DOI: 10.1534/genetics.110.125542

Bayesian Multiple Quantitative Trait Loci Mapping for Recombinant Inbred Intercrosses

Zhongshang Yuan,*,† Fei Zou†,‡,1 and Yanyan Liu*

**School of Mathematics and Statistics, Wuhan University, Wuhan, Hubei 430072, China and* †*Department of Biostatistics,* ‡*Carolina Center for Genome Sciences, University of North Carolina, Chapel Hill, North Carolina 27599*

Manuscript received November 30, 2010
Accepted for publication February 17, 2011

ABSTRACT

The Collaborative Cross (CC) is a renewable mouse resource that mimics the genetic diversity in humans. The recombinant inbred intercrosses (RIX) generated from CC recombinant inbred (RI) lines share similar genetic structures to those of F_2 individuals. In contrast to F_2 mice, genotypes of RIX can be inferred from the genotypes of their RI parents and can be produced repeatedly. Also, RIX mice do not typically share the same degree of relatedness. This unbalanced genetic relatedness requires careful statistical modeling to avoid a large number of false positive findings. For complex traits, mapping multiple genes simultaneously is arguably more powerful than mapping one gene at a time. In this article, we describe how we have developed a Bayesian quantitative trait locus (QTL) mapping method that simultaneously deals with the special genetic architecture of RIX and maps multiple genes. The performance of the proposed method is evaluated by extensive simulations. In addition, for a given set of RI lines, there are numerous ways to generate RIX samples. To provide a general guideline on future RIX studies, we compare several RIX designs through simulations.

THIS study was motivated by the mouse Collaborative Cross (CC) project. As its main thrust, the CC project seeks to overcome several of the severe limitations of human complex trait studies. One major limitation of such studies is that only the simplest genetic models can be resolved with confidence. Many other models are plausible, but difficult or impossible to resolve. Because it is closely related to humans, the mouse serves as a great model organism for studying human diseases. Most human genes have functional mouse counterparts, and genomes of both organisms are organized similarly. Traditional RI lines are generated by crossing only two inbred parental strains, resulting in extensive blind spots where a sizable proportion of the genome is identical by decent (IBD) and a low percentage of genetic variation (15%). In contrast, the CC recombinant inbred (RI) lines are derived from a genetically diverse set of eight founder inbred strains (Chesler *et al.* 2008; Iraqi *et al.* 2008). This selection of founder strains was predicted to result in uniform genome-wide high levels of variation. Genetic variation is randomized in the CC lines so that the causal relationships can be established. Another major advantage of the CC mice is that they are retrievable and extensible, which supports data integration over space and time and at all levels—from molecules and cells to physiological systems and environments.

Recombinant inbred intercross (RIX) panels are created by generating diallel F_1 progeny from a set of RI lines. Given a population of L RI lines, this approach can potentially produce $L(L-1)/2$ genetically distinct RIX individuals. Subsets of RIX can be used to evaluate biological predictions of the ways in which an individual would respond to environmental perturbations and provide statistical support for prediction accuracy. Because RI mice are homozygote at each locus, the genotypes of the derivative RIX mice can be imputed in advance from the genotypes of the parental RI lines. RIX mice with identical genotypes can be regenerated whenever needed. Therefore the CC mice provide us with an ideal platform to generate testable hypotheses that can be used for accurate probability-based whole-organism biological predictions. Preliminary data from the CC strains and their derivative RIX progenies are now being generated. Developing appropriate analysis tools for mapping CC data is critical for the success of the CC project.

Though at the individual level, the genome of each RIX mouse has similar genetic structures to those of F_2 individuals, statistical analyses for F_2 data cannot be applied directly to RIX data. Backcross or F_2 subjects used in traditional quantitative trait locus (QTL) mapping have the same genetic relatedness to each other, which may not be the case for RIX data. Some RIX mice

Supporting information is available online at http://www.genetics.org/cgi/content/full/genetics.110.125542/DC1.

[1]*Corresponding author:* Department of Biostatistics, University of North Carolina, 4115D McGavran-Greenberg Hall, CB 7420, Chapel Hill, NC 27599. E-mail: fzou@bios.unc.edu

may share common parental RI lines, making them more related to each other than the RIX mice that do not share parental RI lines. Past investigations (Tsaih *et al.* 2005; Zou *et al.* 2005) have shown that careful statistical analyses that correctly handle the unbalanced relatedness in RIX are important to achieve valid results.

Over the last decade, mapping genes of human diseases has been an important research area. The vast majority of genes, however, have been limited to simple Mendelian inheritance patterns and well-defined quantitative traits with relatively large and consistent effects. Recent emphasis has shifted from mapping simple Mendelian traits to complex traits with complex genetic causes. Complex traits have proved far more resistant to genetic analysis. Many available QTL mapping methods map only one or a few QTL at a time and may not be efficient for complex traits. Variable selection methods are useful in selecting variables that are not necessarily individually important but that are, in fact, jointly important. By treating multiple QTL mapping as a model/variable selection problem (Broman and Speed 2002), forward and stepwise selection procedures have been proposed to search for multiple QTL. Although simple, these methods have limitations, such as uncertainty about the number of QTL and difficulty in assessing the significance of associated tests due to their sequential model building nature. Bayesian QTL mapping methods (Satagopan *et al.* 1996; Sillanpää and Arjas 1998; Stephens and Fisch 1998; Yi and Xu 2000, 2001; Hoeschele 2007) have been developed, in particular, to detect multiple QTL by treating the number of QTL as a random variable and modeling it specifically using reversible-jump Markov chain Monte Carlo (MCMC) (Green 1995). Due to the variable dimensionality of the parameter spaces associated with different models (different numbers of QTL), one must take care of acceptance probabilities for such changes in dimension, which in practice may not be handled correctly (Ven 2004). One leading approach to variable selection in QTL analysis is the Bayesian analysis based on the composite model space framework (Godsill 2001, 2003), first introduced by Yi (2004) to the genetic mapping community. Reversible-jump MCMC (Green 1995) and stochastic search variable selection (SSVS) (George and McCulloch 1993) are special cases of this analysis. An alternative to variable selection with SSVS is shrinkage methods (*e.g.*, the lasso method of Tibshirani 1996), which shrink the effects of unimportant variables near zero. The Bayesian shrinkage methods of Xu (2003) and Wang *et al.* (2005) place simple normal priors on QTL effects. While they can be modified easily, the specific priors used in Xu (2003) and Wang *et al.* (2005) induce improper posteriors (Ter Braak *et al.* 2005).

In this article, for RIX data, we extend the fixed-effect model of Xu (2003) to a mixed-effects model, where the complex relationship structure among RIX samples is specifically dealt with through random polygenic effects. To avoid improper posteriors, we apply the prior modification of ter Braak *et al.* (2005). To the best of our knowledge, this is the first simultaneous multiple-QTL mapping method specifically designed for RIX data. This article is organized as follows. In statistical method, we first introduce the RIX experiment, and then we propose a mixed-effects model and describe a Bayesian variable selection procedure. In simulation study, extensive simulations are performed to evaluate the proposed model and to compare several RIX designs. Summary comments are given in the discussion.

STATISTICAL METHOD

In this section, after describing the RIX design, we propose the multiple-QTL model and introduce the Bayesian variable selection procedure.

RIX design: The RIX panel is created as F_1 progenies of RI intercross. For L RI lines, a total of $L(L-1)$ reciprocal RIX and $L(L-1)/2$ nonreciprocal RIX can be generated (see Figure 1 in Zou *et al.* 2005). For simplification and without loss of generality, in the sequel, we consider the nonreciprocal RIX. Let the parental RI lines be $RI_1, RI_2, \ldots, RI_L$, respectively, and denote the derived RIX from the $RI_i \times RI_j$ mating as RIX_{ij}, where $i, j = 1, \ldots, L$ with $i < j$ or, alternatively, as RIX_k, where $k = 1, 2, \ldots, L(L-1)/2$ for ease of notation. Let n ($\leq L(L-1)/2$) be the total number of RIX sampled and p be the total number of genetic markers. Further, let the phenotypes (the dependent variable) be y_i ($i = 1, \ldots, n$) and the discrete marker genotypes be m_{ij} ($i = 1, \ldots, n$, $j = 1, \ldots, p$). We set m_{ij} to -1, 0, and 1, referring to genotypes *aa*, *Aa*, and *AA*, respectively.

Model: It is often acknowledged that quantitative traits are controlled by both major genes and polygenes, that is, genes with intermediate and small effects, respectively. The aggregating effect of polygenes may greatly reduce our ability to map major genes if not modeled carefully. Visscher and Haley (1996) modeled polygenic effects for data derived from commonly used experimental crosses. Methods using Wright's relationship matrix to accommodate different correlations between related individuals have also been developed for pedigree data and diallel data (Goldgar 1990; Amos 1994; Zhu and Weir 1996; Xu 1998).

RIX data can be viewed as the last generation of a large pedigree that originated from a set of inbred founder strains. Therefore, methods for analyzing pedigree data can be directly applied. In our model, we assume that major QTL and polygenes affect the trait of interest independently, and the aggregating polygenic effect is normally distributed. For simplicity, we assume that all putative QTL are located on markers, a reasonable assumption with dense maps. For sparse maps, we can employ interval mapping ideas of Wang *et al.* (2005) and Huang *et al.* (2010). Specifically, we fit the mixed-effects model

$$y_i = \mu + \sum_{j=1}^{p}\left[x_{ij}a_j + w_{ij}b_j\right] + \sum_{k=1}^{L} a_{ik}\alpha_k + e_i, \qquad (1)$$

where μ is the overall mean, a_j and b_j are the additive and dominant effects of the jth putative QTL with $x_{ij} = \sqrt{2}m_{ij}$ and $w_{ij} = 1 - 2|m_{ij}|$, a_{ik} is the number of parents of RIX_i that are equal to RI_k, and the random polygenic effect α_k follows $N(0, \sigma_a^2)$ $(k = 1, 2, \ldots, L)$ and the random error e_i follows $N(0, \sigma_0^2)$ $(i = 1, 2, \ldots, n)$ and they are independent of each other. Let $\mathbf{A}$ be an $n \times L$ matrix whose (i, k)th element equals a_{ik}. Clearly, $\sum_{k=1}^{L} a_{ik} \equiv 2$ for all $i = 1, 2, \ldots, n$ since each RIX has two and only two parents. In the model above, we assume that the polygenic effect is additive, which can be easily extended by adding a random dominance polygenic term in model (1).

In the above model, the observed data are $\mathbf{y} = \{y_i\}$, marker genotypes $\mathbf{M} = \{m_{ij}\}$, and parental RI information (*i.e.*, matrix $\mathbf{A}$). The unobserved variables include μ, σ_0^2, and σ_a^2; the regression coefficients $\mathbf{a} = \{a_j\}$, $\mathbf{b} = \{b_j\}$; and the random effects $\boldsymbol{\alpha} = \{\alpha_k\}$. Below we describe the prior distributions of the unobserved variables.

Prior specifications:

1. For μ: $p(\mu) \propto 1$.
2. For the a_j's and b_j's: $a_j \sim N(0, \sigma_j^2)$, $b_j \sim N(0, v_j^2)$.
3. For σ_0^2: $p(\sigma_0^2) \propto 1/\sigma_0^2$.
4. For σ_a^2: $p(\sigma_a^2) \propto (\sigma_a^2)^{\delta-1}$.

Further, for hyperparameters σ_j^2 and v_j^2, we assign the following priors:

5. $p(\sigma_j^2) \propto (\sigma_j^2)^{\delta-1}$ and $p(v_j^2) \propto (v_j^2)^{\delta-1}$.

Here δ $(0 < \delta \le \frac{1}{2})$ is used to ensure that the posterior distribution is proper (ter Braak *et al.* 2005).

Let $\sigma^2 = \left\{\sigma_j^2\right\}$ and $v^2 = \left\{v_j^2\right\}$.

The specific priors above result in known marginal conditional distributions for all variables; therefore simple Gibbs sampling can be performed, which we describe below.

MCMC algorithm for posterior computation: We first initiate μ, $\mathbf{a}$, $\mathbf{b}$, σ_0^2, σ_a^2, $\boldsymbol{\sigma}^2$, and $\mathbf{v}^2$. Since the joint posterior distribution

$$p(\mu, \mathbf{a}, \mathbf{b}, \boldsymbol{\alpha}, \mathbf{v}^2, \boldsymbol{\sigma}^2, \sigma_0^2, \sigma_a^2 \,|\, \mathbf{y}) \propto p(\mathbf{y}|\mu, \mathbf{a}, \mathbf{b}, \boldsymbol{\alpha}, \mathbf{v}^2, \boldsymbol{\sigma}^2, \sigma_0^2, \sigma_a^2) p(\mu, \mathbf{a}, \mathbf{b}, \boldsymbol{\alpha}, \mathbf{v}^2, \boldsymbol{\sigma}^2, \sigma_0^2, \sigma_a^2),$$

where

$$p(\mathbf{y} \,|\, \mu, \mathbf{a}, \mathbf{b}, \boldsymbol{\alpha}, \mathbf{v}^2, \boldsymbol{\sigma}^2, \sigma_0^2, \sigma_a^2) \propto (\sigma_0^2)^{-n/2} \times \exp\left\{ -\frac{\sum_{i=1}^{n}\left(y_i - \mu - \sum_{j=1}^{P}\left[x_{ij}a_j + w_{ij}b_j\right] - \sum_{k=1}^{L} a_{ik}\alpha_k\right)^2}{2\sigma_0^2} \right\},$$

we perform the following Gibbs sampling steps. Superscripts (t) and $(t + 1)$ signify the MCMC iteration steps and we set $t = 0$ for all initial values.

Step 1. Updating μ: $\mu^{(t+1)}$ is drawn from the normal distribution with mean $(1/n)\sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{P}[x_{ij}a_j^{(t)} + w_{ij}b_j^{(t)}] - \sum_{k=1}^{L} a_{ik}\alpha_k^{(t)}\right)$ and variance $\sigma_0^{2(t)}/n$.

Step 2. Updating a_j and b_j $(j = 1, \ldots, p)$, $a_j^{(t+1)}$ is sampled from the normal distribution with mean

$$\left(\sum_{i=1}^{n} x_{ij}^2 + \frac{\sigma_0^{2(t)}}{\sigma_j^{2(t)}}\right)^{-1} \sum_{i=1}^{n} x_{ij}\Big(y_i - \mu^{(t+1)} - \sum_{j=1}^{P} w_{ij}b_j^{(t)} - \sum_{k=1}^{L} a_{ik}\alpha_k^{(t)} - \sum_{l<j} x_{il}a_l^{(t+1)} - \sum_{l>j} x_{il}a_l^{(t)}\Big)$$

and variance

$$\left(\sum_{i=1}^{n} x_{ij}^2 + \frac{\sigma_0^{2(t)}}{\sigma_j^{2(t)}}\right)^{-1} \sigma_0^{2(t)},$$

while $b_j^{(t+1)}$ is drawn from the normal distribution with mean

$$\left(\sum_{i=1}^{n} w_{ij}^2 + \frac{\sigma_0^{2(t)}}{v_j^{2(t)}}\right)^{-1} \sum_{i=1}^{n} w_{ij}\Big(y_i - \mu^{(t+1)} - \sum_{j=1}^{P} x_{ij}a_j^{(t+1)} - \sum_{k=1}^{L} a_{ik}\alpha_k^{(t)} - \sum_{l<j} w_{il}b_l^{(t+1)} - \sum_{l>j} w_{il}b_l^{(t)}\Big)$$

and variance

$$\left(\sum_{i=1}^{n} w_{ij}^2 + \frac{\sigma_0^{2(t)}}{v_j^{2(t)}}\right)^{-1} \sigma_0^{2(t)}.$$

Step 3. Updating σ_0^2: The residual variance is sampled from the scale-inverted χ^2-distribution, $\sum_{i=1}^{n}(y_i - \mu^{(t+1)} - \sum_{j=1}^{P} x_{ij}a_j^{(t+1)} - \sum_{k=1}^{L} a_{ik}\alpha_k^{(t)} - \sum_{j=1}^{P} w_{ij}b_j^{(t+1)})^2/\chi_n^2$.

Step 4. Updating σ_j^2 and v_j^2 $(j = 1, \ldots, p)$, $\sigma_j^{2(t+1)}$ is sampled from the scale-inverted χ^2-distribution, $a_j^{2(t+1)}/\chi_{1-2\delta}^2$, and $v_j^{2(t+1)}$ is sampled from the scale-inverted χ^2-distribution, $b_j^{2(t+1)}/\chi_{1-2\delta}^2$.

Step 5. Updating α_k $(k = 1, \ldots, L)$, $\alpha_k^{(t+1)}$ is drawn from the normal distribution with mean

$$\left(\sum_{i=1}^{n} a_{ik}^2 + \frac{\sigma_0^{2(t+1)}}{\sigma_a^{2(t)}}\right)^{-1} \sum_{i=1}^{n} a_{ik}\Big(y_i - \mu^{(t+1)} - \sum_{j=1}^{P} x_{ij}a_j^{(t+1)} - \sum_{j=1}^{P} w_{ij}b_j^{(t+1)} - \sum_{l<k} a_{il}\alpha_l^{(t+1)} - \sum_{l>k} a_{il}\alpha_l^{(t)}\Big)$$

and variance

$$\left(\sum_{i=1}^{n} a_{ik}^2 + \frac{\sigma_0^{2(t+1)}}{\sigma_a^{2(t)}}\right)^{-1} \sigma_0^{2(t+1)}.$$

TABLE 1

Locations and sizes of the four simulated QTL

QTL no.	Position (cM)	Additive effect	Dominant effect
1	0	2	2
2	250	1	1
3	500	2	0
4	750	0	2

Step 6. Updating σ_a^2, the random effect variance is sampled from the scale-inverted χ^2-distribution, $\sum_{k=1}^{L}\alpha_k^{2(t+1)}/\chi^2_{L-2\delta}$.

After this round of sampling, we complete one sweep of the MCMC and are ready to continue our sampling for the next round by repeating steps 1–6 with the new parameter values. When the chain converges, the sampled parameters approximately follow the joint posterior distribution. For any parameter of interest, one can easily obtain, for example, its posterior means and variances from the joint posterior sample.

SIMULATION STUDY

In this section, we ran extensive simulations to evaluate the performance of the proposed Bayesian method. Specifically, we compared the proposed mixed-effects model with the linear regression model, which ignores the polygenic effect. We also investigated several RIX designs and compared their abilities in mapping major QTL.

Parental RI genotypes were simulated in R/qtl (Broman *et al.* 2003), from which RIX genotypes and phenotypes were generated accordingly. For all simulations, we set the true population mean (μ) and residual variance (σ_0^2) to 5.0 and 10, respectively. A single chromosome with total length 15 M was simulated, on which 301 evenly spaced markers (resulting in 300 5-cM intervals) are located. The number of QTL simulated was 4 or 11, and their corresponding genetic effects and locations are summarized in Tables 1 and 2. Three σ_a^2 values, 0, 1, and 8, were simulated, representing the situations where no, low, and high polygenic effects exist, respectively.

From 100 RI parental lines, a total of 5000 unique nonreciprocal RIX can be produced. For a realistic sample size n that we set to 300 unless specified, we considered the following five designs: (1) the *complete-pair design*, selecting 25 RI lines randomly and generating all 300 possible nonreciprocal RIX; (2) the *loop design* (a clockwise mating scheme described in Zou *et al.* 2005), ordering the 100 RI lines randomly to form a circle and then mating each RI line (clockwise) with the next 3 RI lines after it (*i.e.*, RI_1 mated with RI_2, RI_3, and RI_4 and RI_2 mated with RI_3, RI_4, and RI_5, and so on); (3) the *50 complete-independent design*, ordering the 100 RI lines randomly and mating RI_1 with RI_2, RI_3 with RI_4, RI_5 with RI_6, and so on. The maximal number of RIX that can be generated from this design is $L/2$. To achieve the sample size of 300, we may alternatively employ (4) the *replicated-complete-independent design*, sampling six offspring instead of one from each $RI_i \times RI_j$ mating in design 3, or (5) the *300 complete-independent design*, generating 300 independent RIX samples from 600 RI lines as in design 3.

TABLE 2

Locations and sizes of the 11 simulated QTL

QTL no.	Position (cM)	Additive effect	Dominant effect
1	0	$\sqrt{10}$	$\sqrt{10}$
2	150	$\sqrt{5}$	$\sqrt{5}$
3	300	$\sqrt{2.5}$	$\sqrt{2.5}$
4	450	$\sqrt{2}$	$\sqrt{2}$
5	600	$\sqrt{1.25}$	$\sqrt{1.25}$
6	750	$\sqrt{1.25}$	$\sqrt{1.25}$
7	900	$\sqrt{0.625}$	$\sqrt{0.625}$
8	1050	$\sqrt{0.625}$	$\sqrt{0.625}$
9	1200	$\sqrt{0.25}$	$\sqrt{0.25}$
10	1350	$\sqrt{0.25}$	$\sqrt{0.25}$
11	1500	$\sqrt{0.25}$	$\sqrt{0.25}$

For each simulated datum, the MCMC sampler was run for a total of 63,000 cycles. We set the initial values of μ, **a**, and **b** to 0. The initial values of σ_0^2, σ_a^2, $\boldsymbol{\sigma}^2$, and $\mathbf{v}^2$ were all set to 1. We further set δ to 0.001. The first 3000 cycles were discarded as burn-in, and the remainder of the chain was thinned by keeping 1 of every 50 samples, resulting in a total of 1200 samples for post-MCMC analysis. All the analysis was done in MATLAB and the MATLAB source code is available at http://bios.unc.edu/~fzou/software/BayesianRIX.

We also analyzed each datum with the linear model

$$y_i = \mu + \sum_{j=1}^{p}\left[x_{ij}a_j + w_{ij}b_j\right] + e_i, \qquad (2)$$

the gold model for RIX data when the true polygenic effect is zero. All the priors in this linear model were set the same as their counterparts in model (1).

Case 1. Four QTL and $\sigma_a^2 = 0$: Under this setup, we expect model (2) to perform better than the mixed-effects model (1). The question, more specifically, concerns how much better the linear model (2) is. Figure 1 displays the posterior mean estimates for one simulated datum. The two models perform nearly the same, indicating very little power loss of the proposed mixed-effects model.

Case 2. Four QTL and $\sigma_a^2 = 8$: When the polygenic effect is not zero, Zou *et al.* (2005) and Tsaih *et al.* (2005) have shown that linear models that ignore the unbalanced relatedness of RIX result in high false positive rates. Figure 2 clearly reflects this. The linear model produces large estimated additive effects for many null

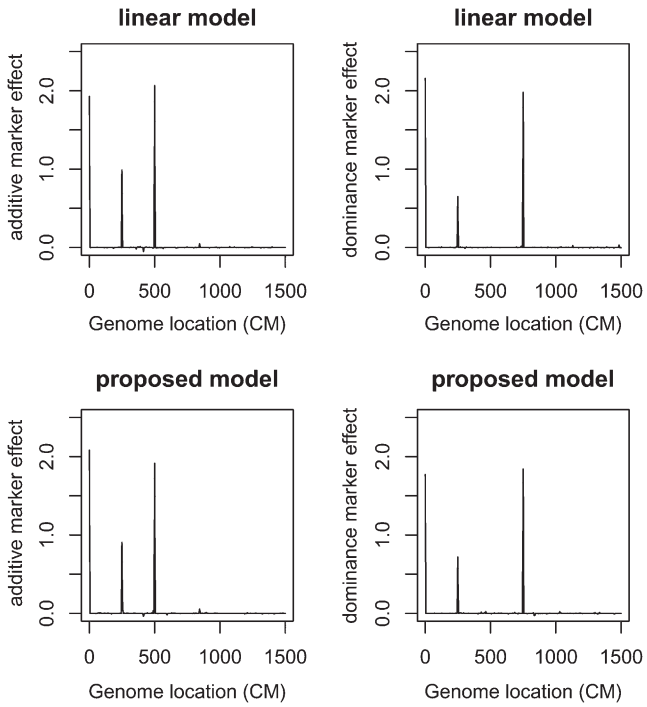

FIGURE 1.—Bayesian QTL estimates from data simulated under case 1 where no polygenic effect exists (*i.e.*, $\sigma_a^2 = 0$). The top two panels are from the linear model, and the bottom two panels are from the proposed mixed-effects model. The height of each vertical line refers the posterior mean estimate of each marker.

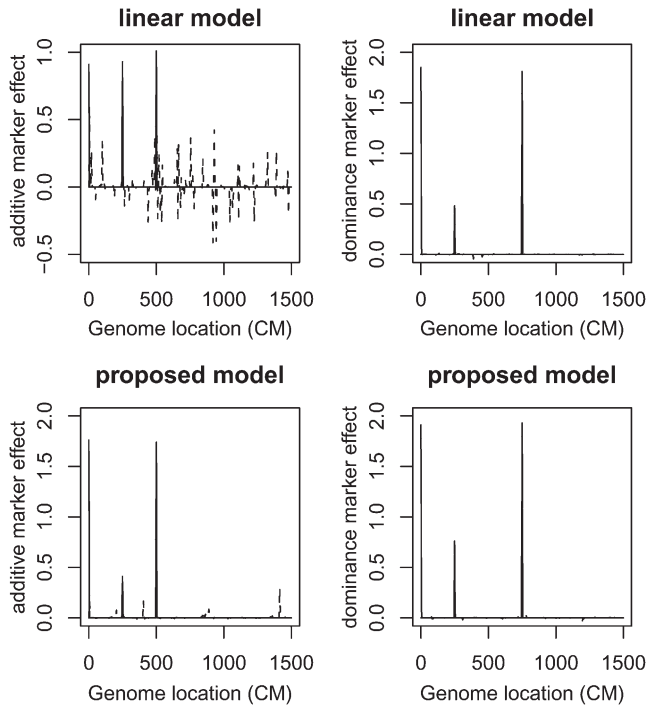

FIGURE 2.—Bayesian QTL estimates from data simulated under case 2 where polygenic effect exists ($\sigma_a^2 = 8$). The solid lines are for the markers located at the simulated QTL and the dashed lines are for the rest of the markers. See Figure 1 legend for details.

markers. In contrast, the four simulated QTL are mapped successfully by the mixed-effects model and most of the null markers have their posterior means close to 0. Table 3 shows that for the four simulated QTL, their estimated genetic effects from the mixed-effects model are close to those simulated. The second QTL located at 250 cM is the weakest and its estimated genetic effects depart most from its true genetic values.

In summary, the mixed-effects model performs substantially better than the linear model. Interestingly, there is no significant difference between the two models for detecting dominance effects. This may be attributed to the fact that the simulated polygenic effect is additive, which is less likely confounded with the dominance QTL effects.

Case 3. Eleven QTL and $\sigma_a^2 = 1$: Under this setup, the heritabilities of 11 QTL decrease from left to right. Figure 3 summarizes the analysis results. The estimated genetic effects of the top 6 QTL are significantly different from 0, and the smallest QTL that the proposed method identified explains ~2% of the phenotypic variation.

All the above results are based on data generated from the loop design. Next, we compared the loop design with the other four designs. For each design, we simulated 100 data sets with the genetic configuration from case 2 (*i.e.*, four QTL and $\sigma_a^2 = 8$). The data from the two complete-independent designs were analyzed by the linear model, the correct model for the data. For analysis of the data from the loop, the 300 replicated-complete-independent, and the complete-pair designs,

TABLE 3

Estimated QTL effects from data under case 2 with four QTL and $\sigma_a^2 = 8$

		Additive effect				Dominant effect			
QTL	Position (cM)	True	a	a_L	a_U	True	b	b_L	b_U
1	0	2	1.76	0.52	2.86	2	1.91	1.54	2.28
2	250	1	0.41	0	1.17	1	0.77	0.40	1.14
3	500	2	1.75	0.65	2.74	0	0	0	0
4	750	0	0	0	0	2	1.93	1.57	2.29

a, posterior mean of additive effect; a_L and a_U, the 5th and 95th percentiles of posterior samples for the additive effect. Similarly, b, b_L, and b_U are defined for the dominant effect.

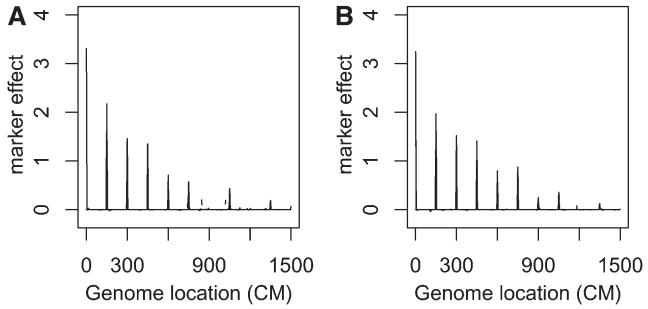

FIGURE 3.—Bayesian QTL estimates from data simulated under case 3. (A) The additive effect; (B) the dominant effect. See Figure 2 legend for details.

we used the mixed-effects model. For the data generated from the 300 replicated-complete-independent design, we also averaged the phenotypes across the six offspring from each $RI_i \times RI_j$ mating and analyzed the 50 independent averages by the linear model. Table 4 summarizes the mean squared error (MSE) of the posterior mean estimates of the additive QTL effects. As expected, the data from the 300 complete-independent design perform the best. The loop design is ranked second best and has much higher efficiency than the rest of the designs. Interestingly, though the sample size of the complete-pair design is 300, its power for mapping major QTL is only slightly higher than that of the 50 complete-independent design where the sample size is 50. This is largely due to the fact that we have a large polygenic effect. In summary, for a fixed sample size, we recommend sampling as many independent RIX as possible. If the number of independent RIX is too small, the loop design should be our next choice. One advantage of the loop design over the independent RIX design is that it allows estimation of the polygenic effect but the independent RIX design does not.

DISCUSSION

Recombinant inbred intercrosses, produced by generating all or a subset of the potential F_1 hybrids between pairs of RI lines, increase the number of available genotypes from L RI lines to $L(L-1)/2$ nonreciprocal RIX and $L(L-1)$ reciprocal RIX. Unlike the parental RI lines whose genotypes are homozygous, the genetic structure of RIX resembles that of F_2 animals, reducing the phenotypic anomalies associated with inbred genomes. One of the great achievements of using RIX animals for QTL mapping is the ongoing CC project (THREADGILL *et al.* 2002). The CC project plans to generate and maintain ~1000 CC RI lines for scientific research. With such large numbers of RI lines available, our ability to map complex traits can be greatly increased. Because RIX genotypes can be directly inferred from the genotypes of their parental RIs, the genotyping effort required for the CC project is to genotype only the CC RI lines. Further, due to the renewability of CC mice, new RIX mice can be selectively produced for testing the accuracies of estimated genetic architectures.

In all our simulations, the parameter δ is set to 0.001 to ensure a proper posterior distribution. We have also analyzed the simulated data with $\delta = 0$. The new analysis results are shown in supporting information, File S1, Figure S1, and Figure S2. Though in theory, the posterior distribution from the priors with $\delta = 0$ is improper (TER BRAAK *et al.* 2005), this adjustment makes very little difference in practice.

In this article, we have compared several RIX designs. For a given sample size, the complete-independent design performs the best. But the complete-independent design is limited by the number of available RI lines and lacks ability in estimating polygenic effects. In contrast, the loop design and the complete-pair design can generate large numbers of RIX. Between the two, the loop design performs better. We highly recommend the use of the loop design when the number of RI lines is limited. The loop design also provides estimation of polygenic effects for the heritability calculation.

For simplicity, our model assumes no maternal or paternal effects and we consider only nonreciprocal RIX. One major advantage of RIX over RI or F_2 populations is that parent-of-origin effects can be tested with reciprocal RIX. Now we briefly describe how to extend the proposed model for parent-of-origin effects.

TABLE 4

Comparison of the five designs based on 100 simulations under case 2

Design	Position (0 cM)	Position (250 cM)	Position (500 cM)
300 complete independent	0.17	0.57	0.19
Loop	1.30	0.89	1.46
Replicated complete independent[a]	3.36	0.94	3.33
Replicated complete independent[b]	3.45	0.95	3.34
Complete pair	3.50	0.98	3.38
50 complete independent	3.54	1.00	3.58

For each QTL, the mean squared error (MSE) of the posterior mean estimate of the additive effect is reported.
[a]Individual phenotypes of RIX samples were analyzed via the proposed mixed-effects model (1).
[b]Average phenotypes of replicated RIX samples were analyzed via the linear model (2).

For the jth QTL with additive parent-of-origin effect, we replace the term $x_{ij}a_j$ in (1) with $x_{ij(p)}a_{jp} + x_{ij(m)}a_{jm}$, where $x_{ij(p)}$ [and $x_{ij(m)}$] = 0 or 1 depending on whether RIX_i gets an A allele from its father (and mother). Any deviation of $|a_{jp} - a_{jm}|$ from 0 suggests a parent-of-origin effect. Polygenic parent-of-origin effects can be similarly modeled. Further, our model basically considers only nucleotide effects, which can be extended as well for modeling founder allelic effects. For example, for the additive effects of the eight CC founder alleles, we can replace the term $x_{ij}a_j + w_{ij}b_j$ in (1) with $\mathbf{x}_{ij}\boldsymbol{\beta}_j$, where $\boldsymbol{\beta}_j = (\beta_{1j}, \ldots, \beta_{8j})^T$ and the kth element of $\mathbf{x}_{ij}$ is 2, 1, or 0 depending on whether RIX_i inherits 2, 1, or 0 copies of the kth founder allele ($k = 1, \ldots, 8$) at the jth locus. Therefore, β_{kj} represents the kth founder allelic effect. Similarly, for a codominant QTL model, we set $\boldsymbol{\beta}_j$ as a vector of length 36 (corresponding to the 36 genotypes formed by the eight CC founder alleles). We can further treat the QTL effects as random as traditionally done for pedigree data. For example, for an additive model, we let $\boldsymbol{\beta}_j \sim N(0, \sigma^2_{aj}\mathbf{I})$. This would potentially increase the mapping power as the number of parameters is dramatically reduced.

The authors are grateful for many constructive comments and suggestions from the reviewers and the associate editor. Support for this work was provided in part by National Institutes of Health grants R01GM074175, R01CA082659, and P50MH090338 to F. Zou and grants from the National Science Foundation of China (10771163) and the China Scholarship Council to Z. Yuan and Y. Liu.

LITERATURE CITED

Amos, C. I., 1994 Robust variance-components approach for assessing genetic linkage in pedigrees. Am. J. Hum. Genet. **54(3):** 535–543.

Broman, K. W., and T. P. Speed, 2002 A model selection approach for the identification of quantitative trait loci in experimental crosses (with discussion). J. R. Stat. Soc. B **64:** 641–656, 731–775.

Broman, K. W., H. Wu, S. Sen and G. A. Churchill, 2003 R/qtl: QTL mapping in experimental crosses. Bioinformatics. **19:** 889–890.

Chesler, E. J., D. R. Miller, L. R. Branstetter, L. D. Galloway, B. L. Jackson *et al.*, 2008 The Collaborative Cross at Oak Ridge National Laboratory: developing a powerful resource for systems genetics. Mamm. Genome. **19:** 382–389.

George, E. I., and R. E. McCulloch, 1993 Variable selection via Gibbs sampling. J. Am. Stat. Assoc. **88:** 881–889.

Godsill, S. J., 2001 On the relationship between Markov chain Monte Carlo methods for model uncertainty. J. Comput. Graph. Stat. **10:** 230–248.

Godsill, S. J., 2003 *Proposal Densities, and Product Space Methods, in Highly Structured Stochastic Systems.* Oxford University Press, London/New York/Oxford.

Goldgar, D. E., 1990 Multipoint analysis of human quantitative genetic variation. Am. J. Hum. Genet. **47:** 957–967.

Green, P. J., 1995 Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika **82:** 711–732.

Huang, H., H. Zhou, F. Cheng, I. Hoeschele and F. Zou, 2010 Gaussian process based Bayesian semiparametric quantitative trait loci interval mapping. Biometrics **66:** 222–232.

Hoeschele, I., 2007 Mapping quantitative trait loci in outbred populations, pp. 623–677 *in Handbook of Statistical Genetics*, Vol. 1, edited by D. J. Balding, M. Bishop and C. Cannings. John Wiley & Sons, New York.

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.

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.

Sillanpaa, M. J., and E. Arjas, 1998 Bayesian mapping of multiple quantitative trait loci from incomplete inbred line cross data. Genetics **148:** 1373–1388.

Stephens, D. A., and R. D. Fisch, 1998 Bayesian analysis of quantitative trait locus data using reversible jump Markov chain Monte Carlo. Biometrics **54:** 1334–1347.

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., K. W. Hunter and R. W. Williams, 2002 Genetic dissection of complex and quantitative traits: from fantasy to reality via a community effort. Mamm. Genome **13:** 175–178.

Tibshirani, R., 1996 Regression shrinkage and selection via the lasso. J. R. Stat. Soc. B **58:** 267–288.

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.

Ven, R. V., 2004 Reversible-jump Markov chain Monte Carlo for quantitative trait loci mapping. Genetics **167:** 1033–1035.

Visscher, P. M., and C. S. Haley, 1996 Detection of putative quantitative trait loci in line crosses under infinitesimal genetic models. Theor. Appl. Genet. **93:** 691–702.

Wang, H., Y. M. Zhang, X. Li, G. L. Masinde, S. Mohan *et al.*, 2005 Bayesian shrinkage estimation of quantitative trait loci parameters. Genetics **170:** 465–480.

Xu, S., 1998 Mapping quantitative loci using multiple families of line crosses. Genetics **148:** 517–524.

Xu, S., 2003 Estimating polygenic effects using markers of the entire genome. Genetics **163:** 789–801.

Yi, N., and S. Xu, 2000 Bayesian mapping of quantitative trait loci under the IBD-based variance. Genetics **156:** 411–422.

Yi, N., and S. Xu, 2001 Bayesian mapping of quantitative trait loci under complicated mating designs. Genetics **157:** 1759–1771.

Yi, N., 2004 A unified Markov chain Monte Carlo framework for mapping multiple quantitative trait loci. Genetics **167:** 967–975.

Zhu, J., and B. S. Weir, 1996 Mixed model approaches for diallel analysis based on a bio-model. Genet. Res. **68:** 233–240.

Zou, F., J. L. Gelfond, D. C. Airey, L. Lu, K. F. Manly *et al.*, 2005 Quantitative trait locus analysis using recombinant inbred intercrosses (RIX): theoretical and empirical considerations. Genetics **170:** 1299–1311.

Communicating editor: S. F. Chenoweth

GENETICS

Supporting Information
http://www.genetics.org/cgi/content/full/genetics.110.125542/DC1

Bayesian Multiple Quantitative Trait Loci Mapping for Recombinant Inbred Intercrosses

Zhongshang Yuan, Fei Zou and Yanyan Liu

DOI: 10.1534/genetics.110.125542

FILE S1

File S1 is available for download at http://www.genetics.org/cgi/content/full/genetics.110.125542/DC1. The compressed folder (.zip) contains the matlab code, instructions for use, and the input data file.

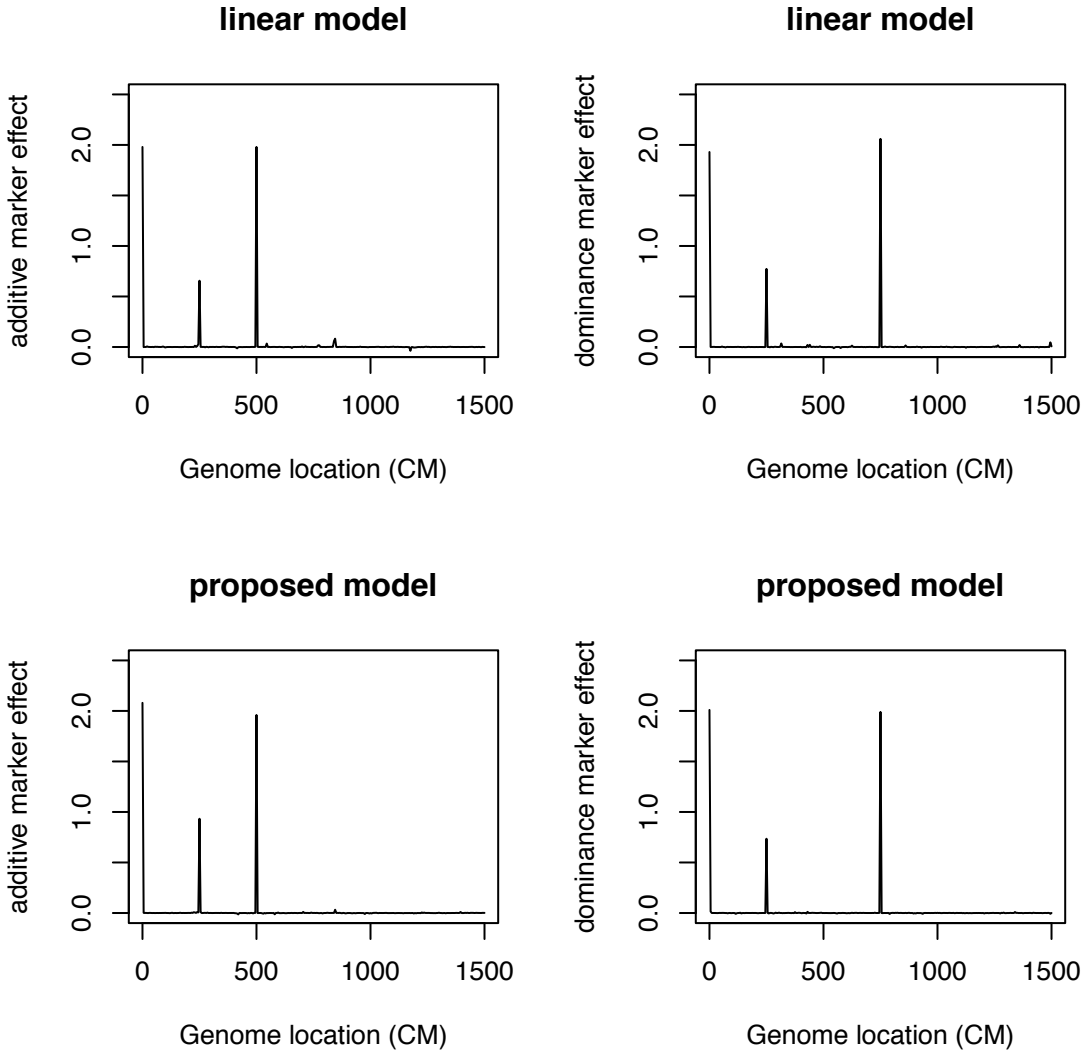

FIGURE S1.—Bayesian QTL estimates from data simulated under Case 1 where no polygenic effect exists (i.e., $\sigma_\alpha^2 = 0$). The parameter δ was set to 0. The top two panels are from the linear model, where the bottom two panels are from the proposed mixed model. The height of each vertical line refers the posterior mean estimate of each marker.

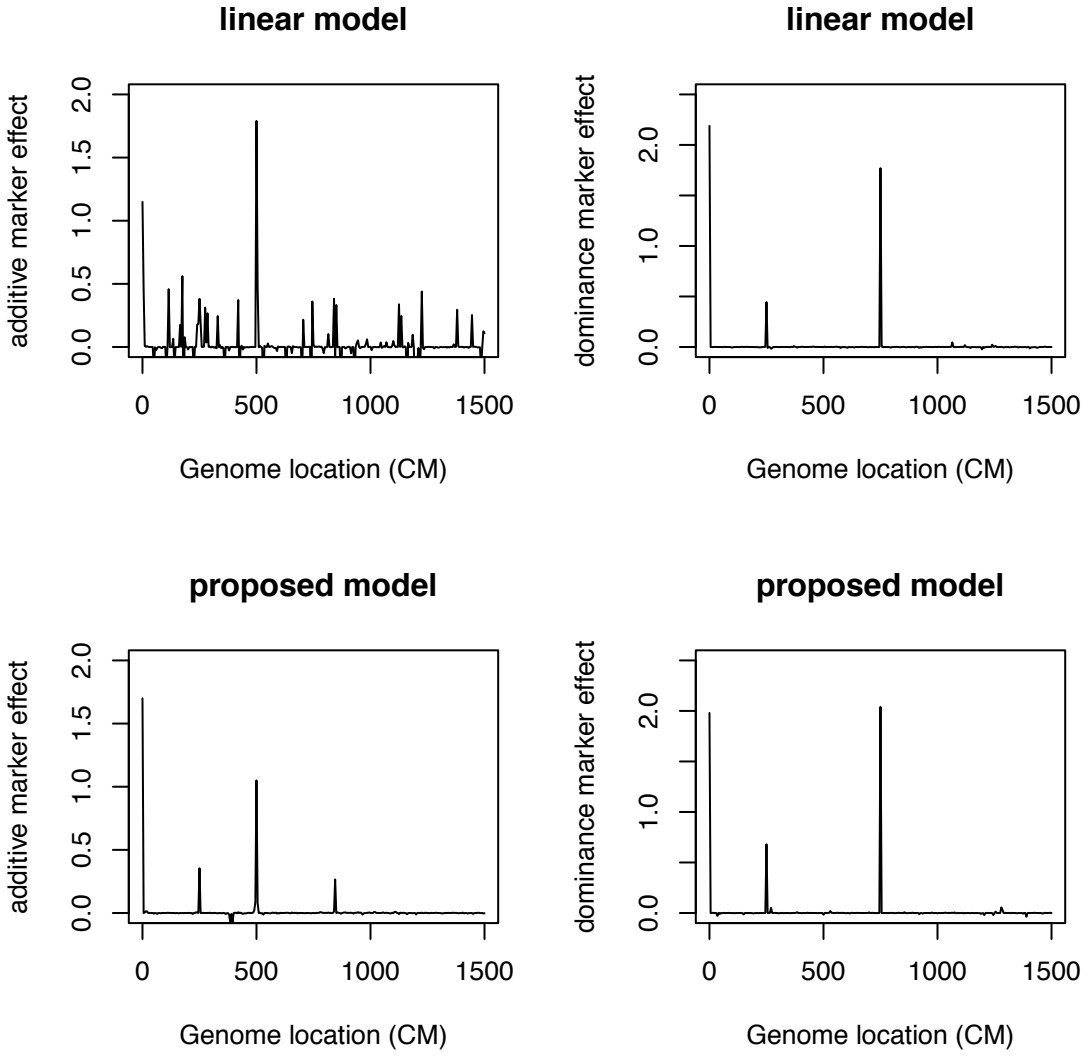

FIGURE S2.—Bayesian QTL estimates from data simulated under Case 2 where polygenic effect exists ($\sigma_\alpha^2 = 8$). The parameter $\delta = 0$. See Figure 1 for the figure legends.