GSE Genetics Selection Evolution

RESEARCH ARTICLE

Open Access

CrossMark

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

Hao Cheng[1,2], Long Qu[3], Dorian J. Garrick[1,4] and Rohan L. Fernando[1*]

Abstract

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. 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 π 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

the locus-specific variance as the proposal distribution. When π 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 δ_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 δ_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 σ_j^2:

$$\left(\alpha_j | \sigma_j^2\right) \begin{cases} = 0 & \text{with probability } \pi \\ \sim N\left(0, \sigma_j^2\right) & \text{with probability } (1-\pi), \end{cases}$$

where σ_j^2 follows a scaled inverted chi-square distribution and π is treated as known. Employing the concept of data augmentation, it is convenient to write the marker effect α_j as $\alpha_j = \beta_j \delta_j$, where we introduce a Bernoulli variable δ_j with probability of success $1 - \pi$ and normally distributed variable β_j with mean zero and variance σ_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 δ_j, β_j are sampled from the joint full-conditional distribution $f\left(\delta_j, \beta_j | \mathbf{y}, \mu, \boldsymbol{\beta}_{-j}, \boldsymbol{\delta}_{-j}, \boldsymbol{\xi}, \sigma_e^2\right)$, where $\boldsymbol{\xi} = \left[\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2\right]$, because δ_j and β_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 β_j when δ_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 β_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 δ_j from $f\left(\delta_j | \mathbf{y}, \mu, \boldsymbol{\beta}_{-j}, \boldsymbol{\delta}_{-j}, \boldsymbol{\xi}, \sigma_e^2\right)$ [8]. However, in BayesB, use of the exact full conditional distribution as the pseudo prior will require MH to sample σ_j^2 and σ_e^2. Thus in this paper, a distribution close to the full conditional is used.

BayesB model with data augmentation model

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

where y_i is the phenotype for individual i, μ 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), β_j is the allele substitution effect for locus j, δ_j is an indicator variable and e_i is the random residual effect for individual i.

Priors
The prior for μ is a constant. The prior for e_i is $e_i | \sigma_e^2 \overset{iid}{\sim} N\left(0, \sigma_e^2\right)$ and $\left(\sigma_e^2 | \nu_e, S_e^2\right) \sim \nu_e S_e^2 \chi_{\nu_e}^{-2}$. The prior for β_j is $\beta_j | \sigma_j^2 \sim N\left(0, \sigma_j^2\right)$. The prior for $\left(\sigma_j^2 | \nu_\beta, S_\beta^2\right) \sim \nu_\beta S_\beta^2 \chi_{\nu_\beta}^{-2}$. The prior for δ_j is

$$\left(\delta_j | \pi\right) \begin{cases} = 1 & probability \, (1 - \pi) \\ = 0 & probability \, \pi. \end{cases}$$

Single-site Gibbs sampler
The full conditional distributions of μ, σ_j^2 and σ_e^2 are well-known [4, 9, 10]. Thus they are presented here without derivations. The full conditional of μ is a normal distribution with mean $\frac{1'\mathbf{e}_\mu}{n}$ and variance $\frac{\sigma_e^2}{n}$, where $\mathbf{e}_\mu = \mathbf{y} - \sum_{j=1}^{k} \mathbf{X}_j \beta_j \delta_j$ and n is the number of individuals. The full conditional distributions of σ_j^2 and σ_e^2 are both scaled inverted chi-square distributions; for σ_j^2, the scale parameter is $\frac{\left(\nu_\beta S_\beta^2 + \beta_j^2\right)}{\nu_\beta + 1}$ and the degrees of freedom parameter are $\nu_\beta + 1$; for σ_e^2, the scale parameter is $\frac{\nu_e S_e^2 + \mathbf{e}'\mathbf{e}}{\nu_e + n}$ and the degrees of freedom parameter are $\nu_e + n$, where $\mathbf{e} = \mathbf{y} - \mathbf{1}\mu - \sum_{j=1}^{k} \mathbf{X}_j \beta_j \delta_j$.

Cheng *et al. Genet Sel Evol (2015) 47:80*

Page 3 of 7

Next, we derive the full conditional distributions of β_j and δ_j. These full conditional distributions are proportional to the joint distribution of all parameters and **y**, which can be written as

$$
\begin{aligned}
&f\left(\boldsymbol{y}, \mu, \boldsymbol{\beta}, \delta, \boldsymbol{\xi}, \sigma_e^2\right) \\
&\propto f\left(\boldsymbol{y} \mid \mu, \boldsymbol{\beta}, \delta, \boldsymbol{\xi}, \sigma_e^2\right) f(\boldsymbol{\beta}|\boldsymbol{\xi}) f(\delta) f(\boldsymbol{\xi}) f\left(\sigma_e^2\right) \\
&\propto \left(\sigma_e^2\right)^{-\frac{n}{2}} \exp\left(-\frac{\boldsymbol{e}'\boldsymbol{e}}{2\sigma_e^2}\right) \\
&\quad \times \prod_{j=1}^{k} \frac{1}{\sqrt{2\pi}\left(\sigma_j^2\right)^{\frac{1}{2}}} \exp\left(-\frac{\beta_j^2}{2\sigma_j^2}\right) \\
&\quad \times \prod_{j=1}^{k} \pi^{(1-\delta_j)}(1-\pi)^{\delta_j} \\
&\quad \times \prod_{j=1}^{k} \frac{\left(S_\beta^2\frac{\nu_\beta}{2}\right)^{\frac{\nu_\beta}{2}}}{\Gamma\left(\frac{\nu_\beta}{2}\right)} \left(\sigma_j^2\right)^{\left(-\frac{\nu_\beta+2}{2}\right)} \exp\left(-\frac{\nu_\beta S_\beta^2}{2\sigma_j^2}\right) \\
&\quad \times \frac{\left(S_e^2\frac{\nu_e}{2}\right)^{\frac{\nu_e}{2}}}{\Gamma\left(\frac{\nu_e}{2}\right)} \left(\sigma_e^2\right)^{\left(-\frac{\nu_e+2}{2}\right)} \exp\left(-\frac{\nu_e S_e^2}{2\sigma_e^2}\right).
\end{aligned}
\tag{2}
$$

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

$$
\begin{aligned}
&f\left(\beta_j|\boldsymbol{y}, \mu, \boldsymbol{\beta}_{-j}, \delta, \boldsymbol{\xi}, \sigma_e^2\right) \\
&\propto f\left(\boldsymbol{y} \mid \mu, \boldsymbol{\beta}, \delta, \boldsymbol{\xi}, \sigma_e^2\right) f\left(\beta_j|\sigma_j^2\right) \\
&\propto \exp\left(-\frac{\boldsymbol{e}'\boldsymbol{e}}{2\sigma_e^2}\right) \exp\left(-\frac{\beta_j^2}{2\sigma_j^2}\right) \\
&\propto \exp\left[-\frac{(\boldsymbol{w}-\boldsymbol{X_j}\beta_j\delta_j)'(\boldsymbol{w}-\boldsymbol{X_j}\beta_j\delta_j)}{2\sigma_e^2}\right] \\
&\quad \times \exp\left(-\frac{\beta_j^2}{2\sigma_j^2}\right),
\end{aligned}
$$

where $\boldsymbol{w} = \boldsymbol{y} - \boldsymbol{1}\mu - \sum_{j'\neq j}\boldsymbol{X}_{j'}\beta_{j'}\delta_{j'}$. When $\delta_j = 1$,

$$
\begin{aligned}
&f\left(\beta_j|\boldsymbol{y}, \mu, \boldsymbol{\beta}_{-j}, \delta, \boldsymbol{\xi}, \sigma_e^2\right) \\
&\propto \exp\left[-\frac{(\boldsymbol{w}-X_j\beta_j)'(\boldsymbol{w}-X_j\beta_j)}{2\sigma_e^2}\right] \exp\left(-\frac{\beta_j^2}{2\sigma_j^2}\right) \\
&\propto \exp\left[-\frac{1}{2}\frac{\left(\beta_j - \frac{\boldsymbol{X}_j'\boldsymbol{w}}{c_j}\right)^2}{\frac{\sigma_e^2}{c_j}}\right]
\end{aligned}
\tag{3}
$$

where $c_j = \boldsymbol{X}_j'\boldsymbol{X}_j + \frac{\sigma_e^2}{\sigma_j^2}$. Now, (3) can be recognized as the kernel of a normal distribution with mean $\frac{\boldsymbol{X}_j'\boldsymbol{w}}{c_j}$ and variance $\frac{\sigma_e^2}{c_j}$. When $\delta_j = 0$,

$$
\begin{aligned}
&f\left(\beta_j|\boldsymbol{y}, \mu, \boldsymbol{\beta}_{-j}, \delta, \boldsymbol{\xi}, \sigma_e^2\right) \\
&\propto \exp\left(-\frac{\boldsymbol{w}'\boldsymbol{w}}{2\sigma_e^2}\right) \exp\left(-\frac{\beta_j^2}{2\sigma_j^2}\right),
\end{aligned}
$$

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

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

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

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

where ELSE stands for all the other parameters and y. This means when $\delta_j = 1$, the sampling of β_j is identical to that in BayesA; when $\delta_j = 0$, β_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 δ_j can be obtained from (2) by dropping all factors free of δ_j, which gives

$$
\begin{aligned}
&\Pr\left(\delta_j = 1 \mid ELSE\right) \\
&\propto f\left(\boldsymbol{y} \mid \mu, \boldsymbol{\beta}, \delta, \boldsymbol{\xi}, \sigma_e^2\right) \Pr\left(\delta_j = 1\right) \\
&\propto \exp\left[-\frac{(\boldsymbol{w}-\boldsymbol{X_j}\beta_j\delta_j)'(\boldsymbol{w}-\boldsymbol{X_j}\beta_j\delta_j)}{2\sigma_e^2}\right](1-\pi) \\
&\propto \exp\left[-\frac{(\boldsymbol{w}-X_j\beta_j)'(\boldsymbol{w}-X_j\beta_j)}{2\sigma_e^2}\right](1-\pi),
\end{aligned}
$$

$$
\begin{aligned}
&\Pr\left(\delta_j = 0 \mid ELSE\right) \\
&\propto f\left(\boldsymbol{y} \mid \mu, \boldsymbol{\beta}, \delta, \boldsymbol{\xi}, \sigma_e^2\right) \Pr\left(\delta_j = 0\right) \\
&\propto \exp\left[-\frac{(\boldsymbol{w}-\boldsymbol{X_j}\beta_j\delta_j)'(\boldsymbol{w}-\boldsymbol{X_j}\beta_j\delta_j)}{2\sigma_e^2}\right]\pi \\
&\propto \exp\left(-\frac{\boldsymbol{w}'\boldsymbol{w}}{2\sigma_e^2}\right)\pi.
\end{aligned}
$$

Cheng *et al. Genet Sel Evol* (2015) 47:80

Page 4 of 7

Thus,

$$\Pr\left(\delta_j = 1 \mid ELSE\right)$$

$$= \frac{(1-\pi)d_j}{(1-\pi)d_j + \pi \exp\left(-\frac{w'w}{2\sigma_e^2}\right)},$$

where $d_j = \exp\left[-\frac{\left(w - X_j\beta_j\right)'\left(w - X_j\beta_j\right)}{2\sigma_e^2}\right]$.

Joint Gibbs sampler

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

$$f\left(\delta_j, \beta_j | y, \mu, \boldsymbol{\beta}_{-j}, \boldsymbol{\delta}_{-j}, \boldsymbol{\xi}, \sigma_e^2\right)$$

$$= f\left(\beta_j | \delta_j, y, \mu, \boldsymbol{\beta}_{-j}, \boldsymbol{\delta}_{-j}, \boldsymbol{\xi}, \sigma_e^2\right)$$

$$\times \Pr\left(\delta_j | y, \mu, \boldsymbol{\beta}_{-j}, \boldsymbol{\delta}_{-j}, \boldsymbol{\xi}, \sigma_e^2\right).$$

Thus, δ_j is first sampled from $f\left(\delta_j | y, \mu, \boldsymbol{\beta}_{-j}, \boldsymbol{\delta}_{-j}, \boldsymbol{\xi}, \sigma_e^2\right)$. Then β_j is sampled from $f\left(\beta_j | \delta_j, y, \mu, \boldsymbol{\beta}_{-j}, \boldsymbol{\delta}_{-j}, \boldsymbol{\xi}, \sigma_e^2\right)$, which is identical to the sampling of β_j in BayesB with single-site Gibbs sampler. The marginal full conditional for δ_j can be written as:

$$f\left(\delta_j | y, \mu, \boldsymbol{\beta}_{-j}, \boldsymbol{\delta}_{-j}, \boldsymbol{\xi}, \sigma_e^2\right)$$

$$\propto f\left(y \mid \delta_j, \mu, \boldsymbol{\beta}_{-j}, \boldsymbol{\delta}_{-j}, \boldsymbol{\xi}, \sigma_e^2\right) \Pr\left(\delta_j\right) \quad (4)$$

$$\propto f\left(w \mid \delta_j, \sigma_j^2, \sigma_e^2\right) \Pr\left(\delta_j\right),$$

where $w = y - \mathbf{1}\mu - \sum_{j' \neq j} X_{j'}\beta_{j'}\delta_{j'} = e + X_j\beta_j\delta_j$. Now $f\left(w \mid \delta_j, \sigma_j^2, \sigma_e^2\right)$ is a multivariate normal distribution with mean $E\left(e + X_j\beta_j\delta_j \mid \delta_j, \sigma_j^2, \sigma_e^2\right)$ and variance $Var\left(X_j\beta_j\delta_j + e \mid \delta_j, \sigma_j^2, \sigma_e^2\right)$. When $\delta_j = 1$, it becomes a multivariate normal with null mean and variance $X_jX_j'\sigma_j^2 + I\sigma_e^2$; when $\delta_j = 0$, it becomes a multivariate normal with null mean and variance $I\sigma_e^2$. Thus, samples can be drawn using

$$f\left(\delta_j = 1 | y, \mu, \boldsymbol{\beta}_{-j}, \boldsymbol{\delta}_{-j}, \boldsymbol{\xi}, \sigma_e^2\right)$$

$$= \frac{h_1 \Pr\left(\delta_j = 1\right)}{h_1 \Pr\left(\delta_j = 1\right) + h_0 \Pr\left(\delta_j = 0\right)}, \quad (5)$$

$$= \frac{1}{1 + \frac{h_0 Pr(\delta_j = 0)}{h_1 \Pr(\delta_j = 1)}} \quad (6)$$

where $h_i = f\left(w \mid \delta_j = i, \sigma_j^2, \sigma_e^2\right)$. However, evaluating the multivariate normal distribution $f\left(w \mid \delta_j, \sigma_j^2, \sigma_e^2\right)$ is computationally intense. An efficient way is to use the univariate distribution of $X_j'w$, which contains all the information from w about β_j, instead of the distribution of w, which is a multivariate. Thus, (5) can be written as

$$f\left(\delta_j = 1 | y, \mu, \boldsymbol{\beta}_{-j}, \boldsymbol{\delta}_{-j}, \boldsymbol{\xi}, \sigma_e^2\right)$$

$$= \frac{m_1 \Pr\left(\delta_j = 1\right)}{m_1 \Pr\left(\delta_j = 1\right) + m_0 \Pr\left(\delta_j = 0\right)},$$

where $m_i = f\left(X_j'w \mid \delta_j = i, \mu, \boldsymbol{\beta}_{-j}, \boldsymbol{\delta}_{-j}, \boldsymbol{\xi}, \sigma_e^2\right)$, m_1 is an univariate normal distribution with null mean and variance $\left(X_j'X_j\right)^2\sigma_j^2 + X_j'X_j\sigma_e^2$, and m_0 is an univariate normal distribution with null mean and variance $X_j'X_j\sigma_e^2$.

Gibbs sampler with pseudo priors

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

$$\beta_j | \delta_j$$

$$= \begin{cases} \sim N\left(0, \sigma_j^2\right) & \delta_j = 1 \\ \sim f\left(\beta_j | \delta_j = 1, y, \mu, \boldsymbol{\beta}_{-j}, \boldsymbol{\delta}_{-j}, \boldsymbol{\xi}, \sigma_e^2\right) & \delta_j = 0. \end{cases}$$

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:

$$E\left(\beta_j\delta_j \mid \mathbf{y}\right)$$

$$= \sum_{\delta_j} \int \beta_j\delta_j f\left(\beta_j, \delta_j \mid \mathbf{y}\right) d\beta_j$$

$$= \frac{\sum_{\delta_j} \int \beta_j\delta_j f\left(\mathbf{y} \mid \beta_j, \delta_j\right) f\left(\beta_j \mid \delta_j\right) \Pr\left(\delta_j\right) d\beta_j}{f(\mathbf{y})} \quad (7)$$

$$= \frac{\int \beta_j f\left(\mathbf{y} \mid \beta_j, \delta_j = 1\right) f\left(\beta_j \mid \delta_j = 1\right) \Pr\left(\delta_j = 1\right) d\beta_j}{f(\mathbf{y})}.$$

The numerator in (7) is free of the pseudo prior: $f\left(\beta_j \mid \delta_j = 0\right)$. Furthermore, it can be seen from the model equation (1) that the value of $\mathbf{y}$ is free of β_j when $\delta_j = 0$. Thus, the marginal distribution of $\mathbf{y}$, the denominator of (7), does not depend on the pseudo prior, which is the distribution of β_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 α_j does not depend on the pseudo prior for β_j.

We show here that, given this pseudo prior, the full conditional for δ_j is identical to the marginal full conditional

Cheng *et al. Genet Sel Evol (2015) 47:80*

Page 5 of 7

distribution of δ_j, $\Pr\left(\delta_j | \boldsymbol{y}, \mu, \boldsymbol{\beta}_{-j}, \boldsymbol{\delta}_{-j}, \boldsymbol{\xi}, \sigma_e^2\right)$, which is used in the joint Gibbs sampler.

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

$$\Pr\left(\delta_j = 1 \mid ELSE\right) = \frac{g_1}{g_1 + g_0}$$
$$= \frac{1}{1 + \frac{g_0}{g_1}}, \tag{8}$$

where

$$g_i = Pr\left(\delta_j = i \mid ELSE\right)$$
$$\propto f\left(\boldsymbol{y} \mid \delta_j = i, \beta_j, \boldsymbol{\delta}_{-j}, \boldsymbol{\beta}_{-j}, \mu, \boldsymbol{\xi}, \sigma_e^2\right)$$
$$\times f\left(\beta_j \mid \delta_j = i, \sigma_j^2\right) \times p\left(\delta_j = i\right)$$
$$\propto f\left(\boldsymbol{w} \mid \delta_j = i, \beta_j, \sigma_e^2\right)$$
$$\times f\left(\beta_j \mid \delta_j = i, \sigma_j^2\right) \times p\left(\delta_j = i\right)$$

with $i = 0$ or 1.

The ratio in the denominator of (8) is:

$$\frac{g_0}{g_1} = f\left(\boldsymbol{w} \mid \delta_j = 0, \beta_j, \sigma_e^2\right) \tag{9}$$

$$\times \frac{f\left(\beta_j \mid \delta_j = 0\right)}{f\left(\boldsymbol{w} \mid \delta_j = 1, \beta_j, \sigma_e^2\right) \times f\left(\beta_j \mid \delta_j = 1, \sigma_j^2\right)}$$
$$\times \frac{Pr\left(\delta_j = 0\right)}{Pr\left(\delta_j = 1\right)}. \tag{10}$$

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

$$p\left(\beta_j \mid \delta_j = 0\right)$$
$$= f\left(\beta_j | \delta_j = 1, \boldsymbol{y}, \mu, \boldsymbol{\beta}_{-j}, \boldsymbol{\delta}_{-j}, \boldsymbol{\xi}, \sigma_e^2\right)$$
$$= f\left(\beta_j | \delta_j = 1, \boldsymbol{w}, \sigma_e^2, \sigma_j^2\right) \tag{11}$$
$$= \frac{f\left(\boldsymbol{w} \mid \delta_j = 1, \beta_j, \sigma_e^2\right) f\left(\beta_j \mid \delta_j = 1, \sigma_j^2\right)}{f\left(\boldsymbol{w} \mid \delta_j = 1, \sigma_j^2, \sigma_e^2\right)}.$$

After replacing the pseudo prior in (10) with (11), it becomes $\dfrac{1}{f\left(\boldsymbol{w} \mid \delta_j = 1, \sigma_j^2, \sigma_e^2\right)}$, 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 σ_j^2 and σ_e^2. Thus, a distribution close to the full conditional is used here. Here, we use a normal distribution with mean $\dfrac{\boldsymbol{X}_j' \boldsymbol{w}}{\boldsymbol{X}_j' \boldsymbol{X}_j + \widetilde{\lambda}}$ and variance $\dfrac{\widetilde{\sigma_e^2}}{\boldsymbol{X}_j' \boldsymbol{X}_j + \widetilde{\lambda}}$, where $\widetilde{\lambda} = \dfrac{\widetilde{\sigma_e^2}}{\widetilde{\sigma_j^2}}$, and $\widetilde{\sigma_e^2}$ and $\widetilde{\sigma_j^2}$ 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 $\boldsymbol{y}$. Here, the joint distribution of all parameters and $\boldsymbol{y}$ can be written as:

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

$$\propto \left(\sigma_e^2\right)^{-\frac{n}{2}} \exp\left(-\frac{\boldsymbol{e}' \boldsymbol{e}}{2\sigma_e^2}\right) \tag{13}$$

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

$$\times \prod_{j=1}^{k} \left[\frac{1}{\sqrt{2\pi}\left(\frac{\widetilde{\sigma_e^2}}{\boldsymbol{X}_j' \boldsymbol{X}_j + \widetilde{\lambda}}\right)^{\frac{1}{2}}}\right]^{1-\delta_j} \tag{15}$$

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

$$\times \prod_{j=1}^{k} \pi^{(1-\delta_j)} (1-\pi)^{\delta_j} \tag{17}$$

$$\times \prod_{j=1}^{k} \frac{\left(S_\beta^2 \frac{\nu_\beta}{2}\right)^{\frac{\nu_\beta}{2}}}{\Gamma\left(\frac{\nu_\beta}{2}\right)} \left(\sigma_j^2\right)^{\left(-\frac{\nu_\beta+2}{2}\right)} \exp\left(-\frac{\nu_\beta S_\beta^2}{2\sigma_j^2}\right) \tag{18}$$

Cheng *et al. Genet Sel Evol (2015) 47:80*

Page 6 of 7

Table 1 Efficiency of alternative MCMC samplers for BayesB

	Alternative MCMC samplers				
	MH	Efficient MH	Single-site Gibbs	Joint Gibbs	Gibbs with pseudo prior
Computing time	90,009	70,714	52,452	44,726	47,043
Effective sample size	25,262	24,588	24,684	26,757	25,036
Effective samples/s	0.280	0.347	0.471	0.598	0.532

Efficiency of alternative MCMC samplers for BayesB. Results are given for the computing time in seconds to obtain 50,000 samples, effective sample size and effective samples/s for BayesB using Metropolis–Hastings (MH), single-site Gibbs sampler, joint Gibbs sampler and Gibbs sampler with pseudo priors

$$\times \frac{\left(S_e^2 \frac{v_e}{2}\right)^{\frac{v_e}{2}}}{\Gamma\left(\frac{v_e}{2}\right)} \left(\sigma_e^2\right)^{\left(-\frac{v_e+2}{2}\right)} \exp\left(-\frac{v_e S_e^2}{2\sigma_e^2}\right). \quad (19)$$

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

$$p(\beta_j | ELSE) = \begin{cases} \sim N\left(\frac{X_j'\boldsymbol{w}}{X_j'X_j + \frac{\sigma_e^2}{\sigma_j^2}}, \frac{\sigma_e^2}{X_j'X_j + \frac{\sigma_e^2}{\sigma_j^2}}\right) & \delta_j = 1, \\ \sim N\left(\frac{X_j'\boldsymbol{w}}{X_j'X_j + \widetilde{\lambda}}, \frac{\widetilde{\sigma_e^2}}{X_j'X_j + \widetilde{\lambda}}\right) & \delta_j = 0. \end{cases}$$

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

$$\Pr\left(\delta_j = 1 \mid ELSE\right)$$
$$\propto f\left(\boldsymbol{y} \mid \mu, \boldsymbol{\beta}, \boldsymbol{\delta}, \boldsymbol{\xi}, \sigma_e^2\right) \Pr\left(\delta_j = 1\right) f\left(\beta_j \mid \delta_j = 1\right),$$

$$\Pr\left(\delta_j = 0 \mid ELSE\right)$$
$$\propto f\left(\boldsymbol{y} \mid \mu, \boldsymbol{\beta}, \boldsymbol{\delta}, \boldsymbol{\xi}, \sigma_e^2\right) \Pr\left(\delta_j = 0\right) f\left(\beta_j \mid \delta_j = 0\right).$$

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

$$\Pr\left(\delta_j = 1 \mid ELSE\right)$$
$$= \frac{(1-\pi)d_j f\left(\beta_j \mid \delta_j = 1\right)}{(1-\pi)d_j f\left(\beta_j \mid \delta_j = 1\right) + \pi \exp\left(-\frac{\boldsymbol{w}'\boldsymbol{w}}{2\sigma_e^2}\right) f\left(\beta_j \mid \delta_j = 0\right)},$$

where $\quad f\left(\beta_j \mid \delta_j = 1\right) = \frac{1}{\sqrt{2\pi}\left(\sigma_j^2\right)^{\frac{1}{2}}} \exp\left(-\frac{\beta_j^2}{2\sigma_j^2}\right) \quad$ and

$$f\left(\beta_j \mid \delta_j = 0\right) = \frac{1}{\sqrt{2\pi}\left(\frac{\widetilde{\sigma_e^2}}{X_j'X_j + \widetilde{\lambda}}\right)^{\frac{1}{2}}} \exp\left[-\frac{\left(\beta_j - \frac{X_j'\boldsymbol{w}}{X_j'X_j + \widetilde{\lambda}}\right)^2}{2\frac{\widetilde{\sigma_e^2}}{X_j'X_j + \widetilde{\lambda}}}\right].$$

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 σ_e^2 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

Cheng *et al. Genet Sel Evol* (2015) 47:80

Page 7 of 7

estimated breeding values and simulated phenotypes, are all equal to 0.296. Posterior means of μ for these four samplers are all equal to 2.508. Posterior means of σ_e^2 for different samplers are almost equal, ranging from 0.955 to 0.957.

Discussion

In the joint Gibbs sampler, δ_j and β_j are sampled jointly, which addresses the problem of dependence between δ_j and β_j. Thus, the joint sampler had the largest effective sample size. However, in the single-site Gibbs sampler, δ_j and β_j are sampled from their full conditionals, and thus due to the dependence between β_j and δ_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, β_j and δ_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 β_j when $\delta_j = 0$. As suggested by Carlin and Chib [7], when the full conditional distribution of β_j when $\delta_j = 1$ is chosen to be the pseudo prior, we have shown that the samples of β_j and δ_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 β_j when $\delta_j = 1$ is used as the pseudo prior in BayesB, the full conditional distributions of σ_e^2 and σ_j^2 are not of known forms because σ_e^2 and σ_j^2 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 σ_e^2 were omitted [8]. Here, we have replaced σ_e^2 and σ_j^2 in the pseudo prior with constants such that the full conditionals for σ_e^2 and σ_j^2 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 δ_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].

Authors' contributions
HC, LQ, DJG, RLF contributed to the development of the statistical methods. HC wrote the program code and conducted the analyses. The manuscript was prepared by RLF and HC. All authors read and approved the final manuscript.

Author details
[1] Department of Animal Science, Iowa State University, Ames 50011, IA, USA. [2] Department of Statistics, Iowa State University, Ames, IA 50011, USA. [3] Department of Mathematics and Statistics, Wright State University, Dayton, OH 45435, USA. [4] Institute of Veterinary, Animal and Biomedical Science, Massey University, Palmerston North, New Zealand.

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, Manfredi 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 9th 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.