $\sum_{i=1}^n a_{il} \equiv 2$ for all $l = 1, 2, \dots, n$, since each CC-RIX has two and only two parents. Therefore, β_{jk} ($j = 1, \dots, p$) represents the k th founder allelic effect of locus j , and α_l ($l = 1, \dots, L$) represents the random polygenic effect of founder strain l . We let α_l follow $N(0, \sigma_a^2)$ ($l = 1, 2, \dots, L$), and the random error e_i follow $N(0, \sigma_e^2)$ ($i = 1, 2, \dots, n$).

At a given locus, PoO effects can only be estimated from individuals with heterogeneous genotypes, since for those with homozygote genotypes, the parental contributions cannot be distinguished. Label the eight-founder alleles A, B, C, D, E, F, G, and H by numbers 1–8, and let the k th element of $\mathbf{z}_{ij} = (z_{ij1}, \dots, z_{ij8})^T$ be equal to 1 or -1 if the maternal or paternal allele of the i th subject at the j th locus equals the k th founder allele, and 0 otherwise. For those CC-RIX samples with a homozygous genotype at a given locus, such as AA or BB, we let $\mathbf{z}_{ij} = \mathbf{0}_{1 \times 8}$. Therefore, $\boldsymbol{\xi}_j = (\xi_{j1}, \dots, \xi_{j8})^T$ represents the PoO effects of the j th locus.

Model (1) can be expressed by the following matrix form:

$$\mathbf{y} = \boldsymbol{\mu} + \sum_{j=1}^p \gamma_{Q_j} \mathbf{x}_j \boldsymbol{\beta}_j + \sum_{j=1}^p \gamma_{P_j} \mathbf{z}_j \boldsymbol{\xi}_j + \mathbf{A} \boldsymbol{\alpha} + \mathbf{e} \quad (2)$$

where $\mathbf{y} = (y_1, \dots, y_n)^T$, $\boldsymbol{\mu} = \mu \mathbf{1}_n$, $\mathbf{x}_j = (\mathbf{x}_{1j}, \dots, \mathbf{x}_{nj})^T$, $\mathbf{z}_j = (\mathbf{z}_{1j}, \dots, \mathbf{z}_{nj})^T$, $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_L)^T$, $\mathbf{e} = (e_1, \dots, e_n)^T$, and $\mathbf{1}_n = (1, \dots, 1)^T$.

To enable selection of QTL and PoO effects, we reexpress the eight-dimensional vectors $\boldsymbol{\beta}_j$ and $\boldsymbol{\xi}_j$ as the expanded parameters $\tilde{\boldsymbol{\beta}}_j$ and $\tilde{\boldsymbol{\xi}}_j$ for $j = 1, \dots, p$, respectively, such that $\tilde{\boldsymbol{\beta}}_j = \gamma_{Q_j} \boldsymbol{\beta}_j$ and $\tilde{\boldsymbol{\xi}}_j = \gamma_{P_j} \boldsymbol{\xi}_j$. This

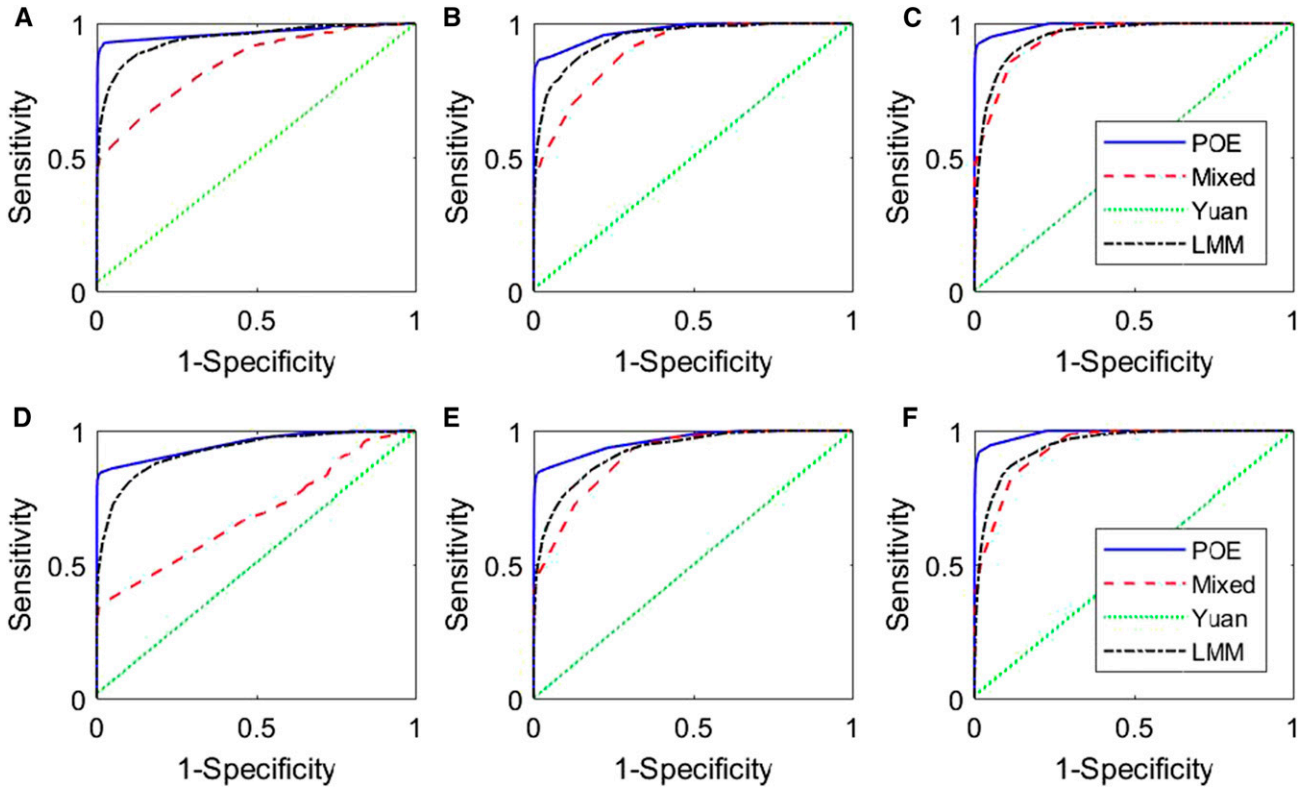


Figure 1 ROC curves. (A–C) for cases 1, 2, and 3, respectively, and (D–F) for cases 1*, 2*, and 3*, respectively. The legends “POE”, “Mixed”, “Yuan” and “LMM” represent our proposed model (2), the mixed model (3), Yuan’s model (5), and the LMM model (4), respectively.

parameter-expansion idea was first proposed by Kuo and Mallick (1998) and has an advantage in selecting important variables and shrinking coefficients.

In the above model (2), the observed data are \mathbf{y} , marker genotypes $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_p)$, PoO information matrix $\mathbf{z} = (\mathbf{z}_1, \dots, \mathbf{z}_p)$, and parental CC-RI information matrix \mathbf{A} . The unobserved variables include μ , $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \dots, \boldsymbol{\beta}_p^T)^T$, $\boldsymbol{\xi} = (\boldsymbol{\xi}_1^T, \dots, \boldsymbol{\xi}_p^T)^T$, $\boldsymbol{\alpha}$, σ_e^2 , σ_a^2 , and the indicators $\boldsymbol{\gamma}_Q = \{\gamma_{Q1}, \dots, \gamma_{Qp}\}$ and $\boldsymbol{\gamma}_P = \{\gamma_{P1}, \dots, \gamma_{Pp}\}$. We assign the following conjugate Gaussian priors to the j^{th} QTL and j^{th} PoO effects as:

$$\boldsymbol{\beta}_j \sim N_8(\mathbf{0}, \sigma_{Qj}^2 \mathbf{I}_8) \quad \text{and} \quad \boldsymbol{\xi}_j \sim N_8(\mathbf{0}, \sigma_{Pj}^2 \mathbf{I}_8) \quad \text{for } j = 1, \dots, p.$$

For convenience, we name the above hierarchical priors for QTL and PoO coefficients as PeG priors in the sequel. We assign conjugate noninformation hyper-priors to the variance parameters

$$P(\sigma_{Qj}^2) \propto \frac{1}{\sigma_{Qj}^2}, \quad P(\sigma_{Pj}^2) \propto \frac{1}{\sigma_{Pj}^2}, \quad P(\sigma_a^2) \propto (\sigma_a^2)^{\delta-1} \quad \text{and} \\ P(\sigma_e^2) \propto \frac{1}{\sigma_e^2}$$

where $\delta(0 < \delta \leq 1/2)$ is used to ensure that the posterior distribution is proper (ter Braak *et al.* 2005). The above PeG priors are equivalent to “spike and slab” point mass mixture Gaussian priors (Mitchell and Beauchamp 1988; George and McCulloch 1993) for the expanded parameters

$$\tilde{\boldsymbol{\beta}}_j \sim \gamma_{Qj} \delta_0 + (1 - \gamma_{Qj}) N_8(\mathbf{0}, \sigma_{Qj}^2 \mathbf{I}_8) \quad \text{for } j = 1, \dots, p.$$

and

$$\tilde{\boldsymbol{\xi}}_j \sim \gamma_{Pj} \delta_0 + (1 - \gamma_{Pj}) N_8(\mathbf{0}, \sigma_{Pj}^2 \mathbf{I}_8) \quad \text{for } j = 1, \dots, p.$$

It is well known that these priors can achieve variable selection. However, compared with mixture priors, the PeG priors can employ a block Gibbs sampler to update the two blocks of parameters, one for $\boldsymbol{\beta}^s$ and $\boldsymbol{\xi}^s$ (corresponding to the selected predictors) and one for $\boldsymbol{\beta}^u$ and $\boldsymbol{\xi}^u$ (corresponding to the unselected predictors) in turn, and therefore can dramatically reduce computational time, especially for high-dimensional data with large p .

For indicator variables γ_{Qj} and γ_{Pj} , we specify Bernoulli priors with inclusion probability $0 < \eta_{Qj} < 1$ and $0 < \eta_{Pj} < 1$ for $j = 1, \dots, p$, respectively. To be more flexible, we further apply hierarchical uniform priors to η_{Qj} and η_{Pj} :

$$\eta_{Qj} \sim U[0, 1], \quad \text{and} \quad \eta_{Pj} \sim U[0, 1], \quad \text{for } j = 1, \dots, p.$$

Lastly, we specify a flat prior on μ as $P(\mu) \propto 1$.

Block Gibbs sampling algorithm for posterior computation

The specific priors above result in known marginal conditional distributions for all variables. The blockwise Gibbs sampling algorithm that we employ can be summarized as follows:

First, we initiate σ_a^2 , σ_e^2 , $\boldsymbol{\sigma}_Q^2 = \{\sigma_{Q1}^2, \dots, \sigma_{Qp}^2\}$, $\boldsymbol{\sigma}_P^2 = \{\sigma_{P1}^2, \dots, \sigma_{Pp}^2\}$, $\boldsymbol{\eta}_Q = \{\eta_{Q1}, \dots, \eta_{Qp}\}$, and $\boldsymbol{\eta}_P = \{\eta_{P1}, \dots, \eta_{Pp}\}$ from uniform distribution $U(0, 1)$, then sample other parameters $\boldsymbol{\beta}$, $\boldsymbol{\xi}$, $\boldsymbol{\alpha}$, and indicators $\boldsymbol{\gamma}_Q$ and $\boldsymbol{\gamma}_P$ from their priors. We then perform the following block Gibbs sampling procedures. Superscripts (t) and ($t + 1$) signify the Markov chain Monte Carlo (MCMC) iterations, and $t = 0$ refers to the initial iteration.

Step 1. Updating μ : $\mu^{(t+1)}$ is drawn from the normal distribution.

$$N\left(\frac{1}{n}\mathbf{1}'_n\left(\mathbf{y}-\sum_{j=1}^p\gamma_{Q_j}^{(t)}\mathbf{x}_j\boldsymbol{\beta}_j^{(t)}-\sum_{j=1}^p\gamma_{P_j}^{(t)}\mathbf{z}_j\xi_j^{(t)}-\mathbf{A}\boldsymbol{\alpha}^{(t)}\right),\frac{\sigma_e^{2(t)}}{n}\right)$$

Step 2. Updating $\boldsymbol{\beta}$: we divide the long vector $\boldsymbol{\beta}^{(t)}$ into two blocks $\boldsymbol{\beta}^{s(t)} = ((\boldsymbol{\beta}_{s_1}^{(t)})', \dots, (\boldsymbol{\beta}_{s_{m_1}}^{(t)})')'$ and $\boldsymbol{\beta}^{u(t)} = ((\boldsymbol{\beta}_{u_1}^{(t)})', \dots, (\boldsymbol{\beta}_{u_{m_0}}^{(t)})')'$, corresponding to the selected ($\gamma_{Q_j}^{(t)} = 1$) and unselected ($\gamma_{Q_j}^{(t)} = 0$) predictors, respectively. We then sample $\boldsymbol{\beta}_{u_h}^{(t+1)} (h = 1, \dots, m_0)$ from their priors, i.e., the eight-dimensional multivariate normal distribution with zero mean and covariance matrix $\sigma_{Q,u_h}^{2(t)}\mathbf{I}_8$, and sample $\boldsymbol{\beta}_{s_h}^{(t+1)} (h = 1, \dots, m_1)$ from the eight-dimensional multivariate normal distribution with mean

$$1/\sigma_e^{2(t)}\boldsymbol{\Sigma}_{\boldsymbol{\beta}_{s_h}}^{(t)}(\mathbf{x}_{s_h})'(\mathbf{y}-\mu^{(t+1)}\mathbf{1}_n-\sum_{j<h}^p\gamma_{Q,s_j}^{(t)}\mathbf{x}_{s_j}\boldsymbol{\beta}_{s_j}^{(t+1)})$$

$$-\sum_{j>h}^p\gamma_{Q,s_j}^{(t)}\mathbf{x}_{s_j}\boldsymbol{\beta}_{s_j}^{(t)}-\sum_{j=1}^p\gamma_{P_j}^{(t)}\mathbf{z}_j\xi_j^{(t)}-\mathbf{A}\boldsymbol{\alpha}^{(t)})$$

and covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\beta}_{s_h}}^{(t)} = \sigma_{Q,s_h}^{2(t)}(\sigma_{Q,s_h}^{2(t)}/\sigma_e^{2(t)}(\mathbf{x}_{s_h})'\mathbf{x}_{s_h} + \mathbf{I}_8)^{-1}$.

Step 3. Updating $\boldsymbol{\xi}$: similar to Step 2, we divide $\boldsymbol{\xi}^{(t)}$ into two parts, $\boldsymbol{\xi}^{s(t)} = ((\xi_{s_1}^{(t)})', \dots, (\xi_{s_{m_1}}^{(t)})')'$ and $\boldsymbol{\xi}^{u(t)} = ((\xi_{u_1}^{(t)})', \dots, (\xi_{u_{m_0}}^{(t)})')'$, corresponding to the selected ($\gamma_{P_j}^{(t)} = 1$) and unselected ($\gamma_{P_j}^{(t)} = 0$) predictors, respectively. Then $\xi_{u_h}^{(t+1)} (h = 1, \dots, m_0)$ is sampled from their priors, i.e., the eight-dimensional multivariate normal distribution with zero mean and covariance matrix $\sigma_{P,u_h}^{2(t)}\mathbf{I}_8$, and $\xi_{s_h}^{(t+1)} (h = 1, \dots, m_1)$ is sampled from the eight-dimensional multivariate normal distribution with mean

$$1/\sigma_e^{2(t)}\boldsymbol{\Sigma}_{\boldsymbol{\xi}_{s_h}}^{(t)}(\mathbf{z}_{s_h})'(\mathbf{y}-\mu^{(t+1)}\mathbf{1}_n-\sum_{j=1}^p\gamma_{Q_j}^{(t)}\mathbf{x}_j\boldsymbol{\beta}_j^{(t+1)})$$

$$-\sum_{j<h}^p\gamma_{P,s_j}^{(t)}\mathbf{z}_{s_j}\xi_{s_j}^{(t+1)}-\sum_{j>h}^p\gamma_{P,s_j}^{(t)}\mathbf{z}_{s_j}\xi_{s_j}^{(t)}-\mathbf{A}\boldsymbol{\alpha}^{(t)})$$

and covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\xi}_{s_h}}^{(t)} = \sigma_{P,s_h}^{2(t)}(\sigma_{P,s_h}^{2(t)}/\sigma_e^{2(t)}(\mathbf{z}_{s_h})'\mathbf{z}_{s_h} + \mathbf{I}_8)^{-1}$.

Step 4. Updating $\boldsymbol{\alpha}$: $\boldsymbol{\alpha}^{(t+1)}$ is drawn from the multivariate normal distribution with mean

$$1/\sigma_e^{2(t)}\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}^{(t)}\mathbf{A}'(\mathbf{y}-\mu^{(t+1)}\mathbf{1}_n-\sum_{j=1}^p\gamma_{Q_j}^{(t)}\mathbf{x}_j\boldsymbol{\beta}_j^{(t+1)}-\sum_{j=1}^p\gamma_{P_j}^{(t)}\mathbf{z}_j\xi_j^{(t+1)})$$

and covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}^{(t)} = \sigma_a^{2(t)}(\sigma_a^{2(t)}/\sigma_e^{2(t)}\mathbf{A}'\mathbf{A} + \mathbf{I}_L)^{-1}$.

Step 5. Updating $\sigma_{Q_j}^2 (1 \leq j \leq p)$: $\sigma_{Q_j}^{2(t+1)}$ is sampled from the scale-inverted χ^2 distribution, $\|\boldsymbol{\beta}_j^{(t+1)}\|^2/\chi_8^2$.

Step 6. Updating $\sigma_{P_j}^2 (1 \leq j \leq p)$: $\sigma_{P_j}^{2(t+1)}$ is sampled from the scale-inverted χ^2 distribution, $\|\xi_j^{(t+1)}\|^2/\chi_8^2$.

Step 7. Updating σ_a^2 : the random-effects variance $\sigma_a^{2(t+1)}$ is sampled from the scale-inverted χ^2 distribution, $\|\boldsymbol{\alpha}^{(t+1)}\|^2/\chi_{L-2\delta}^2$.

Step 8. Updating σ_e^2 : the residual variance $\sigma_e^{2(t+1)}$ is sampled from the scale-inverted χ^2 distribution,

$$\left\|\mathbf{y}-\mu^{(t+1)}\mathbf{1}_n-\sum_{j=1}^p\gamma_{Q_j}^{(t)}\mathbf{x}_j\boldsymbol{\beta}_j^{(t+1)}-\sum_{j=1}^p\gamma_{P_j}^{(t)}\mathbf{z}_j\xi_j^{(t+1)}-\mathbf{A}\boldsymbol{\alpha}^{(t+1)}\right\|^2/\chi_n^2$$

Step 9. Updating $\gamma_{Q_j} (1 \leq j \leq p)$: $\gamma_{Q_j}^{(t+1)}$ is sampled from the Bernoulli distribution with success probability

$p_j^{Q(t)} = \eta_{Q_j}^{(t)}/(\eta_{Q_j}^{(t)} + (1 - \eta_{Q_j}^{(t)})p_{j0}^{Q(t)})$, where

$$p_{j0}^{Q(t)} = \exp\left[-1/2\sigma_e^{2(t+1)}\left(2\left(\mathbf{K}_j^{Q(t)}\right)'\mathbf{x}_j\boldsymbol{\beta}_j^{(t+1)}-\left(\boldsymbol{\beta}_j^{(t+1)}\right)'\left(\mathbf{x}_j'\mathbf{x}_j\right)\boldsymbol{\beta}_j^{(t+1)}\right)\right],$$

and

$$\mathbf{K}_j^{Q(t)} = \mathbf{y} - \mu^{(t+1)}\mathbf{1}_n - \sum_{k<j}\gamma_{Q_k}^{(t+1)}\mathbf{x}_k\boldsymbol{\beta}_k^{(t+1)} - \sum_{k>j}\gamma_{Q_k}^{(t)}\mathbf{x}_k\boldsymbol{\beta}_k^{(t+1)} - \sum_{j=1}^p\gamma_{P_j}^{(t)}\mathbf{z}_j\xi_j^{(t+1)} - \mathbf{A}\boldsymbol{\alpha}^{(t+1)}.$$

Step 10. Updating $\gamma_{P_j} (1 \leq j \leq p)$: $\gamma_{P_j}^{(t+1)}$ is sampled from the Bernoulli distribution with success probability $p_j^{P(t)} = \eta_{P_j}^{(t)}/(\eta_{P_j}^{(t)} + (1 - \eta_{P_j}^{(t)})p_{j0}^{P(t)})$, where

$$p_{j0}^{P(t)} = \exp\left[-1/2\sigma_e^{2(t+1)}\left(2\left(\mathbf{K}_j^{P(t)}\right)'\mathbf{z}_j\xi_j^{(t+1)}-\left(\xi_j^{(t+1)}\right)'\left(\mathbf{z}_j'\mathbf{z}_j\right)\xi_j^{(t+1)}\right)\right],$$

and

$$\mathbf{K}_j^{P(t)} = \mathbf{y} - \mu^{(t+1)}\mathbf{1}_n - \sum_{j=1}^p\gamma_{Q_j}^{(t+1)}\mathbf{x}_j\boldsymbol{\beta}_j^{(t+1)} - \sum_{k<j}\gamma_{P_k}^{(t+1)}\mathbf{z}_k\xi_k^{(t+1)} - \sum_{k>j}\gamma_{P_k}^{(t)}\mathbf{z}_k\xi_k^{(t+1)} - \mathbf{A}\boldsymbol{\alpha}^{(t+1)}.$$

Step 11. Updating $\eta_{Q_j} (1 \leq j \leq p)$: $\eta_{Q_j}^{(t+1)}$ is sampled from the beta distribution $Beta(1 + \gamma_{Q_j}^{(t+1)}, 2 - \gamma_{Q_j}^{(t+1)})$.

Step 12. Updating $\eta_{P_j} (1 \leq j \leq p)$: $\eta_{P_j}^{(t+1)}$ is sampled from the beta distribution $Beta(1 + \gamma_{P_j}^{(t+1)}, 2 - \gamma_{P_j}^{(t+1)})$.

Data availability

Supplemental Material, File S1 contains five supplemental figures. The MATLAB code used to analyze the simulated data is provided in File S2.

SIMULATION STUDY

In this section, we run extensive simulations to evaluate the performance of the proposed Bayesian method. We apply the loop design in Zou *et al.* (2005) and Yuan *et al.* (2011) by ordering the L CC-RI lines randomly and forming them into a circle, and then mating each CC-RI line (clockwise) with the next three CC-RI lines after it (i.e., CC-RI₁ mated with CC-RI₂, CC-RI₃, and CC-RI₄; CC-RI₂ mated with CC-RI₃, CC-RI₄, and CC-RI₁; and so on). For $L = 100$, in this way, we can generate a total of $n = 300$ CC-RIX samples. Parental CC-RI genotypes are simulated in R/qrtl (Broman *et al.* 2003). All the simulation results are based on total 100 replications for each simulation setup.

In model (2), we set the overall mean $\mu = 1$, the variance of random errors $\sigma_e^2 = 1$, the polygenic effect variance $\sigma_a^2 = 1$, and $\delta = 10^{-3}$. Nineteen chromosomes each with a length of 70 cM are simulated, on which total $p = 133, 266$, and 1330 evenly spaced markers (resulting in 10-, 5-, and 1-cM intervals between nearby markers on each chromosome, named cases 1–3) are generated, corresponding to three marker density cases. Among the total p simulated markers, we randomly select five markers, and let the first three markers have QTL effects the last three markers have PoO effects. That is, the first two

genes only have QTL effects, and the last two only have PoO effects. However, the middle-selected marker has both QTL and PoO effects. The corresponding variances of QTL or PoO effects of the selected markers are all set to 1.

To better assess the performance of our method in situations where multiple nearby SNPs jointly affect the outcome, for cases 1–3, we now let the number of causal SNPs in each of the first two QTL be 2 instead of 1. Specifically, we randomly select two nearby SNPs for each of the first two simulated QTL and denote the alleles of the two SNPs A and a . Based on the haplotype frequencies f_{AA}, f_{Aa}, f_{aA} , and f_{aa} (without loss of generality, we assume that $f_{AA} \geq f_{Aa} \geq f_{aA} \geq f_{aa}$), we create three haplotype allelic groups, where haplotype AA is group 1, Aa is group 2, and the other two are group 3, and set the genetic effects of the three groups to 1, 2, and 3, respectively. The three new simulation setups are labeled as cases 1* to 3*.

For each simulated data point, we generate a single long chain with 20,000 cycles, of which the first 10,000 cycles are discarded as burn-in, resulting in a total of 10,000 samples for post-MCMC analysis. All the analysis is done in MATLAB, and the MATLAB source code is submitted as supplemental material (File S2).

For comparison, we fit each simulated data point with the following three models.

The same model as (2), but without the PoO terms

$$y = \mu + \sum_{j=1}^p \gamma_{Qj} x_j \beta_j + A\alpha + e \quad (3)$$

All the priors are set to be the same as their counterparts in model (2). For convenience, we subsequently call model (3) the “mixed model”.

Linear mixed-effects model (LMM) for single gene mapping

$$y = x_j \beta_j + z_j \xi_j + A\alpha + e, 1 \leq j \leq p. \quad (4)$$

Here, random effects $\alpha \sim N(\mathbf{0}, \sigma_a^2 I_L)$ and random errors $e \sim N(\mathbf{0}, \sigma_e^2 I_n)$, and the random effects α are also assumed to be independent of the random errors e as before. Moreover, $\beta_j = (\beta_{j1}, \dots, \beta_{j8})'$ and $\xi_j = (\xi_{j1}, \dots, \xi_{j8})'$ are the fixed QTL and PoO effects, respectively, of the j th tested locus. Given the constraints that $\sum_{k=1}^8 x_{ijk} = 2$ and $\sum_{k=1}^8 z_{ijk} = 0$ for $i = 1, \dots, n, j = 1, \dots, p$, we force the intercept μ in model (4) to be 0 to overcome the identifiability problem, and jointly test the effects of the j th locus as:

$$H_{0j} : \beta_{j1} = \beta_{j2} = \dots = \beta_{j8}, \& \xi_{j1} = \xi_{j2} = \dots = \xi_{j8}.$$

In contrast to H_{1j} , some of the equations in H_{0j} are not satisfied. After obtaining the maximum likelihood estimates of the parameters in model (4), we perform a test with the following log-odds ratio (LOD):

$$LOD_j = \log_{10} \frac{L_{1j}}{L_{0j}}$$

which is equivalent to the log-likelihood ratio test, where L_{0j} and L_{1j} are the likelihood of model (4) under null hypothesis H_{0j} and alternative hypothesis H_{1j} , respectively. Since the above hypothesis test is performed p times, it is necessary to find an appropriate significance threshold to control the multiple testing, which can be obtained, for example, by modified permutation procedures (Zou *et al.* 2005).

Table 1 Simulation settings and AUC values

Case	IS ^a	p	AUC _P ^b	AUC _M ^c	AUC _Y ^d	AUC _L ^e
1	10	133	0.9642	0.8499	0.5155	0.9436
2	5	266	0.9722	0.9022	0.5029	0.9463
3	1	1330	0.9915	0.9489	0.5000	0.9525
1*	10	133	0.9499	0.6885	0.5085	0.9264
2*	5	266	0.9624	0.9056	0.4998	0.9238
3*	1	1330	0.9908	0.9393	0.5040	0.9482

^aInterval space (IS; cM) between nearby markers.

^bAUC values of our proposed model (2).

^cAUC values of the mixed model (3).

^dAUC values of Yuan’s model (5).

^eAUC values of the LMM model (4).

However, in our comparison, we evaluate the receiver operating characteristic (ROC) curve, which only requires the use of LOD scores.

Yuan’s Bayesian method

$$y_i = \mu + \sum_{j=1}^p z_{ij} a_j + \sum_{l=1}^L a_{il} \alpha_l + e_i \quad (5)$$

Here, a_j is the effect of the j th putative QTL with $z_{ij} = \sqrt{2} m_{ij}$, where m_{ij} ($i = 1, \dots, n, j = 1, \dots, p$) equals $-1, 0$, or 1 if the putative QTL genotype is aa, Aa , or AA , respectively. The other parameters are set the same as those in the mixed model. The prior of a_j is set to $N(0, \sigma_j^2)$, $1 \leq j \leq p$ in Yuan *et al.* (2011).

To compare the methods, ROC curves (Fawcett 2006), where true positive rates (also known as sensitivity) are plotted against false positive rates (also known as 1-specificity) evaluated at various threshold cut-offs. To estimate the sensitivity (the proportion of positives that are correctly identified as such) and specificity (the proportion of negatives that are correctly identified as such), we define false and true positive findings as follows. If a detected locus falls no more than 10 cM apart from any simulated genes, we call it a true positive finding, otherwise a false positive finding. Then, combining the outputs of each model for the 100 data sets, we can calculate the corresponding sensitivity and specificity. Specifically, for the j th marker, we record the LOD scores, LOD_j , for the LMM model; the maximum posterior frequency between γ_{Qj} and γ_{Pj} for our proposed method; the posterior frequency γ_{Qj} for the mixed model; and the posterior mean of σ_j^2 for Yuan’s model. The corresponding area under the ROC curve (AUC) is also calculated for each of the four models, and the results are presented in Table 1. Generally speaking, a model with a higher AUC value indicates on average a better performance compared with those with lower AUC scores (Fawcett 2006).

From Table 1, it is clear that our proposed method outperforms the other three methods for all cases, regardless of whether there is a single causal SNP or multiple causal SNPs in each QTL. Yuan’s method fails in all the simulated cases, which is expected as the method only models biallelic SNP effects instead of the founder allelic effects that we simulate. The LMM model outperforms the mixed model, in particular when the marker density is sparse. This phenomenon is further confirmed by Figure 1, where the ROC curves of the proposed method are always higher than the ROC curves of the other three methods; the ROC curves of Yuan’s model fall along the 45-degree line, indicating its low power for mapping genes with founder allelic effects. The ROC curves of the LMM model fall between those of our proposed model and the mixed model in cases 1 and 2, and cases 1* and 2*, but cross

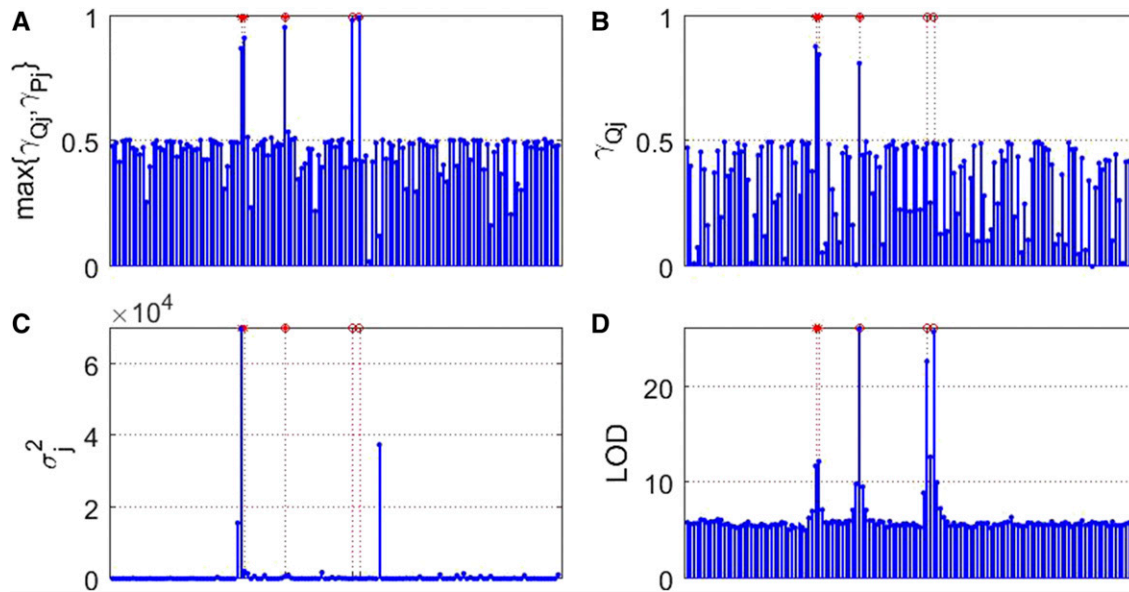


Figure 2 Estimate plot for case 1. X-axis: * indicates QTL location, and o indicates PoO location. Y-axis: (A) $\max\{\gamma_{Q_j}, \gamma_{P_j}\}$ ($1 \leq j \leq p$) in our proposed model (2); (B) γ_{Q_j} ($1 \leq j \leq p$) in the mixed model (3); (C) σ_j^2 ($1 \leq j \leq p$) in Yuan's model (5); and (D) LOD scores for linear mixed-effects model (LMM) (4).

with the ROC curves of the mixed model in cases 3 and 3*. Figure 1 and Table 1 show that the mapping precisions of all the methods increase as the number of markers increases, as expected.

Figure 2 presents the Manhattan plots of the four models based on the average estimate of a total of 100 replications across the whole genome for case 1. Clearly, our proposed method is the winner in mapping genes with founder allelic effects (marked by an asterisk) and PoO effects (marked by a circle). The Manhattan plots for cases 2 and 3, and cases 1*, 2*, and 3*, are included in Figures S1–S5 in File S1, respectively, and show similar patterns.

DISCUSSION

The CC (Threadgill and Churchill 2012), a panel of newly emerged multiparent RI mouse strains, was developed, similar to other MPPs, to provide greater genetic diversity than the traditional RI populations and thereby to improve our power of understanding of complex traits. CC-RIX, F1 crosses that are generated from parental CC RI lines (Zou *et al.* 2005) can serve as excellent mouse models for mapping genes with both traditional genetic effects and epigenetic effects, such as imprinting. This study extends the model of Gong and Zou (2012) and Yuan *et al.* (2011) and, to the best of our knowledge, it is the first use of Bayesian variable selection methods to jointly map multiple QTL with both founder allelic effects and PoO effects for MPP-RIX data, in particular, CC-RIX data.

In this article, it is assumed that the RI lines are equally distanced from each other in terms of genetic distance which is a sensible assumption given the funnel design used for generating CC RI lines. However, we do observe that some RI lines share more or fewer founder alleles than expected. Our limited simulation shows that such genetic unbalance has a negligible effect on mapping genes, but this issue deserves further investigation. Alternatively, as genotypes of the parental RI lines are available, we could modify the design matrices \mathbf{x}_j , \mathbf{z}_j , and \mathbf{A} in model (1) and replace them with local and genome-wide similarity matrices as done in sequence kernel association tests (SKAT) (Ionita-Laza *et al.* 2013) and genome-wide complex trait analysis (GCTA)

(Yang *et al.* 2011). In addition, in our model we assume additive founder allelic effects. This assumption can be easily relaxed to allow genes with both additive and dominance effects (Woods *et al.* 2012; Zhang *et al.* 2014). Because dominant genetic effects are orthogonal to PoO effects, missing the dominant genetic effects does not cause any bias in the PoO effects estimate.

One of the drawbacks of Bayesian methods is their computational cost, especially for data with large numbers of samples and markers. Our model employs a block Gibbs sampler, which dramatically reduces computational time. However, further improvements may be possible. For example, instead of modeling each marker separately, we could jointly model multiple nearby markers and reduce the magnitude of p . Similar ideas have been proposed for genome-wide association study data (Wu *et al.* 2010; Lu *et al.* 2015). This approach is also biologically meaningful, since for some complex traits multiple causal SNPs may be located in a single region, and our simulations (cases 1*–3*) have shown that the proposed method is powerful in mapping regions with multiple causal SNPs. However, grouping nearby markers or SNPs may offer further help in improving mapping power, which deserves more thorough investigation.

ACKNOWLEDGMENTS

We thank two anonymous referees for their valuable comments on our manuscript. This research was partially supported by NIH grant R01 GM105785, and by the scholarship from China Scholarship Council (CSC) under the Grant CSC No.201506270042.

LITERATURE CITED

- Barlow, D. P., 2011 Genomic imprinting: a mammalian epigenetic discovery model. *Annu. Rev. Genet.* 45: 379–403.
- Basten, C. J., B. S. Weir, and Z.-B. Zeng, 1999 *QTL Cartographer: A Reference Manual and Tutorial for QTL Mapping*. Department of Statistics, North Carolina State University, Raleigh, NC.
- Broman, K. W., 2001 Review of statistical methods for QTL mapping in experimental crosses. *Lab Anim.* (NY) 30: 44–52.

- Broman, K. W., H. Wu, S. Sen, and G. A. Churchill, 2003 R/qtl: QTL mapping in experimental crosses. *Bioinformatics* 19: 889–890.
- Cavanagh, C., M. Morell, I. Mackay, and W. Powell, 2008 From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants. *Curr. Opin. Plant Biol.* 11: 215–221.
- Collaborative Cross Consortium, 2012 The genome architecture of the collaborative cross mouse genetic reference population. *Genetics* 190: 389–401.
- Complex Trait Consortium, 2004 The collaborative cross, a community resource for the genetic analysis of complex traits. *Nat. Genet.* 36: 1133–1137.
- Crowley, J. J., Y. Kim, A. B. Lenarcic, C. R. Quackenbush, C. J. Barrick *et al.*, 2014 Genetics of adverse reactions to haloperidol in a mouse diallel: a drug–placebo experiment and Bayesian causal analysis. *Genetics* 196: 321–347.
- de Koning, D.-J., and L. M. McIntyre, 2017 Back to the future: multiparent populations provide the key to unlocking the genetic basis of complex traits. *Genetics* 206: 527–529.
- Fawcett, T., 2006 An introduction to ROC analysis. *Pattern Recognit. Lett.* 27: 861–874.
- George, E. I., and R. E. McCulloch, 1993 Variable selection via Gibbs sampling. *J. Am. Stat. Assoc.* 88: 881–889.
- Gong, Y., and F. Zou, 2012 Varying coefficient models for mapping quantitative trait loci using recombinant inbred intercrosses. *Genetics* 190: 475–486.
- Hallin, J., K. Märtens, A. I. Young, M. Zackrisson, F. Salinas *et al.*, 2016 Powerful decomposition of complex traits in a diploid model. *Nat. Commun.* 7: 13311.
- Huang, B. E., K. L. Verbyla, A. P. Verbyla, C. Raghavan, V. K. Singh *et al.*, 2015 MAGIC populations in crops: current status and future prospects. *Theor. Appl. Genet.* 128: 999–1017.
- Ionita-Laza, I., S. Lee, V. Makarov, J. D. Buxbaum, and X. Lin, 2013 Sequence kernel association tests for the combined effect of rare and common variants. *Am. J. Hum. Genet.* 92: 841–853.
- Iraqi, F. A., G. Churchill, and R. Mott, 2008 The collaborative cross, developing a resource for mammalian systems genetics: a status report of the Wellcome trust cohort. *Mamm. Genome* 19: 379–381.
- 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., and Z.-B. Zeng, 1997 General formulas for obtaining the MLEs and the asymptotic variance-covariance matrix in mapping quantitative trait loci when using the EM algorithm. *Biometrics* 53: 653–665.
- Kao, C.-H., Z.-B. Zeng, and R. D. Teasdale, 1999 Multiple interval mapping for quantitative trait loci. *Genetics* 152: 1203–1216.
- King, E. G., C. M. Merkes, C. L. McNeil, S. R. Hooper, S. Sen *et al.*, 2012 Genetic dissection of a model complex trait using the *Drosophila* synthetic population resource. *Genome Res.* 22: 1558–1566.
- Kover, P. X., W. Valdar, J. Trakalo, N. Scarcelli, I. M. Ehrenreich *et al.*, 2009 A multiparent advanced generation inter-cross to fine-map quantitative traits in *Arabidopsis thaliana*. *PLoS Genet.* 5: e1000551.
- Kuo, L., and B. Mallick, 1998 Variable selection for regression models. *Sankhya Ser. B* 60: 65–81.
- Ladejobi, O., J. Elderfield, K. A. Gardner, R. C. Gaynor, J. Hickey *et al.*, 2016 Maximizing the potential of multi-parental crop populations. *Appl. Transl. Genomics* 11: 9–17.
- Lander, E. S., and D. Botstein, 1989 Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121: 185–199.
- Lawson, H. A., J. M. Cheverud, and J. B. Wolf, 2013 Genomic imprinting and parent-of-origin effects on complex traits. *Nat. Rev. Genet.* 14: 609–617.
- Lenarcic, A. B., K. L. Svenson, G. A. Churchill, and W. Valdar, 2012 A general Bayesian approach to analyzing diallel crosses of inbred strains. *Genetics* 190: 413–435.
- Lu, Z., H. Zhu, R. C. Knickmeyer, P. F. Sullivan, W. N. Stephanie *et al.*, 2015 Multiple SNP-sets analysis for genome-wide association studies through Bayesian latent variable selection. *Genet. Epidemiol.* 39: 664–677.
- Manly, K. F., and J. M. Olson, 1999 Overview of QTL mapping software and introduction to map manager QT. *Mamm. Genome* 10: 327–334.
- Mitchell, T. J., and J. J. Beauchamp, 1988 Bayesian variable selection in linear regression. *J. Am. Stat. Assoc.* 83: 1023–1032.
- Mott, R., W. Yuan, P. Kaisaki, X. Gan, J. Cleak *et al.*, 2014 The architecture of parent-of-origin effects in mice. *Cell* 156: 332–342.
- Rasmussen, A. L., A. Okumura, M. T. Ferris, R. Green, F. Feldmann *et al.*, 2014 Host genetic diversity enables Ebola hemorrhagic fever pathogenesis and resistance. *Science* 346: 987–991.
- ter Braak, C. J. F., M. P. Boer, and M. C. A. M. Bink, 2005 Extending Xu’s Bayesian model for estimating polygenic effects using markers of the entire genome. *Genetics* 170: 1435–1438.
- Threadgill, D. W., and G. A. Churchill, 2012 Ten years of the collaborative cross. *Genetics* 190: 291–294.
- Threadgill, D. W., D. R. Miller, G. A. Churchill, and F. P.-M. de Villena, 2011 The collaborative cross: a recombinant inbred mouse population for the systems genetic era. *ILAR J.* 52: 24–31.
- Tsaih, S.-W., L. Lu, D. C. Airey, R. W. Williams, and G. A. Churchill, 2005 Quantitative trait mapping in a diallel cross of recombinant inbred lines. *Mamm. Genome* 16: 344–355.
- Vered, K., C. Durrant, R. Mott, and F. A. Iraqi, 2014 Susceptibility to *Klebsiella pneumoniae* infection in collaborative cross mice is a complex trait controlled by at least three loci acting at different time points. *BMC Genomics* 15: 865.
- Wei, J., and S. Xu, 2016 A random-model approach to QTL mapping in multiparent advanced generation intercross (MAGIC) populations. *Genetics* 202: 471–486.
- Woods, L. C. S., K. L. Holl, D. Oreper, Y. Xie, S.-W. Tsaih *et al.*, 2012 Fine-mapping diabetes-related traits, including insulin resistance, in heterogeneous stock rats. *Physiol. Genomics* 44: 1013–1026.
- Wu, M. C., P. Kraft, M. P. Epstein, D. M. Taylor, S. J. Chanock *et al.*, 2010 Powerful SNP-set analysis for case-control genome-wide association studies. *Am. J. Hum. Genet.* 86: 929–942.
- Yang, J., S. H. Lee, M. E. Goddard, and P. M. Visscher, 2011 GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* 88: 76–82.
- Yu, J., J. B. Holland, M. D. McMullen, and E. S. Buckler, 2008 Genetic design and statistical power of nested association mapping in maize. *Genetics* 178: 539–551.
- Yuan, F., F. Zou, and Y. Liu, 2011 Bayesian multiple quantitative trait loci mapping for recombinant inbred intercrosses. *Genetics* 188: 189–195.
- Zeng, Z.-B., 1994 Precision mapping of quantitative trait loci. *Genetics* 136: 1457–1468.
- Zhang, Z., W. Wang, and W. Valdar, 2014 Bayesian modeling of haplotype effects in multiparent populations. *Genetics* 198: 139–156.
- Zou, F., J. A. L. Gelfond, D. C. Airey, L. Lu, K. F. Manly *et al.*, 2005 Quantitative trait locus analysis using recombinant inbred intercrosses: theoretical and empirical considerations. *Genetics* 170: 1299–1311.

Communicating editor: D. J. de Koning