Bayesian LASSO for Quantitative Trait Loci Mapping

Nengjun Yi*,[1] and Shizhong Xu[†]

*Department of Biostatistics, Section on Statistical Genetics, University of Alabama, Birmingham, Alabama 35294
and [†]Department of Botany and Plant Sciences, University of California, Riverside, California 92521

ABSTRACT

The mapping of quantitative trait loci (QTL) is to identify molecular markers or genomic loci that influence the variation of complex traits. The problem is complicated by the facts that QTL data usually contain a large number of markers across the entire genome and most of them have little or no effect on the phenotype. In this article, we propose several Bayesian hierarchical models for mapping multiple QTL that simultaneously fit and estimate all possible genetic effects associated with all markers. The proposed models use prior distributions for the genetic effects that are scale mixtures of normal distributions with mean zero and variances distributed to give each effect a high probability of being near zero. We consider two types of priors for the variances, exponential and scaled inverse-χ^2 distributions, which result in a Bayesian version of the popular least absolute shrinkage and selection operator (LASSO) model and the well-known Student's t model, respectively. Unlike most applications where fixed values are preset for hyperparameters in the priors, we treat all hyperparameters as unknowns and estimate them along with other parameters. Markov chain Monte Carlo (MCMC) algorithms are developed to simulate the parameters from the posteriors. The methods are illustrated using well-known barley data.

Q UANTITATIVE traits are controlled by multiple quantitative trait loci (QTL). The genetic effects of QTL and the phenotypic value of a quantitative trait are usually described by a linear model. Since the locations of QTL are not known *a priori*, we often use markers to represent QTL. Some markers may be closely linked to one or more QTL, and thus they may show strong association with the trait. Most markers, however, may not be directly linked to QTL, and thus no association will be expected between these markers and the trait. Interval mapping or the equivalent single-marker analysis does not provide accurate estimates of QTL effects because the single-QTL model used by the method is seldom the correct model. The multiple-QTL model is the reasonable choice for mapping quantitative traits (KAO *et al.* 1999).

When all markers are included in the QTL analysis, the model may be oversaturated. Variable selection may be conducted to include and exclude markers in the model (SILLANPÄÄ and ARJAS 1998; YI *et al.* 2003, 2005; YI 2004; YI and SHRINER 2008). An alternative approach is a shrinkage method that includes all variables in the model and uses informative prior distributions to shrink trivial effects toward zero. Ridge regression (HOERL and KENNARD 1970) is a shrinkage procedure and can be obtained using independent and identical normal

priors centered at zero, with the degree of shrinkage controlled by the prior variance. The least absolute shrinkage and selection operator (LASSO) is another shrinkage method that has been widely used in regression analysis for large models (TIBSHIRANI 1996). It minimizes the residual sum of squares constraining the sum of absolute values of the regression coefficients, $t \geq \sum |\beta_j|$ for $t \geq 0$, if the response and the predictors are standardized. This special constraint allows some estimated regression coefficients to be exactly zero. This selective or individualized shrinkage allows LASSO to handle extremely large models. Mathematically, the LASSO estimates of regression coefficients can be achieved by

$$\min_{\beta,\lambda} \left[\left(y - \sum X_j \beta_j \right)^T \left(y - \sum X_j \beta_j \right) + \lambda \sum |\beta_j| \right],$$

where $\lambda \geq 0$ is a Lagrange multiplier, which relates implicitly to the bound t and controls the degree of shrinkage. The LASSO estimates of coefficients can be efficiently computed via the LARS algorithm of EFRON *et al.* (2004). The LASSO procedure can be interpreted as a Bayesian posterior mode estimate when assigning an independent double-exponential prior to each β_j (*e.g.*, TIBSHIRANI 1996; YUAN and LIN 2005; PARK and CASELLA 2007). Recently, XU (2007) successfully applied the LASSO method to estimate the epistatic effects of QTL. However, the LASSO provides no estimate for the residual error variance and no interval estimate for a

[1] *Corresponding author:* Department of Biostatistics, University of Alabama, Birmingham, AL 35294-0022.   E-mail: nyi@ms.soph.uab.edu

regression coefficient and needs to predetermine the parameter λ (TIBSHIRANI 1996; EFRON *et al.* 2004).

There are some other shrinkage methods that have been applied to mapping multiple QTL. The method of XU (2003) simultaneously fits maker effects of the entire genome in a regression model and assigns each effect a normal prior with mean 0 and an effect-specific variance (also see WANG *et al.* 2005; HOTI and SILLANPÄÄ 2006; HUANG *et al.* 2007). The variance parameters are further assigned the noninformative Jeffrey's prior, which in turn induces improper priors on the coefficients. MEUWISSEN *et al.* (2001) and XU (2007) used scaled inverse-χ^2 distributions with predetermined hyperparameters as priors for the effect-specific variance parameters. Rather than presetting the hyperparameters in the priors that control the degree of shrinkage, it would be appealing to treat the hyperparameters as unknowns and estimate them from the data so that the model can shrink the coefficients as much as can be justified by the data. PARK and CASELLA (2007) set up a hyperprior for the hyperparameter in the LASSO model and developed a Gibbs sampler to fit the hierarchical model. R. B. O'HARA and M. J. SILLANPÄÄ (unpublished results) conducted a comparison with several Bayesian model selection and shrinkage methods, including Bayesian LASSO and the method of XU (2003).

In this article, we propose several Bayesian hierarchical models for mapping multiple QTL that simultaneously fit and estimate all possible genetic effects associated with all markers across the entire genome. We use prior distributions for the genetic effects that are scale mixtures of normal distributions with mean zero and unknown effect-specific variances. We consider two types of priors for the variances, exponential and scaled inverse-χ^2 distributions, which result in a Bayesian version of the LASSO model and the well-known Student's t model, respectively. We treat all hyperparameters as unknowns and estimate them along with other parameters. We fit the models in a fully Bayesian approach, employing the Markov chain Monte Carlo (MCMC) simulation to generate posterior samples from the joint posterior distribution. Our methods give not only point estimates but also interval estimates of all parameters and provide natural means of assessing model uncertainty.

MULTIPLE-QTL MODEL

We describe our method primarily for a mapping population with only two segregating genotypes, *e.g.*, a backcross, double-haploid (DH), or recombinant inbred line (RIL). The method can be extended to other types of population. Assume that we observe many markers across the genome. The aim of QTL mapping is to identify which markers are tightly linked to genes with detectable effects and to estimate the magnitudes of the effects. For a continuously distributed trait, the observed phenotypic value y_i of individual i can be described by the linear regression model

$$y_i = \mu + \sum_{j=1}^{p} x_{ij}\beta_j + e_i \triangleq \mu + X_i\beta + e_i, \quad i = 1, 2, \cdots, n,$$

(1)

where p is the number of markers, μ is the overall mean, x_{ij} denotes the genotype of marker j for individual i and is defined as -0.5 and 0.5 for the two genotypes in the mapping population, the coefficient β_j represents the main effect of marker j, e_i is the residual error assumed to follow a $N(0, \sigma^2)$ distribution, $X_i = (x_{i1}, \cdots, x_{ip})$, and $\beta = (\beta_1, \cdots, \beta_p)^T$.

In practice, some marker genotypes may be missing. There are two commonly used methods to deal with missing marker data. The first approach takes the uncertainty in missing genotypes into account by treating the missing genotypes as unknowns and sampling them in the MCMC update procedure. The second approach is similar to the regression model of HALEY and KNOTT (1992) and replaces all missing genotypes by their expected values conditioning on the observed marker data, using the multipoint method (JIANG and ZENG 1997). Although the second method ignores the uncertainty of the estimated marker genotypes, it has a big computational advantage over the first method. Both approaches can be incorporated into our Bayesian method. For simplicity and computational convenience, this study considers only the second method.

PRIOR DISTRIBUTIONS

Data are usually sufficient to estimate the overall mean μ and the residual variance σ^2. Thus, we can use any reasonable noninformative prior distributions for these parameters. For example, we can use an independent, flat prior for μ, *i.e.*, $\pi(\mu) \sim 1$, and a noninformative scale-invariant prior $1/\sigma^2$ for σ^2.

QTL data typically contain a large number of markers, leading to many coefficients in model (1). However, most of the coefficients are expected to have no or only weak effects on the phenotype. Only a few coefficients may have notable effects (see, *e.g.*, XU 2003). To incorporate this prior knowledge into our analysis, therefore, we set up a prior distribution that gives each coefficient β_j a high probability of being near zero and in the meantime gives each coefficient a chance to take a large effect.

A commonly used prior that has the above properties is the double exponential (also called Laplace) distribution

$$\pi(\beta) = \prod_{j=1}^{p} \frac{\lambda}{2} e^{-\lambda |\beta_j|}$$

(2)

(TIBSHIRANI 1996; PARK and CASELLA 2007), where λ is a hyperparameter. With the double-exponential prior,

the posterior mode estimate of the coefficients β is the LASSO estimate (TIBSHIRANI 1996; PARK and CASELLA 2007). Another wide-tailed distribution is the well-known Student's t distribution,

$$\pi(\beta) = \prod_{j=1}^{p} t_{\nu}(\beta_j \,|\, \alpha, \, s^2), \qquad (3)$$

where the hyperparameters ν, α, and s^2 are the degrees of freedom, the location, and the scale parameters, respectively (see, *e.g.*, SORENSEN and GIANOLA 2002; GELMAN *et al.* 2003; VARONA *et al.* 2005).

Both the double-exponential distribution and Student's t distribution can be presented as a two-level hierarchical model (see, *e.g.*, ANDREWS and MALLOWS 1974; GRIFFIN and BROWN 2006). Compared with their original forms, the two-level hierarchical models are easier to deal with both analytically and computationally. In the two-level hierarchical models, the first level assumes that the coefficients β_j follow independent normal distributions with mean zero and unknown variances τ_j^2,

$$\beta \,|\, \tau^2 \sim \prod_{j=1}^{p} N(\beta_j \,|\, 0, \, \tau_j^2), \qquad (4)$$

where $\tau^2 = (\tau_1^2, \cdots, \tau_p^2)$. At the second level, we assume that the variances τ_j^2 follow some specified independent prior distributions

$$\tau^2 \,|\, \theta \sim \prod_{j=1}^{p} \pi(\tau_j^2 \,|\, \theta), \qquad (5)$$

where θ includes all the hyperparameters. The above two-level priors result in a scale mixture of normal distributions for the coefficients β_j,

$$\beta \sim \pi(\beta \,|\, \theta) = \prod_{j=1}^{p} \int_0^{\infty} N(\beta_j \,|\, 0, \, \tau_j^2) \pi(\tau_j^2 \,|\, \theta) \, d\tau_j^2, \qquad (6)$$

where $\pi(\tau_j^2 \,|\, \theta)$ is referred to as the mixing distribution. From this, we see that two-level hierarchical models introduce different variance parameters for different coefficients and hence induce different amounts of shrinkage in the estimates of different coefficients. The variance components are not the parameters of interest, but they are useful intermediate quantities to facilitate better inferences for the individual regression coefficients (*e.g.*, GELMAN 2005).

The double-exponential distribution (2) can be presented as a mixture of normal distributions (4) with the mixing distribution $\pi(\tau_j^2 \,|\, \theta)$ being an exponential distribution $\mathrm{Expon}(\lambda^2/2)$ or equivalently a gamma distribution $\mathrm{Gamma}(1, \lambda^2/2)$,

$$\pi\!\left(\tau_j^2 \,|\, \lambda\right) = \mathrm{Expon}\!\left(\tau_j^2 \,|\, \frac{\lambda^2}{2}\right) = \frac{\lambda^2}{2} e^{-\lambda^2 \tau_j^2/2}, \qquad (7)$$

where $\lambda^2/2$ is the inverse scale that needs to be carefully preset or estimated from the data.

For any values of λ, the mode of the exponential distribution $\mathrm{Expon}(\lambda^2/2) = 0$, meaning that the "most likely" values of τ_j^2 and thus β_j are zero. However, the variance of τ_j^2 is $4/\lambda^4$, strongly depending on the value of λ. Instead of presetting a value for λ, it is appealing to assign a prior distribution to λ also so that λ can be estimated along with other parameters. This procedure obviates the choice of λ and automatically accounts for the uncertainty in its selection that affects the estimates of regression coefficients (PARK and CASELLA 2007). For convenience, we regard $\lambda^2/2$, instead of λ, as the parameter of interest. Following PARK and CASELLA (2007), we assign a conjugate gamma prior $\mathrm{Gamma}(a, \, b)$, $a > 0$, $b > 0$, to the parameter $\lambda^2/2$. Assigning a prior distribution to $\lambda^2/2$ provides a way to estimate $\lambda^2/2$. We set a and b as small values (*e.g.*, $a = 0.1$ and $b = 0.1$) so that the prior for $\lambda^2/2$ is essentially noninformative.

Student's t distribution (3) can be expressed as a mixture of normal distributions (4) with the mixing distribution $\pi(\tau_j^2 \,|\, \theta)$ being a conjugate scaled inverse-χ^2 distribution $\mathrm{Inv}\text{-}\chi^2(\nu, \, s^2)$ or equivalently an inverse gamma distribution $\mathrm{Inv}\text{-}\mathrm{gamma}(\nu/2, (\nu/2)s^2)$,

$$\pi(\tau_j^2 \,|\, \nu, \, s^2)$$
$$= \mathrm{Inv}\text{-}\chi^2(\tau_j^2 \,|\, \nu, \, s^2) \propto (\tau_j^2)^{-(\nu/2+1)} \exp\!\left(-\nu s^2 / \left(2\tau_j^2\right)\right), \qquad (8)$$

where the hyperparameters $\nu > 0$ and $s^2 > 0$ represent the degrees of freedom and the scale of the distribution, respectively.

The amount of shrinkage in the prior (8) depends on the values of the hyperparameters ν and s^2. The improper distribution $\pi(\tau_j^2) \propto \tau_j^{-2}$, which is equivalent to the uniform distribution on $\log \tau_j^2$, has the form of the inverse-χ^2 density with $\nu = 0$ and $s^2 = 0$ and can be taken as a limit of the proper inverse-χ^2 or inverse-gamma densities. This improper prior in turn induces an improper prior for β_j of the form $\pi(\beta_j) \propto |\beta_j|^{-1}$ (GRIFFIN and BROWN 2006). In the class of inverse-χ^2 densities, $\mathrm{Inv}\text{-}\chi^2(0, 0)$ has the constant degree of shrinkage in the coefficient estimates. XU (2003) actually used this improper prior $\pi(\tau_j^2) \propto \tau_j^{-2}$ in his shrinkage QTL mapping.

We here treat ν and s^2 as unknown parameters, obviating the choice of (ν, s^2) and automatically accounting for the uncertainty in its selection that affects the estimates of the regression coefficients. We assign a uniform density on $1/\nu$ for the range $(0, 1]$ and a uniform distribution on s for the range $(0, A]$ with A being a large number (see GELMAN *et al.* 2003).

The prior (4) on the coefficients β_j is independent of the residual variance and has been commonly used in the literature. PARK and CASELLA (2007) recently pro-

posed an alternative version of the prior on β_j that is dependent on the residual variance σ^2,

$$\beta \mid \tau^2 \sim \prod_{j=1}^{p} N\left(\beta_j \mid 0, \sigma^2 \tau_j^2\right), \qquad (9)$$

where $\tau^2 = (\tau_1^2, \cdots, \tau_p^2)$. At the second level, the prior distributions on τ_j^2 take the same form as (7) or (8). This conditional prior may have some advantages; for example, it results in that λ is unitless and posteriors induced from (9) generally do not have more than one mode (PARK and CASELLA 2007).

We now have introduced two different priors for β_j and two different priors for τ_j^2. The two priors for β_j are $\pi_1(\beta_j) = N(\beta_j \mid 0, \tau_j^2)$ and $\pi_2(\beta_j) = N(\beta_j \mid 0, \sigma^2\tau_j^2)$, respectively. The two priors for τ_j^2 are $\pi_1(\tau_j^2) =$ Expon$(\tau_j^2 \mid \lambda^2/2)$ and $\pi_2(\tau_j^2) = $ Inv-$\chi^2(\tau_j^2 \mid \nu, s^2)$, respectively. This yields four possible combinations and thus four different models. These four models are called model I with prior $\pi_1(\beta_j)\pi_1(\tau_j^2)$, model II with prior $\pi_2(\beta_j)\pi_1(\tau_j^2)$, model III with prior $\pi_1(\beta_j)\pi_2(\tau_j^2)$, and model IV with prior $\pi_2(\beta_j)\pi_2(\tau_j^2)$.

POSTERIOR COMPUTATION

We fit the models using the MCMC algorithm, applied to the joint posterior distribution of all the parameters $(\mu, \beta, \sigma^2, \tau^2, \theta)$. The joint posterior distribution can be expressed as

$$\pi(\mu, \beta, \sigma^2, \tau^2, \theta \mid y)$$
$$\propto \prod_{i=1}^{n} \pi(y_i \mid \mu, \beta, \sigma^2)\pi(\mu)\pi(\sigma^2) \prod_{j=1}^{p} \pi\left(\beta_j \mid \tau_j^2\right)\pi\left(\tau_j^2 \mid \theta\right)\pi(\theta),$$
$$(10)$$

where $y = (y_1, \cdots, y_n)^T$, $\beta = (\beta_1, \cdots, \beta_p)^T$, $\tau^2 = (\tau_1^2, \cdots, \tau_p^2)$, θ represents λ for models I and II and (ν, s^2) for models III and IV, $\pi(y_i \mid \mu, \beta, \sigma^2)$ is the normal density function $N(\mu + X_i\beta, \sigma^2)$, and other terms in the right-hand side are the priors defined in the last section. For notational convenience, we suppress the dependence on X here and afterward.

With the two-level hierarchical models as set up in the last section, the joint posterior distribution $\pi(\mu, \beta, \sigma^2, \tau^2, \theta \mid y)$ can be simulated using the MCMC algorithm, alternately updating μ from the normal distribution $\pi(\mu \mid y, \beta, \sigma^2)$, updating each β_j from the normal distribution $\pi(\beta_j \mid y, \mu, \beta_{-j}, \sigma^2, \tau^2)$, where β_{-j} represents all elements of β except β_j, updating σ^2 from the inverse-χ^2 distribution $\pi(\sigma^2 \mid y, \mu, \beta)$, updating each τ_j^2 from the inverse Gaussian (for models I and II) or inverse-χ^2 distribution (for models III and IV), and updating $\lambda^2/2$ from the gamma distribution $\pi(\lambda^2 \mid \tau^2)$ (for models I and II), or updating s^2 from the gamma distribution $\pi(s^2 \mid \tau^2, \nu)$ and updating ν from $\pi(\nu \mid y, \mu, \beta, \sigma^2, \tau^2)$, using a Metropolis step (for models III and IV).

The two-level hierarchical modeling allows us to easily derive the above conditional posterior distributions (see APPENDIX A). For all the four proposed models, however, the above MCMC algorithms can be performed directly using the Bayesian software WinBUGS14 (SPIEGELHALTER *et al.* 2002), and thus we actually do not need to derive the conditional posteriors. However, the expressions of the conditional posteriors allow us to develop faster, purpose-built programs. The WinBUGS code implementing these four hierarchical models can be obtained by request from the first author.

The MCMC algorithm proceeds to draw each unknown from its fully conditional posterior distribution, given the current values of all other unknowns and the observed data. Each iteration of the MCMC algorithm cycles through all elements of $(\mu, \beta, \sigma^2, \tau^2, \theta)$. This process continues for a large number of iterations to obtain random samples from the joint posterior distribution. We run multiple parallel chains with overdispersed starting points and use the potential scale reduction factor $\hat{R}$ that compares the between- and within-sequence variances for any scalar estimands, to monitor convergence (GELMAN *et al.* 2003). We use the condition of $\hat{R} < 1.1$ for all scalar estimands of interest as the criteria of convergence. Once our simulations have converged, we continue to draw thousands of simulations and treat them as a sample from the joint posterior distribution. We thin the sequences by keeping every kth simulation draw from each sequence and discarding the rest (*e.g.*, $k = 100$).

POSTERIOR SUMMARY AND INTERPRETATION

The proposed Bayesian hierarchical models can provide various graphical and tabular summaries that assess the contribution of individual loci while adjusting for effects of all other possible loci. One of such summaries is the estimate of the genetic effects β_j associated with each marker. Furthermore, using the posterior samples of β_j and the observed values x_{ij}, we can also calculate the proportion of the phenotypic variance explained by each effect (*i.e.*, heritability), $h_j^2 = V_j\beta_j^2/V_y$, where V_y is the phenotypic variance and V_j is the sample variance of $\{x_{ij}; i = 1, \cdots, n\}$. The heritability h_j^2 is independent of the scale of the phenotype, while the effect size β_j depends on it. The fully Bayesian analysis usually uses the posterior median or mean to estimate β_j and h_j^2 and also can provide the posterior interval for the estimate.

The Bayesian hierarchical models shrink the "insignificant" coefficients to zero, and hence the posterior estimates of β_j and h_j^2 can guide variable selection. However, these posterior summaries do not automatically provide a criterion to declare that an effect is "in" the model. HOTI and SILLANPÄÄ (2006) suggest setting a threshold value, c, such that the standardized effect β_j is included if $|\beta_j| > c$. Note that the summary statistic, heritability, is essentially the same as the standardized
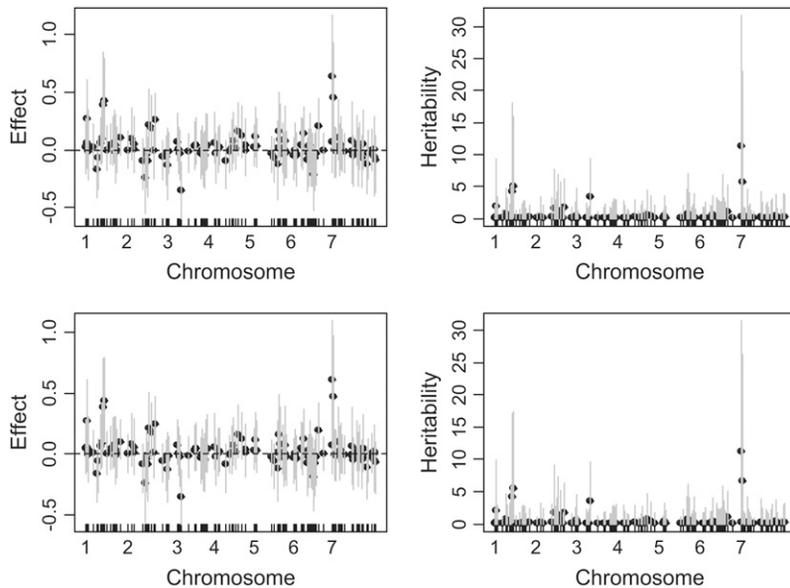
FIGURE 1.—Posterior medians (points) and 95% intervals (shaded lines) for genetic effects and the proportion of the phenotypic variance explained by each effect (*i.e.*, heritability). (Top) Plots drawn from model I; (bottom) plots for model II. Inner tick marks on the *x*-axis represent the marker positions.

effects in HOTI and SILLANPÄÄ (2006). YANG and XU (2007) proposed using a Wald test statistic.

We here construct an alternative measurement that is similar to the widely used LOD score in the traditional QTL mapping approach if the phenotype is affected by a single QTL (LANDER and BOTSTEIN 1989), following the idea of YANDELL *et al.* (2007) (also see the vignette in the package R/qtlbim). We call this statistic the approximate Bayesian LOD score. The approximate Bayesian LOD score can be conveniently compared with the LOD curve obtained from the traditional interval-mapping approach. Our Bayesian LOD score considers the contribution of a given locus to the LOD after adjusting effects for all other possible loci. For marker *j*, the Bayesian LOD score (BLOD) is calculated as

$$\text{BLOD}_j = 2\log_{10}\left(\frac{\prod_{i=1}^{n} N(y_i \mid \mu + X_i\beta, \sigma^2)}{\prod_{i=1}^{n} N\left(y_i \mid \mu + X_i\beta - x_{ij}\beta_j, \sigma_j^2\right)}\right),$$

(11)

where the values of μ, β, and σ^2 are the simulated draws in a certain iteration, and $\sigma_j^2 = (1/(n-2))\sum_{i=1}^{n}(y_i - \mu - X_i\beta + x_{ij}\beta_j)^2$, the mean of the conditional posterior $\pi(\sigma^2 \mid y, \mu, \beta_{-j}) = \text{Inv-}\chi^2\left(n, (1/n)\sum_{i=1}^{n}(y_i - \mu - X_i\beta + x_{ij}\beta_j)^2\right)$. Note that although we adjust for the residual error variance, we use the simulated draws under the full model (including $x_{ij}\beta_j$) to calculate the likelihood function for the reduced model (without $x_{ij}\beta_j$).

REAL DATA ANALYSIS

The well-known barley data set from the North American Barley Genome Mapping Project (TINKER *et al.* 1996) was analyzed using the four proposed Bayesian hierarchical models. This data set has been extensively analyzed

using different methods (TINKER *et al.* 1996; XU 2003, 2007; YI *et al.* 2003, 2005; ZHANG *et al.* 2005; XU and JIA 2007). The data were collected from a doubled-haploid population that contained $n = 145$ lines; each was grown in 25 different environments. The phenotype analyzed was the average value of kernel weight across environments. A total of $p = 127$ mapped markers covering a genome of 1500 cM (seven chromosomes) were used in the analysis. Observed marker genotypes were coded as -0.5 for one genotype and $+0.5$ for the other genotype. The data contain $\sim$5% missing marker genotypes. The missing marker genotypes were replaced by their expected values conditioning on the observed marker data using the multipoint method. We simultaneously modeled all 127 marker effects in the linear regression (1).

For all four proposed models, the MCMC simulations were performed in the Bayesian software Win-BUGS14 (SPIEGELHALTER *et al.* 2002), linked from the R statistical computing software (STURTZ *et al.* 2005; R DEVELOPMENT CORE TEAM 2006), which we can use to conveniently plot the results. For each analysis, we ran three parallel simulation sequences with starting points randomly generated from the prior distributions. The models with the dependent prior (9) (models III and IV) were found to converge slightly quicker than those with the independent prior (4) (models I and II). For all four models, the potential scale reduction factor $\hat{R}$ values fell below 1.1 for all parameters after $\sim$5000 iterations, indicating that the chains converged adequately (data not shown). For each of the three parallel sequences, therefore, we ran a total of 20,000 iterations and discarded the first halves to allow adequate convergence. The remaining 10,000 posterior samples of each sequence were thinned to 1000 for estimating the posterior distribution of quantities of interest by keeping every 10th simulation draw from each sequence and
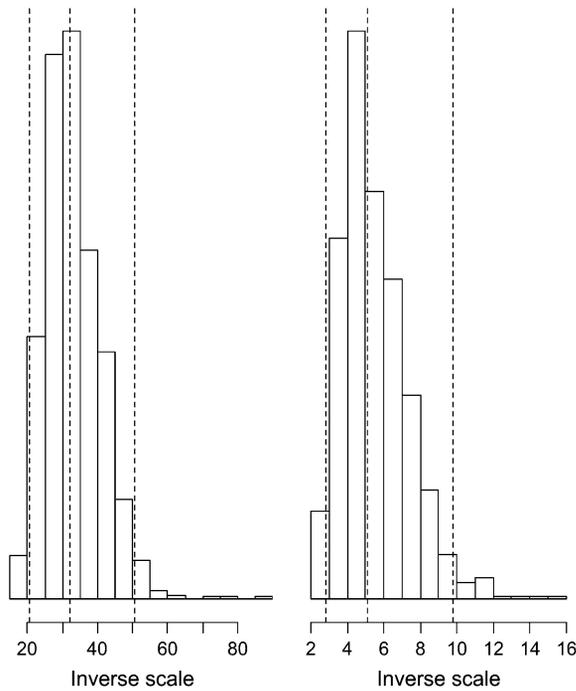
FIGURE 2.—Histogram of the posterior samples for the inverse scale of the exponential prior on the variances. The dotted lines represent the posterior 5, 50, and 95% quantiles. The left and right plots show inferences for models I and II, respectively.

discarding the rest. On a dual processor 2.4-GHz machine, each analysis took ~4.5 hr.

To investigate whether or not posterior inferences are affected by the values of the shape a and the inverse scale b in the gamma prior Gamma(a, b), we analyzed models I and II with four different values of (a, b), $a = b = 0.01, 0.1, 0.5, 1$. Under the four analyses, the posterior

estimates of the parameters (μ, β, σ^2, τ^2) were identical, but the inverse scale $\lambda^2/2$ differed slightly.

Figure 1 shows the posterior medians and the 95% posterior intervals for the marker effects β_j and the proportion of the phenotypic variance explained by each effect (*i.e.*, heritability) h_j^2. Models I and II produced almost identical results, both being able to identify markers with large effects. The analyses showed four chromosomes (7, 1, 3, and 2) with strong evidence of QTL. The QTL on chromosome 3 was found to have an opposite effect on the phenotype to the other detected QTL and was missed in the analysis of XU (2003), who used the improper prior $\pi(\tau_j^2) \propto \tau_j^{-2}$. The difference may also be caused by the different ways of treating missing marker genotypes for the two methods. As can be seen in Figure 1, the plots of the heritability can provide a clearer picture than those of the effect.

Figure 2 shows the histograms of the inverse scale $\lambda^2/2$ for the analyses of models I and II with $a = b = 0.1$. We can see that the hyperparameter $\lambda^2/2$ can be estimated with quite high precision. For model I, the inverse scale parameter was estimated at 32.0, with a 95% posterior interval of [20.6, 50.5]. However, model II provided a much smaller estimate, with a posterior median of 5.1 and a 95% posterior interval of [2.8, 9.8]. We know that in the exponential distribution a larger inverse scale forces stronger shrinkage. Figure 1 shows that models I and II perform at the same level of shrinkage in the coefficient estimates. This can also be shown by looking at the estimated σ^2, which was 0.3, <1.

Figure 3 depicts the posterior medians and the 95% posterior intervals for marker effect β_j and the heritability h_j^2 from the analyses of models III and IV. Models III and IV produced identical results, similar to those from models I and II. However, the Inv-χ^2 prior distributions on τ_j^2 produced slightly larger posterior me-
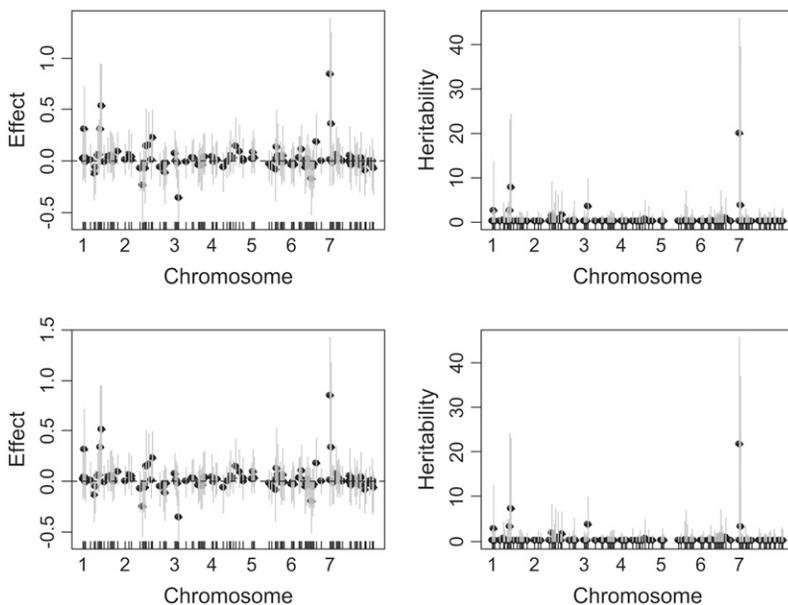
FIGURE 3.—Posterior medians (points) and 95% intervals (shaded lines) for genetic effects and the proportion of the phenotypic variance explained by each effect (*i.e.*, heritability). (Top) Plots drawn from model III; (bottom) plots for model IV. Inner tick marks on the *x*-axis represent the marker positions.
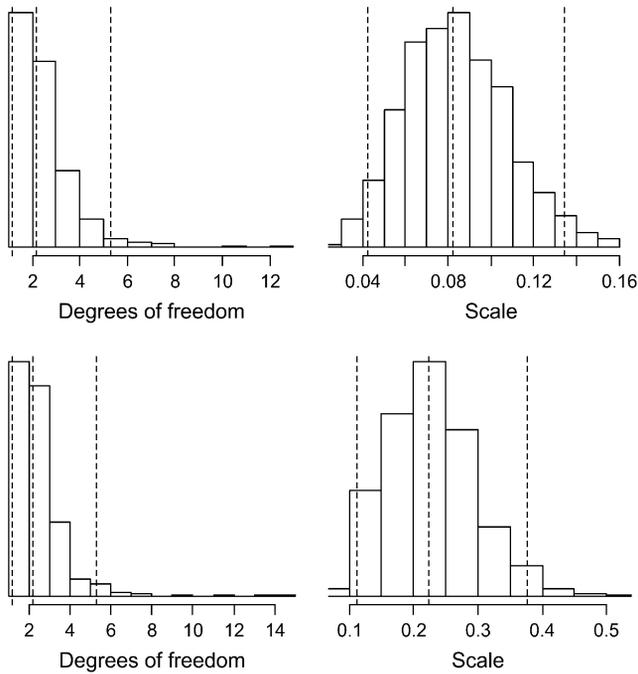
FIGURE 4.—Histogram of the posterior samples for the degrees of freedom and scale of the Inv-χ^2 prior on variances. The dotted lines represent the posterior 5, 50, and 95% quantiles. The top and bottom plots show inferences for models III and IV, respectively.

dians and wider posterior intervals for β_j and h_j^2. The degrees of freedom and scale parameters can be estimated from the data. Figure 4 shows the histograms of these two hyperparameters for the analyses of models III and IV. The two models produced similar estimates for the degrees of freedom, with the posterior median at 2.1

and the 95% posterior interval of [1.2, 5.3]. However, the estimate of the scale parameter for model IV was three times as large as that for model III.

Figure 5 (top) shows the posterior medians and the 95% posterior intervals for the Bayesian LOD score for model I. The other three models produced similar estimates (data not shown). For comparison, Figure 5 (bottom) also shows the LOD score curve calculated by the traditional interval mapping based on a single-QTL model. If we use the standard threshold value of 3.2, the interval mapping identified only two regions (on chromosomes 7 and 1, respectively), which have larger effects in our analyses. Our Bayesian models produced much higher estimates of LOD score and detected more significant QTL.

In the proposed two-level hierarchical models, the amount of shrinkage in the coefficient estimates is largely determined by the hyperparameters in the priors. It is expected that optimal values of the hyperparameters depend on the data and the number of variables included, which makes it difficult to preset their values. To illustrate this, we analyzed the data with only markers on chromosomes 1, 3, and 7 included in the model. Figure 6 displays the plots of marker effects, showing that we were still able to detect the same QTL as in our previous analyses. As shown in Figure 7, however, the posterior estimates of the hyperparameters were different from those in the previous analyses.

DISCUSSION

We have presented four Bayesian hierarchical models that can be used to simultaneously fit and estimate all
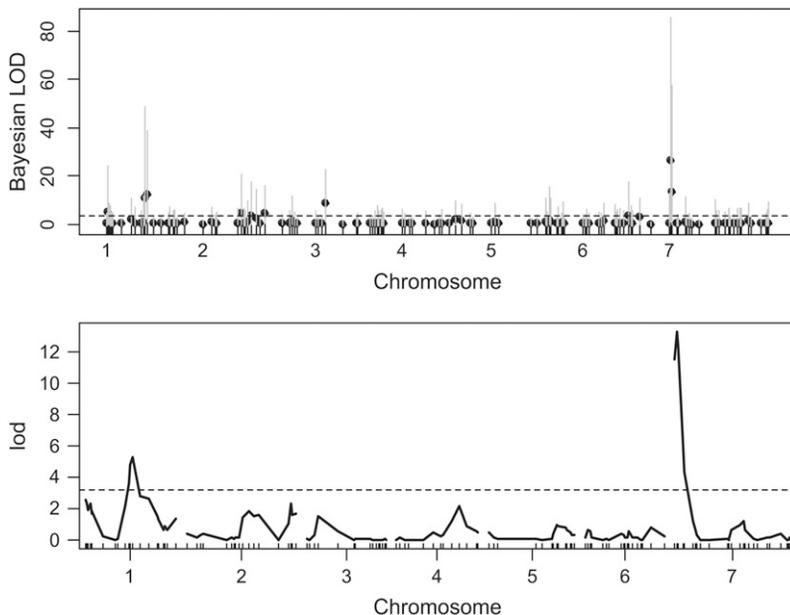
FIGURE 5.—The top plot shows posterior medians (points) and 95% intervals (shaded lines) for the approximate Bayesian LOD score from model I, and the bottom plot shows the LOD score curve from traditional interval mapping. The dotted lines represent the standard threshold value of 3.2. Inner tick marks on the x-axis represent the marker positions.
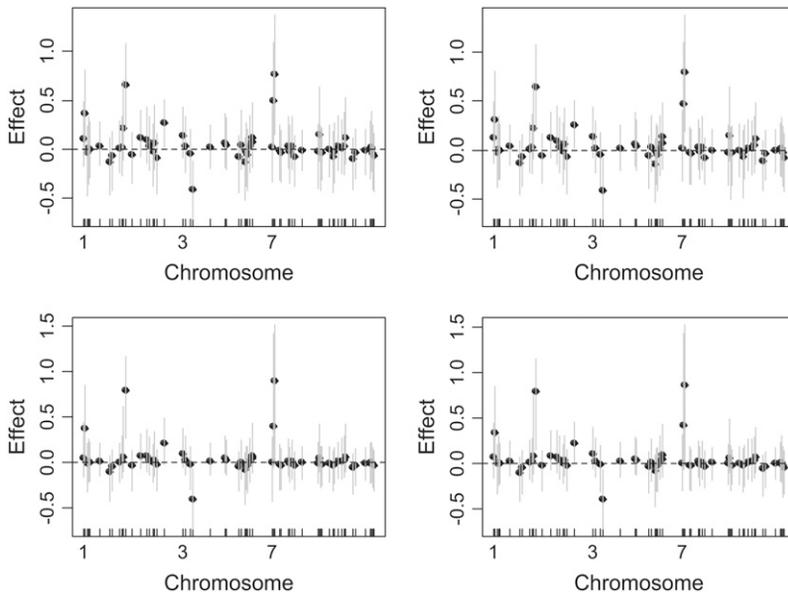
FIGURE 6.—Posterior medians (points) and 95% intervals (shaded lines) for genetic effects from all four models using markers on chromosomes 1, 3, and 7. Inner tick marks on the x-axis represent the marker positions.

possible genetic effects associated with all molecular markers across the entire genome. We found that these four methods perform equally well. It is known *a priori* that most markers have no or little effect on the phenotype, but if a marker has an effect, it could be large. To incorporate this prior knowledge into our analysis, we set up a continuous form of prior distribution for genetic effects that gives each effect a high probability of being near zero and also enables us to accommodate large effects. We have considered two such priors—double exponential distribution and Student's *t* distribution—

that can be expressed as scale mixtures of normal distributions with mean zero and unknown variances distributed as exponential and scaled inverse-χ^2 distributions, respectively. The former prior results in a Bayesian LASSO model (TIBSHIRANI 1996; YUAN and LIN 2005; PARK and CASELLA 2007), and the latter includes Jeffrey's noninformative prior as a special case (GELMAN 2006).

Both the exponential and the scaled inverse-χ^2 distributions include hyperparameters that determine the degrees of shrinkage in the estimates of the variances
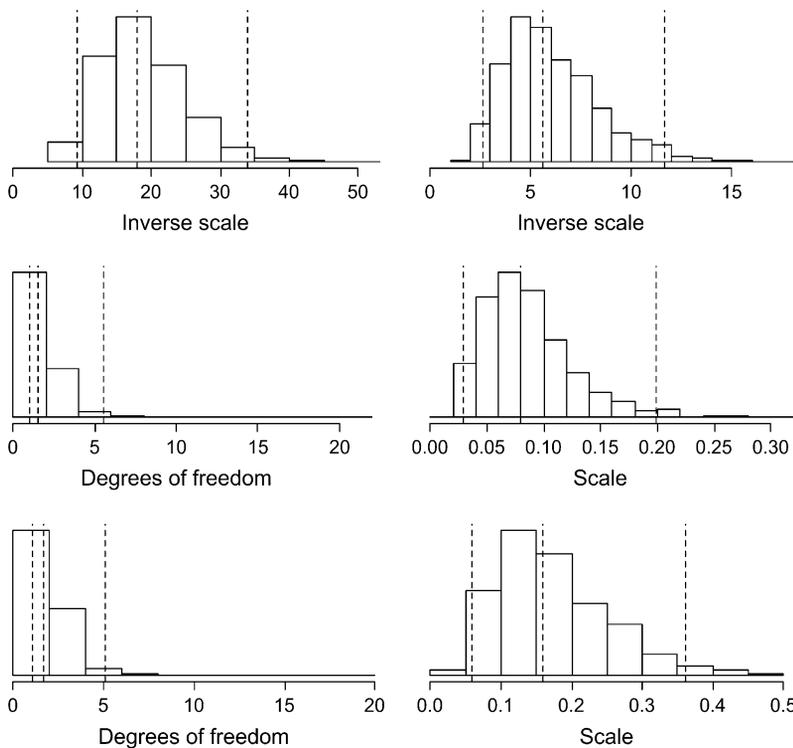
FIGURE 7.—Histogram of the posterior samples for the hyperparameters of priors on variances. The dotted lines represent the posterior 5, 50, and 95% quantiles. The top plots show inferences for models I and II, respectively, the middle plots show inferences from model III, and the bottom plots show inferences from model IV.

and thus the genetic effects. A common practice is to preset values for the hyperparameters using methods like cross validation (TIBSHIRANI 1996; MEUWISSEN *et al.* 2001; XU 2003; BAO and MALLICK 2004; XU 2007). However, optimal values depend on the data and thus may be difficult to choose. A novel aspect of our approach is to treat all hyperparameters as unknowns and estimate them along with other parameters. This procedure allows us to control the amount of shrinkage by taking advantage of the characteristics of the data. In contrast, Jeffrey's prior includes no hyperparameter but always forces a constant degree of shrinkage. XU (2003) showed that Jeffrey's prior yields good performance, but we observed that it converges more slowly than our proposed methods.

We fit the models in a fully Bayesian approach, employing the MCMC simulation to generate posterior samples from the joint posterior distribution, which can be used to make various posterior inferences. Although computationally intensive, it is easy to implement and provides not only point estimates but also interval estimates of all parameters. We have proposed graphical tools to summarize the posterior samples, notably the approximate Bayesian LOD score that can be easily compared with the results from traditional interval mapping. However, we realize that these tools still are informal and descriptive, due to the lack of formal threshold value to select markers. A formal choice of threshold value will be a topic of future research.

The fully Bayesian approach enables us to obtain much richer inferences about the models than most non-Bayesian analyses. In practice, however, it is desirable to have a quicker calculation that merely looks for posterior modes rather than fully investigating the posterior distribution. The original LASSO estimates for linear regression coefficients are equivalent to Bayesian posterior mode estimates when the coefficients have independent and identical double-exponential priors (TIBSHIRANI 1996). The two-level hierarchical formulation enables us to find posterior modes using various algorithms, *e.g.*, conditional maximization (ZHANG and XU 2005), the EM algorithm, and its extensions (FIGUEIREDO 2003; FOSTER *et al.* 2007; XU 2007). However, these methods have been developed on the basis of preset hyperparameters. Further investigation is necessary to extend the mode-searching algorithms to the proposed hierarchical models.

The proposed models have been used to fit all markers and estimate their effects. The methods can be easily extended to detect QTL within marker intervals. The idea is to insert loci within each marker interval as possible positions of QTL and put a reasonable number of loci on each chromosome with positions treated as parameters (WANG *et al.* 2005). An additional Metropolis step is then used to update the positions. Future research also will consider epistatic effects and extensions to complicated experimental crosses. Com-

putationally efficient algorithms are an essential feature for the practical analysis of these more complicated cases. We are in the process of developing purpose-specific computer programs that would significantly reduce the computing time. Our hierarchical models assign marker effects independent priors with a single prior mean. However, it is desirable to develop hierarchical priors to allow shrinkage of coefficients toward multiple prior means with the locations of these means unknown and to accommodate the spatial covariance structure along markers (GIANOLA *et al.* 2003).

LITERATURE CITED

ANDREWS, D. F., and C. L. MALLOWS, 1974 Scale mixtures of normal distributions. J. R. Stat. Soc. Ser. B **36:** 99–102.

BAO, K., and B. K. MALLICK, 2004 Gene selection using a two-level hierarchical Bayesian model. Bioinformatics **20:** 3423–3430.

CHHIKARA, R. S., and L. FOLKS, 1989 *The Inverse Gaussian Distribution: Theory, Methodology, and Applications.* Marcel Dekker, New York.

EFRON, B., T. HASTIE, I. JOHNSTONE and R. TIBSHIRANI, 2004 Least angle regression. Ann. Stat. **32:** 407–499.

FIGUEIREDO, M. A. T., 2003 Adaptive sparseness for supervised learning. IEEE Trans. Patt. Anal. Machine Intell. **25:** 1150–1159.

FOSTER, S. D., A. P. VERBYLA and W. S. PITCHFORD, 2007 Incorporating LASSO effects into a mixed model for QTL detection. J. Agric. Biol. Environ. Stat. **12**(2): 300–314.

GELMAN, A., 2005 Analysis of variance: why it is more important than ever (with discussion). Ann. Stat. **33:** 1–53.

GELMAN, A., 2006 Prior distributions for variance parameters in hierarchical models. Bayesian Anal. **1:** 515–533.

GELMAN, A., J. CARLIN, H. STERN and D. RUBIN, 2003 *Bayesian Data Analysis.* Chapman & Hall, London.

GIANOLA, D., M. PEREZ-ENCISO and M. A. TORO, 2003 On marker-assisted prediction of genetic value: beyond of the ridge. Genetics **163:** 347–365.

GRIFFIN, J. E., and P. J. BROWN, 2006 Alternative prior distributions for variable selection with very many more variables than observations. Technical report. University of Warwick, Coventry, UK.

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:** 315–324.

HOERL, A. E., and R. W. KENNARD, 1970 Ridge regression: biased estimation for nonorthogonal problems. Technometrics **12:** 55–67.

HOTI, F., and M. J. SILLANPÄÄ, 2006 Bayesian mapping of genotype × expression interactions in quantitative and qualitative traits. Heredity **97:** 4–18.

HUANG, H., C. D. EVERSLEY, D. W. THREADGILL and F. ZOU, 2007 Bayesian multiple quantitative trait loci mapping for complex traits using markers of the entire genome. Genetics **176:** 2529–2540.

JIANG, C., and Z.-B. ZENG, 1997 Mapping quantitative trait loci with dominant and missing markers in various crosses from two inbred lines. Genetica **101:** 47–58.

KAO, C. H., Z.-B. ZENG and R. D. TEASDALE, 1999 Multiple interval mapping for quantitative trait loci. Genetics **152:** 1203–1216.

LANDER, E. S., and D. BOTSTEIN, 1989 Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics **121:** 185–199.

MEUWISSEN, T. H. E., B. J. HAYES and M. E. GODDARD, 2001 Prediction of total genetic value using genome-wide dense marker maps. Genetics **157:** 1819–1829.

PARK, T., and G. CASELLA, 2007 The Bayesian Lasso. Technical report. University of Florida, Gainesville, FL.

R Development Core Team, 2006 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna. http://www.R-project.org.

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

Sorensen, D., and D. Gianola, 2002 *Likelihood, Bayesian and MCMC Methods in Quantitative Genetics*. Springer, New York.

Spiegelhalter, D., A. Thomas, N. Best and D. Lunn, 2002 *BUGS: Bayesian Inference Using Gibbs Sampling*, Version 1.4. MRC Biostatistics Unit, Cambridge, UK. www.mrc-bsu.cam.ac.uk/bugs/.

Sturtz, S., U. Ligges and A. Gelman, 2005 R2WinBUGS: a package for running WinBUGS from R. J. Stat. Software **12:** 1–16.

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

Tinker, N. A., D. E. Mather, B. G. Rossnagel, K. J. Kasha and A. Kleinhofs, 1996 Regions of the genome that affect agronomic performance in two-row barley. Crop Sci. **36:** 1053–1062.

Varona, L., W. Mekkawy, D. Gianola and A. Blasco, 2005 A whole genome analysis using robust asymmetric distributions. Genet. Res. **88:** 143–151.

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

Xu, S., 2007 An empirical Bayes method for estimating epistatic effects of quantitative trait loci. Biometrics **63:** 513–521.

Xu, S., and Z. Jia, 2007 Genome-wide analysis of epistatic effects for quantitative traits in barley. Genetics **175:** 1955–1963.

Yandell, B. S., T. Mehta, S. Banerjee, D. Shriner, R. Venkataraman *et al.*, 2007 R/qtlbim: QTL with Bayesian interval mapping in experimental crosses. Bioinformatics **23:** 641–643.

Yang, R., and S. Xu, 2007 Bayesian shrinkage analysis of quantitative trait loci for dynamic traits. Genetics **176:** 1169–1185.

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

Yi, N., and D. Shriner, 2008 Advances in Bayesian multiple QTL mapping in experimental designs. Heredity **100:** 240–252.

Yi, N., V. George and D. B. Allison, 2003 Stochastic search variable selection for identifying quantitative trait loci. Genetics **164:** 1129–1138.

Yi, N., B. S. Yandell, G. A. Churchill, D. B. Allison, E. J. Eisen *et al.*, 2005 Bayesian model selection for genome-wide epistatic QTL analysis. Genetics **170:** 1333–1344.

Yuan, M., and Y. Lin, 2005 Efficient empirical Bayes variable selection and estimation in linear models. J. Am. Stat. Assoc. **100:** 1215–1225.

Zhang, Y.-M., and S. Xu, 2005 A penalized maximum likelihood method for estimating epistatic effects of QTL. Heredity **95:** 96–104.

Zhang, M., K. L. Montooth, M. T. Wells, A. G. Clark and D. Zhang, 2005 Mapping multiple quantitative trait loci by Bayesian classification. Genetics **169:** 2305–2318.

APPENDIX A: CONDITIONAL POSTERIOR DISTRIBUTIONS

The two-level hierarchical models presented in this study have explicit forms of conditional posterior distributions for parameters and hyperparameters. These conditional posterior distributions are required to perform the MCMC algorithm. The posterior distributions are presented in this section.

Conditional on the parameters (τ^2, θ), the models are the standard weighted linear regressions, and thus the conditional posterior distributions of (μ, β, σ^2) are

$$\mu \mid y, \beta, \sigma^2, \tau^2, \theta \sim N\left(\frac{1}{n}\sum_{i=1}^{n}(y_i - X_i\beta), \frac{1}{n}\sigma^2\right) \tag{A1}$$

$$\sigma^2 \mid y, \mu, \beta, \tau^2, \theta \sim \text{Inv-}\chi^2\left(n, \frac{1}{n}\sum_{i=1}^{n}(y_i - \mu - X_i\beta)^2\right) \tag{A2}$$

$$\beta_j \mid y, \mu, \beta_{-j}, \sigma^2, \tau^2, \theta \sim N\left(\frac{\sum_{i=1}^{n} x_{ij}\left(y_i - \mu - \sum_{k\neq j}^{p} x_{ik}\beta_k\right)}{\sum_{i=1}^{n} x_{ij}^2 + V_j}, \frac{\sigma^2}{\sum_{i=1}^{n} x_{ij}^2 + V_j}\right), \quad j = 1, \cdots, p, \tag{A3}$$

where β_{-j} represents all elements of β except β_j, and $V_j = \sigma^2/\tau_j^2$ for models I and III or $V_j = 1/\tau_j^2$ for models II and IV.

For models I and II, the conditional posterior distribution of τ_j^{-2} is inverse Gaussian,

$$\tau_j^{-2} \mid y, \mu, \beta, \sigma^2, \lambda^2 \sim \text{InvGauss}\left(\sqrt{\frac{\lambda^2\sigma^2}{\beta_j^2}}, \lambda^2\right), \quad j = 1, \cdots, p. \tag{A4}$$

A relatively simple algorithm is available for simulating from the inverse Gaussian distribution (Chhikara and Folks 1989). With the conjugate prior Gamma(a, b), the conditional posterior distribution of the hyperparameter λ^2 in the prior (7) is gamma

$$\lambda^2 \mid y, \mu, \beta, \sigma^2, \tau^2 \sim \text{gamma}\left(p + a, \sum_{j=1}^{p} \tau_j^2/2 + b\right), \quad j = 1, \cdots, p. \tag{A5}$$

For models III and IV, the conditional posterior distribution of τ_j^2 is inverse-χ^2,

$$\tau_j^2 \mid y, \mu, \beta, \sigma^2, \nu, s^2 \sim \text{Inv-}\chi^2\left(\nu + 1, \frac{\nu s^2 + \beta_j^2}{\nu + 1}\right), \quad j = 1, \cdots, p. \tag{A6}$$

The conditional posterior distribution of the scale parameter s^2 in the prior (8) is gamma

$$s^2 \mid y, \mu, \beta, \sigma^2, \nu \sim \text{gamma}\left(\frac{p\nu}{2}, \frac{\nu}{2}\sum_{j=1}^{p}\frac{1}{\tau_j^2}\right). \tag{A7}$$

The conditional posterior of the degrees of freedom ν has no standard form. Therefore, a Metropolis step is required to update ν in each iteration.