A fast and efficient Gibbs sampler for BayesB in whole-genome analyses

Hao Cheng¹², Long Qu³, Dorian J. Garrick¹⁴ and Rohan L. Fernando¹*

Abstract

Background: In whole-genome analyses, the number of marker covariates is often much larger than the number of observations. Bayesian multiple regression models are widely used in genomic selection to address this problem of $p \gg n$. The primary difference between these models is the prior assumed for the effects of the covariates. Usually in the BayesB method, a Metropolis–Hastings (MH) algorithm is used to jointly sample the marker effect and the locus-specific variance, which may make BayesB computationally intensive. In this paper, we show how the Gibbs sampler without the MH algorithm can be used for the BayesB method.

Results: We consider three different versions of the Gibbs sampler to sample the marker effect and locus-specific variance for each locus. Among the Gibbs samplers that were considered, the most efficient sampler is about 2.1 times as efficient as the MH algorithm proposed by Meuwissen et al. and 1.7 times as efficient as that proposed by Habier et al.

Conclusions: The three Gibbs samplers presented here were twice as efficient as Metropolis–Hastings samplers and gave virtually the same results.

Background

In whole-genome analyses, the number $p$ of marker covariates is often much larger than the number $n$ of observations. Bayesian multiple regression models are widely used in genomic selection to address this problem of $p \gg n$. The primary difference between these models is the prior assumed for the effects of the covariates. These priors and their effects on inference have been recently reviewed by Gianola [1]. In most Bayesian analyses of whole-genome data, inferences are based on Markov chains constructed to have a stationary distribution equal to the posterior distribution of the unknown parameters of interest [2]. This is often done by employing a Gibbs sampler where samples are drawn from the full-conditional distributions of the parameters [3].

It can be shown that in BayesA introduced by Meuwissen et al. [4], the prior for each marker effect follows a scaled $t$ distribution [5]. However when the prior for the marker effect is specified as a $t$ distribution, its full-conditional is not of a known form. Fortunately, this prior can also be specified as a normally distributed marker effect conditional on a locus-specific variance, which is given a scaled inverted chi-square distribution. When marginalized over the variance, this gives a $t$ distribution for the marker effect [5]. Thus, the posterior for the marker effect would be identical under both these priors. The second form of the prior, however, is more convenient because it results in the full-conditional for the marker effect having a normal distribution.

BayesA is a special case of BayesB, also introduced by Meuwissen et al. [4], where the prior for each marker effect follows a mixture distribution with a point mass at zero with probability $\pi$ and a univariate-$t$ distribution with probability $1 - \pi$ [5]. When $\pi = 0$, BayesB becomes BayesA. When the marker effect is non-null, as in BayesA, the second form of the prior leads to the full-conditional of the marker effect being normal. Nevertheless, Meuwissen et al. [4] used a Metropolis–Hastings (MH) algorithm to jointly sample the marker effect and the locus-specific variance because they argued that “the Gibbs sampler will not move through the entire sampling space” for BayesB. In their MH algorithm, they use the prior distribution of

*Correspondence: rohan@iastate.edu
¹ Department of Animal Science, Iowa State University, Ames 50011, IA, USA
Full list of author information is available at the end of the article

© 2015 Cheng et al. This article is 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 you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.
the locus-specific variance as the proposal distribution. When $\pi$ is high, the proposed values for the marker effect will be zero with high probability. Thus, for each locus, 100 cycles of MH algorithm were used in their paper, which makes BayesB computationally intensive. Habier et al. [6] used an alternative proposal, where the marker effect was zero with probability 0.5, that leads to a more efficient MH algorithm. For each locus, five cycles of MH were used to sample marker effects in this efficient MH method.

In this paper, we will show how Gibbs samplers without the MH algorithm can be used for the BayesB method. Recall that by introducing a locus-specific variance into BayesA, the full-conditional for the marker effects becomes normal. Similarly, in this paper we show that by introducing a variable $\delta_j$ in BayesB, indicating whether the marker effect for a locus is zero or non-zero, the marker effect and locus-specific variance can be sampled using Gibbs.

We consider three different versions of the Gibbs sampler to sample each marker effect, locus-specific variance and its indicator variable $\delta_j$. The objectives of this paper are to introduce these samplers and study their performance.

Methods

Statistical methods

BayesB introduced by Meuwissen et al. [4] assumes that each locus-specific variance follows a mixture distribution. However, following Gianola [5], we prefer to specify the mixture at the level of the marker effect instead of the locus specific variance. In this formulation, the prior for the marker effect is a mixture with a point mass at zero and a univariate normal distribution conditional on $\sigma_j^2$:

$$
\begin{align*}
(\alpha_j | \sigma_j^2) & \sim \left\{ 
\begin{array}{ll}
0 & \text{with probability } \pi \\
\sim N(0, \sigma_j^2) & \text{with probability } (1 - \pi),
\end{array}
\right.
\end{align*}
$$

where $\sigma_j^2$ follows a scaled inverted chi-square distribution and $\pi$ is treated as known. Employing the concept of data augmentation, it is convenient to write the marker effect $\alpha_j$ as $\alpha_j = \beta_j \delta_j$, where we introduce a Bernoulli variable $\delta_j$ with probability of success $1 - \pi$ and normally distributed variable $\beta_j$ with mean zero and variance $\sigma_j^2$, which has a scaled inverted chi-square distribution. As shown below, Gibbs sampling can be used to draw samples for these unknowns.

**Gibbs samplers for BayesB**

Here, we present three Gibbs samplers for BayesB. The first is a single-site Gibbs sampler, where all parameters are sampled from their full conditional distributions. The second is a joint Gibbs sampler, where $\delta_j, \beta_j$ are sampled from the joint full-conditional distribution $f(\delta_j, \beta_j | y, \mu, \beta_{-j}, \delta_{-j}, \xi, \sigma_j^2)$, where $\xi = [\sigma_2^2, \sigma_3^2, \ldots, \sigma_k^2]$ because $\delta_j$ and $\beta_j$ are highly dependent. Carlin and Chib [7] have shown that the prior used for parameters that are not in the model does not affect the Bayes factor. Thus, this prior, which they call a pseudo prior, can be chosen to improve mixing of the sampler. Following Carlin and Chib, the third sampler is a Gibbs sampler where a pseudo prior is used for $\beta_j$ when $\delta_j$ is zero. Godsill [8] has shown that the marginal posterior for a parameter in the model does not depend on the choice of pseudo priors. It has been suggested to choose the full conditional distribution for $\beta_j$ when it is in the model as the pseudo prior [7, 8]. This choice is justified by showing that using this pseudo prior is equivalent to sampling $\beta_j$ from $f(\delta_j, \beta_j | y, \mu, \beta_{-j}, \delta_{-j}, \xi, \sigma_j^2)$ [8]. However, in BayesB, use of the exact full conditional distribution as the pseudo prior will require MH to sample $\sigma_j^2$ and $\sigma^2_e$. Thus in this paper, a distribution close to the full conditional is used.

**BayesB model with data augmentation model**

$$
y_i = \mu + \sum_{j=1}^{k} X_{ij} \beta_j \delta_j + e_i,
$$

where $y_i$ is the phenotype for individual $i$, $\mu$ is the overall mean, $k$ is the number of SNPs, $X_{ij}$ is the genotype covariate at locus $j$ for animal $i$ (coded as 0, 1, 2), $\beta_j$ is the allele substitution effect for locus $j$, $\delta_j$ is an indicator variable and $e_i$ is the random residual effect for individual $i$.

**Priors**

The prior for $\mu$ is a constant. The prior for $e_i$ is $e_i | \sigma^2_e \sim N(0, \sigma^2_e)$ and $(\sigma^2_e | \nu_e, S^2_e) \sim \nu_e S^2_e \chi^2_{\nu_e}$. The prior for $\beta_j$ is $\beta_j | \sigma^2_j \sim N(0, \sigma^2_j)$. The prior for $(\sigma^2_j | \nu_j, S^2_j) \sim \nu_j S^2_j \chi^2_{\nu_j}$ is zero. Thus, this prior, which they call a pseudo prior, can be used for $\delta_j$.

**Single-site Gibbs sampler**

The full conditional distributions of $\mu, \sigma^2_j$ and $\sigma^2_e$ are well-known [4, 9, 10]. Thus they are presented here without derivations. The full conditional of $\mu$ is a normal distribution with mean $\frac{1}{n} \sum_{i=1}^{n} y_i$ and variance $\sigma^2_e / \nu$, where $e = y - \sum_{j=1}^{k} X_{ij} \beta_j \delta_j$ and $n$ is the number of individuals. The full conditional distributions of $\sigma^2_j$ and $\sigma^2_e$ are both scaled inverted chi-square distributions; for $\sigma^2_j$, the scale parameter is $\frac{(\nu_j S^2_j + \beta_j^2)}{\nu_j + 1}$ and the degrees of freedom parameter are $\nu_j + 1$; for $\sigma^2_e$, the scale parameter is $\frac{\nu_e S^2_e}{\nu_e + n}$ and the degrees of freedom parameter are $\nu_e + n$, where $e = y - \sum_{j=1}^{k} X_{ij} \beta_j \delta_j$. 


Next, we derive the full conditional distributions of $\beta_j$ and $\delta_j$. These full conditional distributions are proportional to the joint distribution of all parameters and $y$, which can be written as

$$ f(y, \mu, \beta, \delta, \xi, \sigma_e^2) $$

$$ \propto f(y | \mu, \beta, \delta, \xi, \sigma_e^2) f(\beta | \xi) f(\delta) f(\xi) f(\sigma_e^2) $$

$$ \propto \left(\sigma_e^2\right)^{-\frac{n}{2}} \exp \left(-\frac{y'\hat{e}}{2\sigma_e^2}\right) \times \prod_{j=1}^{k} \frac{1}{\sqrt{2\pi} \left(\sigma_j^2\right)^{\frac{1}{2}}} \exp \left(-\frac{\beta_j^2}{2\sigma_j^2}\right) \times \prod_{j=1}^{k} \pi^{(1-\delta_j)}(1-\pi)^{\delta_j} \times \prod_{j=1}^{k} \left(\frac{S_j^2}{\sigma_j^2}\right)^{-\frac{n}{2}} \Gamma\left(\frac{v}{2}\right) \left(\frac{v}{2}\right)^{\frac{v}{2}} \exp \left(-\frac{v\beta_j^2}{2\sigma_j^2}\right) \times \prod_{j=1}^{k} \left(\frac{S_j^2}{\sigma_j^2}\right)^{-\frac{n}{2}} \Gamma\left(\frac{v}{2}\right) \left(\frac{v}{2}\right)^{\frac{v}{2}} \exp \left(-\frac{v\beta_j^2}{2\sigma_j^2}\right).$$

The full conditional distribution of $\beta_j$ is now obtained by dropping factors that do not involve $\beta_j$, which gives

$$ f(\beta_j | y, \mu, \beta_{-j}, \delta, \xi, \sigma_e^2) $$

$$ \propto f(y | \mu, \beta, \delta, \xi, \sigma_e^2) f(\beta_j | \sigma_j^2) $$

$$ \propto \exp \left(-\frac{y'\hat{e}}{2\sigma_e^2}\right) \exp \left(-\frac{\beta_j^2}{2\sigma_j^2}\right) $$

$$ \propto \exp \left(-\frac{\left(\mathbf{w} - X_j\beta_j\delta_j\right)' \left(\mathbf{w} - X_j\beta_j\delta_j\right)}{2\sigma_j^2}\right) \times \exp \left(-\frac{\beta_j^2}{2\sigma_j^2}\right),$$

where $\mathbf{w} = y - 1\mu - \sum_{j \neq j} X_j \beta_j \delta_j$. When $\delta_j = 1$,

$$ f(\beta_j | y, \mu, \beta_{-j}, \delta, \xi, \sigma_e^2) $$

$$ \propto \exp \left(-\frac{\left(\mathbf{w} - X_j\beta_j\right)' \left(\mathbf{w} - X_j\beta_j\right)}{2\sigma_j^2}\right) \exp \left(-\frac{\beta_j^2}{2\sigma_j^2}\right) $$

$$ \propto \exp \left(-\frac{1}{2} \left(\frac{\beta_j - X_j'w}{\sigma_j^2}\right)^2 \right) \sigma_{c_j}^2. $$

Now, (3) can be recognized as the kernel of a normal distribution with mean $\frac{X_j'w}{c_j}$ and variance $\frac{\sigma_{c_j}^2}{c_j}$. When $\delta_j = 0$,

$$ f(\beta_j | y, \mu, \beta_{-j}, \delta, \xi, \sigma_e^2) $$

$$ \propto \exp \left(-\frac{\beta_j^2}{2\sigma_j^2}\right), $$

and dropping the factor $\exp\left(-\frac{w'w}{2\sigma_e^2}\right)$, which is free of $\beta_j$, gives

$$ f(\beta_j | y, \mu, \beta_{-j}, \delta, \xi, \sigma_e^2) \propto \exp \left(-\frac{\beta_j^2}{2\sigma_j^2}\right), $$

which is the kernel of a normal distribution with null mean and variance $\sigma_j^2$. Thus,

$$ p(\beta_j | \text{ELSE}) = \begin{cases} \sim N \left(\frac{X_j'w}{c_j}, \frac{\sigma_{c_j}^2}{c_j}\right) & \text{when } \delta_j = 1, \\ \sim N \left(0, \sigma_j^2\right) & \text{when } \delta_j = 0, \end{cases} $$

where ELSE stands for all the other parameters and $y$. This means when $\delta_j = 1$, the sampling of $\beta_j$ is identical to that in BayesA; when $\delta_j = 0$, $\beta_j$ is sampled from its prior. This is different from the original implementation of BayesB introduced by Meuwissen et al. [4].

Similarly, the full conditional distribution of $\delta_j$ can be obtained from (2) by dropping all factors free of $\delta_j$, which gives

$$ \Pr(\delta_j = 1 | \text{ELSE}) $$

$$ \propto f(y | \mu, \beta, \delta, \xi, \sigma_e^2) \Pr(\delta_j = 1) $$

$$ \propto \exp \left[-\frac{(\mathbf{w} - X_j\beta_j\delta_j)'(\mathbf{w} - X_j\beta_j\delta_j)}{2\sigma_{c_j}^2}\right] \left(1 - \pi\right) \times \exp \left[-\frac{(\mathbf{w} - X_j\beta_j)'(\mathbf{w} - X_j\beta_j)}{2\sigma_j^2}\right] \left(1 - \pi\right), $$

$$ \Pr(\delta_j = 0 | \text{ELSE}) $$

$$ \propto f(y | \mu, \beta, \delta, \xi, \sigma_e^2) \Pr(\delta_j = 0) $$

$$ \propto \exp \left[-\frac{(\mathbf{w} - X_j\beta_j\delta_j)'(\mathbf{w} - X_j\beta_j\delta_j)}{2\sigma_{c_j}^2}\right] \pi \times \exp \left[-\frac{(\mathbf{w} - X_j\beta_j)'(\mathbf{w} - X_j\beta_j)}{2\sigma_j^2}\right] \pi.$$
Thus,
\[
\Pr(\delta_j = 1 \mid \text{ELSE}) = \frac{(1-\pi)d_j}{(1-\pi)d_j + \pi \exp\left(-\frac{w_\delta w}{2\sigma^2}\right)},
\]
where \(d_j = \exp\left(-\frac{(w-X_j\beta_j)^2}{2\sigma^2}\right)\).

**Joint Gibbs sampler**

The same priors as in single-site Gibbs sampler are used here. The only difference is that \(\delta_j, \beta_j\) are sampled from their joint full conditional distribution, which can be written as the product of the full conditional distribution of \(\beta_j\) given \(\delta_j\) and marginal full conditional distribution of \(\delta_j\):

\[
f(\delta_j, \beta_j | y, \mu, \beta_{-j}, \delta_{-j}, \xi, \sigma^2_e) = f(\beta_j | \delta_j, y, \mu, \beta_{-j}, \delta_{-j}, \xi, \sigma^2_e) \times \Pr(\delta_j | y, \mu, \beta_{-j}, \delta_{-j}, \xi, \sigma^2_e).
\]

Thus, \(\delta_j\) is first sampled from \(f(\delta_j | y, \mu, \beta_{-j}, \delta_{-j}, \xi, \sigma^2_e)\). Then \(\beta_j\) is sampled from \(f(\beta_j | \delta_j, y, \mu, \beta_{-j}, \delta_{-j}, \xi, \sigma^2_e)\), which is identical to the sampling of \(\beta_j\) in BayesB with single-site Gibbs sampler. The marginal full conditional for \(\delta_j\) can be written as:

\[
f(\delta_j | y, \mu, \beta_{-j}, \delta_{-j}, \xi, \sigma^2_e) 
\propto f(y | \delta_j, y, \mu, \beta_{-j}, \delta_{-j}, \xi, \sigma^2_e) \Pr(\delta_j) 
\propto f(w | \delta_j, \sigma^2_e) \Pr(\delta_j),
\]

where \(w = y - \mu - \sum_{j \neq j} X_j \beta_j \delta_j = e + X_j \beta \delta_j\). Now \(f(w | \delta_j, \sigma^2_e)\) is a multivariate normal distribution with mean \(E(e + X_j \beta \delta_j | \delta_j, \sigma^2_e, \sigma^2_e)\) and variance \(\text{Var}(X_j \beta \delta_j + e | \delta_j, \sigma^2_e, \sigma^2_e)\). When \(\delta_j = 1\), it becomes a multivariate normal with null mean and variance \(X_jX_j^\prime \sigma^2_e + I \sigma^2_e\), when \(\delta_j = 0\), it becomes a multivariate normal with null mean and variance \(I \sigma^2_e\). Thus, samples can be drawn using

\[
f(\delta_j = 1 | y, \mu, \beta_{-j}, \delta_{-j}, \xi, \sigma^2_e) = \frac{h_1 \Pr(\delta_j = 1)}{1 + h_0 \Pr(\delta_j = 0)}.
\]

where \(h_i = f(w | \delta_i = i, \sigma^2_e, \sigma^2_e)\). However, evaluating the multivariate normal distribution \(f(w | \delta_j, \sigma^2_e, \sigma^2_e)\) is computationally intense. An efficient way is to use the univariate distribution of \(X_jw\), which contains all the information from \(w\) about \(\beta_j\), instead of the distribution of \(w\), which is a multivariate. Thus, (5) can be written as

\[
f(\delta_j = 1 | x_\delta y, \mu, \beta_{-j}, \delta_{-j}, \xi, \sigma^2_e) = \frac{m_1 \Pr(\delta_j = 1)}{m_1 \Pr(\delta_j = 1) + m_0 \Pr(\delta_j = 0)},
\]

where \(m_1 = f(X_j^\prime X_j | \delta_j = 0, \mu, \beta_{-j}, \delta_{-j}, \xi, \sigma^2_e)\), \(m_1\) is an univariate normal distribution with null mean and variance \((X_j^\prime X_j)^2 + X_jX_j^\prime \sigma^2_e\), and \(m_0\) is an univariate normal distribution with null mean and variance \(X_j^\prime X_j \sigma^2_e\).

**Gibbs sampler with pseudo priors**

Here, following Carlin and Chib [7], a pseudo prior is used for \(\beta_j\) when \(\delta_j\) is zero. They proposed to use the full conditional distribution of \(\beta_j\) when \(\delta_j = 1\) as the pseudo prior for \(\beta_j\) when \(\delta_j = 0\) [7, 8], which results in the prior for \(\beta_j\) as

\[
\beta_j | \delta_j = \begin{cases} \sim N(0, \sigma^2_e) & \delta_j = 1 \\ \sim f(\beta_j | \delta_j = 1, y, \mu, \beta_{-j}, \delta_{-j}, \xi, \sigma^2_e) & \delta_j = 0. \\
\end{cases}
\]

We show below that the posterior mean of the marker effect, \(\alpha_j = \beta_j \delta_j\), does not depend on the pseudo prior. This posterior mean can be written as:

\[
E(\beta_j | \delta_j | y) = \sum_{\delta_j} \int \beta_j f(\beta_j | \delta_j | y) d\beta_j
\]

\[
= \sum_{\delta_j} \int \beta_j f(\beta_j | \delta_j) f(\beta_j | \delta_j) \Pr(\delta_j) d\beta_j
\]

\[
= \frac{\int \beta_j f(y | \beta_j, \delta_j = 1) f(\beta_j | \delta_j = 1) \Pr(\delta_j = 1) d\beta_j}{f(y)}.
\]

The numerator in (7) is free of the pseudo prior: \(f(\beta_j | \delta_j = 0)\). Furthermore, it can be seen from the model equation (1) that the value of \(y\) is free of \(\beta_j\) when \(\delta_j = 0\). Thus, the marginal distribution of \(y\), the denominator of (7), does not depend on the pseudo prior, which is the distribution of \(\beta_j\) when \(\delta_j = 0\). As both the numerator and denominator of (1) are free of the pseudo prior, it follows that the posterior mean of \(\alpha_j\) does not depend on the pseudo prior for \(\beta_j\).

We show here that, given this pseudo prior, the full conditional for \(\delta_j\) is identical to the marginal full conditional
distribution of \( \delta_j \), \( \Pr (\delta_j | y, \mu, \beta_{-j}, \delta_{-j}, \xi, \sigma_e^2) \), which is used in the joint Gibbs sampler.

The full conditional probability of \( \delta_j = 1 \) can be written as:

\[
\Pr (\delta_j = 1 \mid \text{ELSE}) = \frac{g_1}{g_1 + g_0} = \frac{1}{1 + \frac{g_0}{g_1}}, \tag{8}
\]

where

\[
g_i = \Pr (\delta_j = i \mid \text{ELSE}) = \frac{f (w \mid \delta_j = i, \beta_j, \delta_{-j}, \mu, \xi, \sigma_e^2)}{f (w \mid \delta_j = 0)} \times f (\beta_j \mid \delta_j = 0) \times \Pr (\delta_j = 0) \times \frac{\Pr (\delta_j = 0)}{\Pr (\delta_j = 1)}. \tag{9}
\]

In the above equation, (9) is identical to \( h_0 \) in (5) used in the joint Gibbs sampler, because \( f (w \mid \delta_j = 0, \beta_j, \sigma_e^2) \) is also a multivariate normal with mean \( I \sigma_e^2 \). We show below that (10) is identical to \( h_1^{-1} \). Our proposed prior for \( \beta_j \) when \( \delta_j = 0 \) (the pseudo prior) can be written as:

\[
p (\beta_j \mid \delta_j = 0) = f (\beta_j \mid \delta_j = 1, y, \mu, \beta_{-j}, \delta_{-j}, \xi, \sigma_e^2) = f (\beta_j \mid \delta_j = 1, w, \sigma_e^2, \sigma_e^2)
\]

\[
= \frac{f (w \mid \delta_j = 1, \beta_j, \sigma_e^2) f (\beta_j \mid \delta_j = 1, \sigma_e^2)}{f (w \mid \delta_j = 1, \sigma_e^2, \sigma_e^2)}. \tag{11}
\]

After replacing the pseudo prior in (10) with (11), it becomes

\[
\Pr (\delta_j = 1 \mid \text{ELSE}) = \frac{1}{f (w \mid \delta_j = 1, \sigma_e^2, \sigma_e^2)}, \text{ which is identical to } h_1^{-1}.
\]

Thus, the ratio \( \frac{g_0}{g_1} \) in (8) is identical to \( \frac{h_0 \Pr (\delta_j = 1)}{h_1 \Pr (\delta_j = 0)} \) in (6), which proves the full conditional probability (8) of \( \delta_j = 1 \), when the proposed prior is used, is identical to (6), the marginal full conditional probability of \( \delta_j = 1 \), which is used in the joint Gibbs sampler.

Use of the exact full conditional distribution as the pseudo prior in BayesB, however, will require MH to sample \( \sigma_e^2 \) and \( \sigma_f^2 \). Thus, a distribution close to the full conditional is used here. Here, we use a normal distribution with mean \( \overline{\sigma_e^2} \), and variance \( \frac{\overline{\sigma_e^2}}{\sigma_f^2} \), where \( \overline{\sigma_e^2} = \frac{\sigma_e^2}{\sigma_f^2} \) and \( \sigma_e^2 \) and \( \sigma_f^2 \) are means of the prior distributions for the residual and the marker effect variances, respectively.

Next, we will show the derivation of the full conditionals, which are proportional to the joint distribution of all parameters and \( y \). Here, the joint distribution of all parameters and \( y \) can be written as:

\[
f (y, \mu, \beta, \delta, \xi, \sigma_e^2) \propto f (y \mid \mu, \beta, \delta, \xi, \sigma_e^2) f (\beta \mid \delta, \xi) f (\delta) f (\xi) f \left( \sigma_e^2 \right) \tag{12}
\]

\[
\propto (\sigma_e^2)^{-n} \exp \left( - \frac{e e}{2 \sigma_e^2} \right) \tag{13}
\]

\[
\times \prod_{j=1}^{k} \left[ \frac{1}{\sqrt{2\pi} (\sigma_e^2)^{\frac{1}{2}}} \exp \left( - \frac{\beta_j^2}{2 \sigma_e^2} \right) \right]^{\delta_j} \tag{14}
\]

\[
\times \prod_{j=1}^{k} \left[ \frac{1}{\sqrt{2\pi} (\sigma_e^2)^{\frac{1}{2}}} \right]^{1-\delta_j} \tag{15}
\]

\[
\times \prod_{j=1}^{k} \left[ \exp \left[ - \frac{(\beta_j - X_j^w)^2}{2 \sigma_e^2 X_j + \lambda} \right] \right]^{1-\delta_j} \tag{16}
\]

\[
\times \prod_{j=1}^{k} \left[ \frac{1}{\Gamma (\frac{\nu}{2})} \right]^{1-\delta_j} \tag{17}
\]

\[
\times \prod_{j=1}^{k} \left( \frac{\nu}{2} \right)^{-\frac{\nu}{2}} (\sigma_f^2)^{-\frac{\nu}{2} - 1} \exp \left( - \frac{\nu \beta_j^2 \sigma_f^2}{2 \sigma_e^2} \right). \tag{18}
\]
\[
x = \left( \frac{\sigma^2_y}{\sigma^2_f} \right)^{\frac{n}{2}} \left( \frac{\sigma^2_f}{\sigma^2_e} \right)^{\frac{n}{2}} \exp \left( -\frac{v_\beta S^2_f + \delta_f^2}{2\sigma^2_e} \right). \tag{19}
\]

It is easy to see that the full conditional distribution of \( \sigma^2_e \), which does not involve \( \delta_j \), is the same as that in the single-site Gibbs sampler. Although \( \mu \) also appears in \( w \) in (16), (16) has no effect on the full conditional of \( \mu \) because the columns of \( X \), which are always centered, are orthogonal to the column vectors of ones so that \( X_j^T \mu = 0 \). Thus, the full conditional of \( \mu \) is the same as that in the single-site Gibbs sampler. When \( \delta_j = 1 \), the full conditional distribution of \( \beta_j \) is identical to that in the single-site Gibbs sampler. When \( \delta_j = 0 \), (16) is the only part that includes \( \beta_j \). Thus the full conditional distribution of \( \beta_j \) is:

\[
p(\beta_j | \text{ELSE}) = \begin{cases} 
N \left( \frac{X_j^T w + \sigma^2_e}{\sigma_j^2} X_j^T X_j \delta_j = 1, \\
N \left( \frac{X_j^T w}{\sigma^2_e} X_j + \frac{\sigma^2_f}{\sigma^2_e} \right) \sigma_j^2 \delta_j = 0.
\end{cases}
\]

When \( \delta_j = 1 \), the full conditional distribution of \( \sigma^2_e \) is the same as that in the single-site Gibbs sampler. When \( \delta_j = 0 \), (18) is the only part that contains \( \sigma^2_f \), which means it should be sampled from its prior. Thus when \( \delta_j = 1 \), the full conditional distribution of \( \sigma^2_f \) is a scaled inverted chi-square distribution with scale parameter \( v_\beta S^2_f + \delta_f^2 \) and degrees of freedom parameter \( v_\beta \); when \( \delta_j = 0 \), it is a scaled inverted chi-square distribution with scale parameter \( \sigma^2_f \) and degrees of freedom parameter \( v_\beta \). The full conditional distribution of \( \delta_j \) can be obtained from the joint distribution of all parameters and \( y \) by dropping all factors free of \( \delta_j \), which gives

\[
\Pr (\delta_j = 1 | \text{ELSE}) \propto f (y | \mu, \beta, \delta, \xi, \sigma^2_e) \Pr (\delta_j = 1) f (\beta_j | \delta_j = 1),
\]

\[
\Pr (\delta_j = 0 | \text{ELSE}) \propto f (y | \mu, \beta, \delta, \xi, \sigma^2_e) \Pr (\delta_j = 0) f (\beta_j | \delta_j = 0).
\]

Compared to the full conditional distributions for \( \delta_j \) in the single-site Gibbs sampler, the difference is the extra factor \( f (\beta_j | \delta_j = 1) \) and \( f (\beta_j | \delta_j = 0) \), because they cannot be canceled out as in the single-site Gibbs sampler. Thus,

\[
\Pr (\delta_j = 1 | \text{ELSE}) = \frac{(1 - \pi) d f (\beta_j | \delta_j = 1)}{(1 - \pi) d f (\beta_j | \delta_j = 1) + \pi \exp \left( -\frac{w^T w}{2\sigma^2_e} f (\beta_j | \delta_j = 0) \right)},
\]

where \( f (\beta_j | \delta_j = 1) = \frac{1}{\sqrt{2\pi} (\sigma_j^2)^{\frac{3}{2}}} \exp \left( -\frac{1}{2} \left( \frac{\beta_j - \frac{X_j^T w}{X_j^T X_j + 2} \sigma_j^2}{\sigma_j^2} \right)^2 \right) \) and

\[
f (\beta_j | \delta_j = 0) = \frac{1}{\sqrt{2\pi} (\sigma_j^2)^{\frac{3}{2}}} \exp \left( -\frac{1}{2} \left( \frac{\beta_j - \frac{X_j^T w}{X_j^T X_j + 2} \sigma_j^2}{\sigma_j^2} \right)^2 \right).
\]

Data

Real genotypic data and simulated phenotypic data were used here to compare BayesB using MH, efficient MH or the three different Gibbs samplers as described above. The genotypic data included 3961 individuals with 55,734 SNPs. The heritability of the simulated trait was 0.25. The training data contained 3206 individuals and the remaining individuals were used for testing. A chain of length of 50,000 was used to estimate parameters of interest. Prediction accuracies were calculated using different samplers. The effective sample sizes [11], which estimate the number of independent samples from a chain, were calculated for \( \sigma^2_e \) to compare convergence rates for different methods. Computing time for different methods with the same number of iterations were also compared.

Results

The number of effective samples per second of computing time was obtained for BayesB using MH, efficient MH or the three different Gibbs samplers. These three Gibbs samplers were almost twice as efficient as Metropolis–Hastings (Table 1). The prediction accuracies for different samplers, which are calculated as the correlation between
estimated breeding values and simulated phenotypes, are all equal to 0.296. Posterior means of \( \mu \) for these four samplers are all equal to 2.508. Posterior means of \( \sigma_e^2 \) for different samplers are almost equal, ranging from 0.955 to 0.957.

**Discussion**

In the joint Gibbs sampler, \( \delta_j \) and \( \beta_j \) are sampled jointly, which addresses the problem of dependence between \( \delta_j \) and \( \beta_j \). Thus, the joint sampler had the largest effective sample size. However, in the single-site Gibbs sampler, \( \delta_j \) and \( \beta_j \) are sampled from their full conditionals, and thus due to the dependence between \( \beta_j \) and \( \delta_j \), the single-site Gibbs sampler had the smallest effective sample size. These differences in effective sample size, however, were small.

In the Gibbs sampler with pseudo priors, \( \beta_j \) and \( \delta_j \) are also sampled from their full conditionals. Recall that we have shown that the posterior mean of the marker effects does not depend on the pseudo prior. Furthermore, Godsill [8] has shown that the marginal posterior for parameters in the model do not depend on the pseudo prior, which is the prior for \( \beta_j \) when \( \delta_j = 0 \). As suggested by Carlin and Chib [7], when the full conditional distribution of \( \beta_j \) when \( \delta_j = 1 \) is chosen to be the pseudo prior, we have shown that the samples of \( \beta_j \) and \( \delta_j \) are identically distributed to those from the joint Gibbs sampler. Thus, the Gibbs sampler with pseudo priors will have a similar effective sample size as the joint Gibbs sampler.

However, when the full conditional distribution for \( \beta_j \) when \( \delta_j = 1 \) is used as the pseudo prior in BayesB, the full conditional distributions of \( \sigma^2_e \) and \( \sigma^2_\delta \) are not of known forms because \( \sigma^2_\delta \) and \( \sigma^2_e \) are in the pseudo prior for the marker effect. In contrast to BayesB, in the model used by Godsill [8] to justify the use of full conditional distributions as the pseudo priors, for simplicity, hyper-parameters such as \( \sigma^2_e \) were omitted [8]. Here, we have replaced \( \sigma^2_\delta \) in the pseudo prior with constants such that the full conditionals for \( \sigma^2_e \) and \( \sigma^2_\delta \) have scaled inverted chi-square distributions. This modification will give a pseudo prior whose distribution is close to that of the full conditional. In the Gibbs sampler with this pseudo prior, the effective sample size was smaller than in the joint Gibbs sampler but still larger than in the single-site Gibbs sampler.

**Conclusions**

When a MH algorithm is used to jointly sample the marker effect and the locus-specific variance, the BayesB method is computationally intensive. After introducing a variable \( \delta_j \), indicating whether the marker effect for a locus is zero or non-zero, the marker effect and locus-specific variance can be sampled using Gibbs sampler without MH. Among the Gibbs samplers that were considered here, the joint Gibbs sampler is the most efficient. This sampler is about 2.1 times as efficient as the MH algorithm proposed by Meuwissen et al. [4] and 1.7 times as efficient as that proposed by Habier et al. [6].

**References**

1. Gianola D. Priors in whole-genome regression: the bayesian alphabet returns. Genetics. 2013;157:1819–29.
2. Norris JR. Markov chains. Cambridge series on statistical and probabilistic mathematics. New York: Cambridge University Press; 1997.
3. Sorensen DA, Gianola D. Likelihood, Bayesian, and MCMC methods in quantitative genetics. New York: Springer; 2002.
4. Meuwissen THE, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. Genetics. 2001;157:1819–29.
5. Gianola D, de los Campos G, Hill WG, Manfred E, Fernando R. Additive genetic variability and the bayesian alphabet. Genetics. 2009;183:347–63.
6. Habier D, Fernando RL, Kizilkaya K, Garrick DJ. Extension of the bayesian alphabet for genomic selection. BMC Bioinformatics. 2011;12:186.
7. Carlin BP, Chib S. Bayesian model choice via Markov chain Monte Carlo methods. J R Stat Soc Series B Stat Methodol. 1995;57:473–84.
8. Godsill SJ. On the relationship between Markov chain Monte Carlo methods for model uncertainty. J Comput Graph Stat. 2001;10:230–48.
9. Habier D, Fernando RL, Kizilkaya K, Garrick DJ. Extension of the Bayesian alphabet for genomic selection. In: Proceedings of the 8th world congress on genetics applied to livestock production, 1–6 August 2010, Leipzig, 2010. http://www.kongressband.de/wcgalp2010/assets/pdf/0468.pdf.
10. Fernando RL, Garrick D. Bayesian methods applied to GWAS. Methods Mol Biol. 2013;1019:237–74.
11. Geyer CJ. Practical Markov chain Monte Carlo. Stat Sci. 1992;7:473–511.