

# Multiple-trait Bayesian Regression Methods with Mixture Priors for Genomic Prediction

Hao Cheng<sup>\*,\*\*,\*1</sup>, Kadir Kizilkaya<sup>†</sup>, Jian Zeng<sup>‡</sup>, Dorian Garrick<sup>§</sup> and Rohan Fernando<sup>\*</sup>

<sup>\*</sup>Department of Animal Science, Iowa State University, Ames, Iowa, <sup>\*\*</sup>Department of Statistics, Iowa State University, Ames, Iowa, <sup>†</sup>Department of Animal Science, Adnan Menderes University, Aydin, Turkey, <sup>‡</sup>Program in Complex Trait Genomics, Institute for Molecular Bioscience, University of Queensland, St Lucia, Brisbane, Australia, <sup>§</sup>Institute of Veterinary, Animal, and Biomedical Science, Massey University, Palmerston North, New Zealand

**ABSTRACT** Bayesian multiple-regression methods incorporating different mixture priors for marker effects are widely used in genomic prediction. Improvement in prediction accuracies from using those methods, such as BayesB, BayesC and BayesC $\pi$ , have been shown in single-trait analyses with both simulated data and real data. These methods have been extended to multi-trait analyses, but only under a specific limited circumstance that assumes a locus affects all the traits or none of them. In this paper, we develop and implement the most general multi-trait BayesCII and BayesB methods allowing a broader range of mixture priors. Further, we compare them to single-trait methods and the “restricted” multi-trait formulation using real data. In those data analyses, significant higher prediction accuracies were sometimes observed from these new broad-based multi-trait Bayesian multiple-regression methods. The software tool JVAS offers routines to perform the analyses.

**KEYWORDS** multi-trait; mixture priors; genomic prediction

## Introduction

Genomic prediction was proposed by Meuwissen et al. (Meuwissen et al. 2001) to incorporate whole-genome data into genetic evaluation. In genomic prediction, all the marker or haplotype effects are estimated simultaneously, and these estimates can then be used to predict breeding values of individuals not in the training population used to estimate the effects.

Bayesian multiple-regression methods incorporating mixture priors for marker effects are widely used in genomic prediction. For example, BayesB with locus specific variances accommodates models where markers have zero effects with probability  $\pi$  (Meuwissen et al. 2001; Cheng et al. 2015). Another mixture model, BayesC, assumes a common locus variance for all markers, and its extension known as BayesC $\pi$  further treats  $\pi$  as an unknown parameter with a uniform prior distribution (Habier et al. 2011).

Bayesian multiple-regression methods were first proposed for single-trait analyses but have been extended to some particular forms of multi-trait analyses (Calus and Veerkamp 2011; Jia and Jannink 2012). Those extensions have pertained to a particular,

somewhat restrictive mixture model. The “restricted” multi-trait BayesCII presented by Jia et al. (Jia and Jannink 2012) assumes a variant affects none of the traits or has simultaneous effects on all traits. This assumption of genetic architecture in that multi-trait BayesCII circumstance is violated if some loci have no effect on at least one of the traits while having an effect on at least one of the other traits.

In this paper, we present a more general class of multi-trait BayesCII and BayesB methods for which the previous multi-trait model is a special case. The new methods are compared to the previous multi-trait methods and to single-trait methods with real data.

## Materials and Methods

### Multi-trait Marker Effects Model

For simplicity and without loss of generality, we will assume a general mean as the only fixed effect, and write the multi-trait model for individual  $i$  from  $n$  genotyped individuals as

$$y_i = \mu + \sum_{j=1}^p m_{ij} \alpha_j + e_i,$$

where  $y_i$  is a vector of phenotypes of  $t$  traits for individual  $i$ ,  $\mu$  is a vector of overall means for  $t$  traits,  $m_{ij}$  is the genotype covariate

at locus  $j$  for individual  $i$ ,  $p$  is the number of genotyped loci,  $\alpha_j$  is a vector of allele substitution effects of  $t$  traits for locus  $j$ , and  $e_i$  is a vector of random residuals of  $t$  traits for individual  $i$ . The fixed effects, or general mean in this case, are assigned flat priors. The residuals,  $e_i$ , are a priori assumed to be independently and identically distributed multivariate normal vectors with null mean and covariance matrix  $R$ , which in turn is assumed to have an inverse Wishart prior distribution,  $W_t^{-1}(\mathbf{S}_e, \nu_e)$ .

### Multi-trait BayesCPI model

**Priors for marker effects** The prior for  $\alpha_{jk}$ , the allele substitution or marker effect of trait  $k$  for locus  $j$ , is a mixture with a point mass at zero and a univariate normal distribution conditional on  $\sigma_k^2$ :

$$\alpha_{jk} \mid \pi_k, \sigma_k^2 \begin{cases} \sim N(0, \sigma_k^2) & \text{probability } (1 - \pi_k) \\ 0 & \text{probability } \pi_k \end{cases}$$

and the covariance between effects for traits  $k$  and  $k'$  at the same locus, i.e.,  $\alpha_{jk}$  and  $\alpha_{jk'}$  is

$$\text{cov}(\alpha_{jk}, \alpha_{jk'} \mid \sigma_{kk'}) = \begin{cases} \sigma_{kk'} & \text{if both } \alpha_{jk} \neq 0 \text{ and } \alpha_{jk'} \neq 0 \\ 0 & \text{otherwise} \end{cases}.$$

Employing the concept of data augmentation, the vector of marker effects at a particular locus  $\alpha_j$  can be written as  $\alpha_j = \mathbf{D}_j \beta_j$ , where  $\mathbf{D}_j$  is a diagonal matrix with elements  $\text{diag}(\mathbf{D}_j) = \delta_j = (\delta_{j1}, \delta_{j2}, \delta_{j3}, \dots)$ , where  $\delta_{jk}$  is an indicator variable indicating whether the marker effect of locus  $j$  for trait  $k$  is zero or non-zero, and  $\beta_j$  follows a multivariate normal distribution with null mean

$$\text{and covariance matrix } \mathbf{G} = \begin{bmatrix} \sigma_1^2 & \cdots & \sigma_{1t} \\ \vdots & \ddots & \vdots \\ \sigma_{1t} & \cdots & \sigma_t^2 \end{bmatrix}. \text{ The covariance}$$

matrix  $\mathbf{G}$  is a priori assumed to follow an inverse Wishart distribution,  $W_t^{-1}(\mathbf{S}_\beta, \nu_\beta)$ . Thus the multi-trait BayesCPI model with data augmentation is written as

$$\mathbf{y}_i = \boldsymbol{\mu} + \sum_{j=1}^p m_{ij} \mathbf{D}_j \beta_j + \mathbf{e}_i. \quad (1)$$

In the most general case, any marker effect might be zero for any possible combination of  $t$  traits resulting in  $2^t$  possible combinations of  $\delta_j$ . For example, in a  $t=2$  trait model, there are  $2^2 = 4$  combinations of  $\delta_j$ , namely  $\delta_1 = (0, 0)$ ,  $\delta_2 = (0, 1)$ ,  $\delta_3 = (1, 0)$ ,  $\delta_4 = (1, 1)$ . In the special case of this model described by (Jia and Jannink 2012), only  $\delta_1 = (0, 0)$  and  $\delta_4 = (1, 1)$  have non-zero probability. Suppose in general we use numerical labels "1", "2", ..., "l" for the  $2^t$  possible outcomes for  $\delta_j$ , then the prior for  $\delta_j$  is a categorical distribution

$$p(\delta_j = "i") = \Pi_1 I(\delta_j = "1") + \Pi_2 I(\delta_j = "2") + \dots + \Pi_l I(\delta_j = "l"),$$

where  $\Pi_i$  is the probability that the vector  $\delta_j = "i"$  and  $\sum_{i=1}^l \Pi_i = 1$ .

A Dirichlet distribution with all parameters equal to one, i.e., a uniform distribution, can be used for the prior for  $\Pi = (\Pi_1, \Pi_2, \dots, \Pi_l)$ . As shown below, a Gibbs sampler can be used to draw samples for all the parameters in this model.

**Gibbs sampler I for multi-trait BayesCPI** Suppose the prior for  $\delta_j$  is a categorical distribution whose support is for all  $2^t$  possible outcomes of  $\delta_j$ . For convenience, from now on let "1" denote trait  $k$  and "2" the other  $t - 1$  traits. In our sampling scheme,  $\beta_{j1}$  and  $\delta_{j1}$  are sampled from their joint full conditional distributions, which can be written as the product of the full conditional distribution of  $\beta_{j1}$  given  $\delta_{j1}$  and the marginal full conditional distribution of  $\delta_{j1}$ . Let  $\boldsymbol{\theta}$  denote all other parameters except  $\delta_{j1}$  and  $\beta_{j1}$ , then our sampling scheme can be written as

$$f(\beta_{j1}, \delta_{j1} \mid \boldsymbol{\theta}, \mathbf{y}) = f(\beta_{j1} \mid \delta_{j1}, \boldsymbol{\theta}, \mathbf{y}) f(\delta_{j1} \mid \boldsymbol{\theta}, \mathbf{y}).$$

The full conditional distributions of  $\beta_{j1}$ ,  $\delta_{j1}$ ,  $\Pi$ ,  $\mathbf{G}$  and  $\mathbf{R}$  for Gibbs sampler I, which were derived in the Appendix, are given below.

The full conditional distributions of  $\beta_{j1}$  is

$$p(\beta_{j1} \mid \delta_{j1}, \boldsymbol{\theta}, \mathbf{y}) = \begin{cases} N(\hat{\beta}_{j1}^0, (\mathbf{G}^{11})^{-1}) & \text{when } \delta_{j1} = 0 \\ N(\hat{\beta}_{j1}^1, (\mathbf{C}_{j,11}^1)^{-1}) & \text{when } \delta_{j1} = 1 \end{cases},$$

with

$$\begin{aligned} \hat{\beta}_{j1}^0 &= -(\mathbf{G}^{11})^{-1} \mathbf{G}^{12} \beta_{j2}, \\ \hat{\beta}_{j1}^1 &= (\mathbf{C}_{j,11}^1)^{-1} (r_{j1} - \mathbf{C}_{j,12}^1 \beta_{j2}), \\ \mathbf{C}_{j,11}^1 &= \mathbf{G}^{11} + \mathbf{R}^{11} \sum_{i=1}^n m_{ij}^2, \\ \mathbf{C}_{j,12}^1 &= \mathbf{G}^{12} + \mathbf{R}^{12} \mathbf{D}_{j2} \sum_{i=1}^n m_{ij}^2, \\ r_{j1} &= \left( \sum_{i=1}^n \mathbf{w}'_i m_{ij} \right) \left[ \mathbf{R}^{21} \right], \end{aligned}$$

where  $\mathbf{w}_i = \mathbf{y}_i - \boldsymbol{\mu}_i - \sum_{j' \neq j} m_{ij'} \mathbf{D}_{j'} \beta_{j'}$ .

The marginal full conditional probability of  $\delta_{j1} = 1$  is

$$f(\delta_{j1} = 1 \mid \boldsymbol{\theta}, \mathbf{y}) = \left\{ 1 + \left( \frac{\text{Pr}(\delta_{j1} = 0, \delta_{j2} \mid \Pi)}{\text{Pr}(\delta_{j1} = 1, \delta_{j2} \mid \Pi)} H \right)^{-1} \right\}^{-1},$$

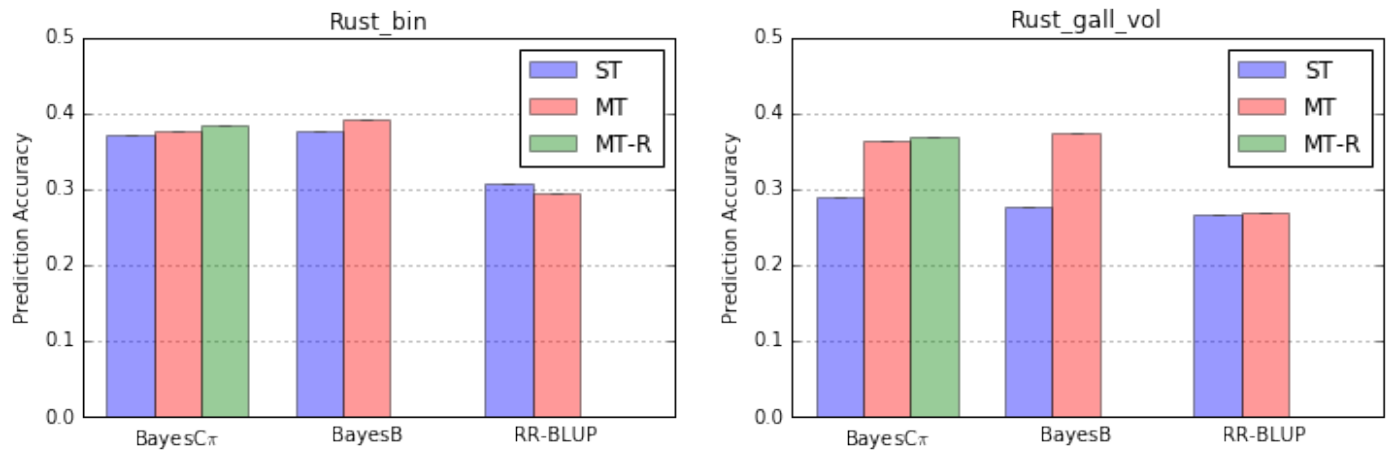
where  $H = \exp \left\{ -\frac{1}{2} \left( \log \mathbf{C}_{j,11}^1 - \hat{\beta}_{j1}^{12} \mathbf{C}_{j,11}^1 \right) - \left( -\frac{1}{2} \left( \log \mathbf{G}^{11} - \hat{\beta}_{j1}^{02} \mathbf{G}^{11} \right) \right) \right\}$ .

The full conditional distribution for  $\Pi$  can be written as

$$f(\Pi \mid \beta, \mathbf{D}, \mathbf{G}, \mathbf{R}, \mathbf{y}) \propto \text{Dirichlet}(n_1 + 1, n_2 + 1, \dots),$$

where  $n_i$  is the number of markers with  $\delta_j = "i"$ .

The full conditional distributions for  $\mathbf{R}$ , the covariance matrix for residuals, is an inverse Wishart distribution,  $W_t^{-1}(\mathbf{S}_e + \mathbf{e}'\mathbf{e}, \nu_e + n)$ , where  $\mathbf{e}$  is the  $n \times t$  matrix for residuals with the  $i$ th row as  $\mathbf{e}'_i$ . The full conditional distribution for  $\mathbf{G}$ , the covariance matrix for  $\beta_j$ , is an inverse Wishart distribution,  $W_t^{-1}(\mathbf{S}_\beta + \beta' \beta, \nu_\beta + p)$ , where  $\beta$  is the  $p \times t$  matrix with the  $i$ th row as  $\beta'_i$ .



**Figure 1** Comparison of single-trait and multi-trait methods for Rust\_bin and Rust\_gall\_vol traits.

**Gibbs sampler II for multi-trait BayesC $\pi$**  The Gibbs sampler above requires that all  $2^t$  outcomes for  $\delta_j$  have non-zero prior probabilities, i.e. none of  $\Pi_i$  can be zero. If some  $\Pi_i$  are zero, the markov chain generated from Gibbs sampler I may not be irreducible. Another more general Gibbs sampler that does not require all  $\Pi_i$  to be non-zero is proposed below.

The full conditional distributions of  $\beta_j$ ,  $\delta_j$ ,  $\Pi$ ,  $G$ ,  $R$  for Gibbs sampler II, which were derived in the Appendix, are given below.

Let  $\theta$  denote all other parameters except  $\beta_j$  and  $\delta_j$ , then our sampling scheme can be written as

$$f(\beta_j, \delta_j | \theta, y) = f(\delta_j | \theta, y) f(\beta_j | \delta_j, \theta, y).$$

The full conditional distribution of  $\beta_j$  is

$$f(\beta_j | \delta_j, \theta, y) \propto N(C_j^{-1} r_j, C_j^{-1}),$$

where  $C_j = D_j' R^{-1} D_j \sum_{i=1}^n m_{ij}^2 + G^{-1}$  and  $r_j = (\sum_{i=1}^n w_i' m_{ij}) R^{-1} D_j$ .

The marginal full conditional probability of  $\delta_j = "i"$  is

$$f(\delta_j = "i" | \theta, y) = \frac{f(y | \delta_j = "i", \theta) f(\delta_j = "i" | \Pi)}{\sum_{i \in \{ "1", "2", \dots, "I" \}} f(y | \delta_j = "i", \theta) f(\delta_j = "i" | \Pi)},$$

where

$$f(y | \delta_j, \theta) = |C_j^{-1}|^{\frac{1}{2}} \exp \left\{ -\frac{1}{2} r_j' C_j^{-1} r_j \right\}.$$

This Gibbs sampler can accommodate the restricted multi-trait BayesC $\pi$  that was proposed by Jia et al. (Jia and Jannink 2012), which only allows  $\delta_j$  to be a vector of all ones or a vector of all zeros.

### Multi-trait BayesB Model

The multi-trait BayesC $\pi$  model proposed above can be modified to accommodate the multi-trait BayesB model. Model equation (1) can also be used for the multi-trait BayesB method. The differences in multi-trait BayesB method is that the prior for  $\beta_j$  is a multivariate normal distribution with null mean and locus-specific covariance matrix  $G_j$ . The locus-specific covariance matrix  $G_j$  is a priori assumed to follow an inverse Wishart distribution,  $W_t^{-1}(S_\beta, \nu_\beta)$ .

The derivations of the full conditional distributions of parameters of interest for Gibbs samplers are shown in the Appendix. In the multi-trait BayesB model, the full conditional distributions for all parameters except  $G_j$  are similar to the multi-trait BayesC $\pi$  model. The full conditional distribution for  $G_j$ , the covariance matrix for  $\beta_j$ , is a inverse Wishart distribution,  $W_t^{-1}(S_\beta + \beta_j \beta_j', \nu_\beta + 1)$ .

### Data analyses

Published genotypic and deregressed phenotypic data for Loblolly Pine (*Pinus Taeda* L.) were used (Resende et al. 2012). Two disease traits, namely Rust\_bin and Rust\_gall\_vol were analyzed. The reported heritability was 0.21 for Rust\_bin and 0.12 for Rust\_gall\_vol. Loci with missing genotypes were imputed as the mean of the observed genotype covariates at that locus and loci with a missing rate >50% were excluded. After these quality control edits, 4,828 SNPs on 807 individuals with phenotypes and genotypes on both traits remained.

Prediction accuracy was calculated as the correlation between the vector of deregressed phenotypes and the vector of estimated breeding values. Cross-validation using 10-folds formed the basis for comparing our general multi-trait BayesC $\pi$  model (MT-BayesC $\pi$ ) to a similar model where the prior for  $\beta_j$  is a multivariate normal rather than a mixture of multivariate normals (MT-BayesC0), the restricted multi-trait BayesC $\pi$  proposed by Jia et al. (MT-BayesC $\pi$ -R), multi-trait BayesB with known  $\Pi$  (MT-BayesB) and the usual single trait formulations of the mixture models (ST-BayesC0, ST-BayesC $\pi$ , ST-BayesB). The constant  $\Pi$  used in BayesB were estimated using BayesC $\pi$  methods. All analyses were performed using JWAS (Cheng et al. 2016), a publicly-available package for single-trait and multi-trait whole-genome analyses written in the freely-

available Julia language. Since BayesC0 is equivalent to random regression best linear unbiased prediction (RR-BLUP), ST-BayesC0 and MT-BayesC0 are denoted as ST-RR-BLUP and MT-RR-BLUP below. The prior for the residual covariance matrix  $\mathbf{R}$  in all multi-trait methods was an inverse Wishart distribution,  $W^{-1} \left( \begin{bmatrix} 0.003 & 0 \\ 0 & 0.003 \end{bmatrix}, 6 \right)$ , for which the mean of  $\mathbf{R}$  is  $\begin{bmatrix} 0.001 & 0 \\ 0 & 0.001 \end{bmatrix}$ . The prior for the marker effects covariance matrix  $\mathbf{G}$  in MT-BayesCII and MT-BayesCII-R was an inverse Wishart distribution,  $W^{-1} \left( \begin{bmatrix} 0.003 & 0 \\ 0 & 0.003 \end{bmatrix}, 6 \right)$ , for which the mean of  $\mathbf{G}$  was  $\begin{bmatrix} 0.001 & 0 \\ 0 & 0.001 \end{bmatrix}$ . The priors for the residual variance and marker effects variance in single-trait analyses were a scaled inverted chi-squared distribution with scale parameter  $S^2 = 0.0005$  and degrees of freedom  $\nu = 4$ , for which the mean of the prior was also 0.001. Marker effect variances estimated from BayesCII were used to construct the priors for marker effect variances in the BayesB methods.

## Results

The prediction accuracies from all methods for Rust\_bin and Rust\_gall\_vol are in figure 1. The prediction accuracies from all single-trait analyses using JWAS are similar to those in (Resende *et al.* 2012). ST-BayesC $\pi$  showed higher prediction accuracies than ST-RR-BLUP for both traits (Resende *et al.* 2012). The prediction accuracies from ST-BayesB were similar to those from ST-BayesC $\pi$ , when both marker effect variances and  $\pi$  estimated from ST-BayesC $\pi$  were used in ST-BayesB.

The analyses of Rust\_bin exhibited no significant difference between multi-trait and single-trait analyses within each method (ST-RR-BLUP versus MT-RR-BLUP; ST-BayesC $\pi$  versus MT-BayesCII; ST-BayesC $\pi$  versus MT-BayesCII-R; ST-BayesB versus MT-BayesB).

In contrast, analyses for the lower heritability Rust\_gall\_vol with MT-BayesCII showed significantly higher accuracies than ST-BayesC $\pi$ . MT-BayesCII and MT-BayesCII-R showed similar prediction accuracies. The posterior means of  $\Pi$  for both methods were shown in table 1. The performance of MT-BayesB were similar to MT-BayesCII, when both marker effect variances and  $\Pi$  estimated from MT-BayesCII were used. Similar prediction accuracies were observed in MT-RR-BLUP and ST-RR-BLUP for trait Rust\_gall\_vol.

## Discussion

In the single trait analyses, accuracies from ST-BayesC $\pi$  and ST-BayesB were higher than those from ST-RR-BLUP, suggesting that these two traits are influenced by a few QTL with large effects. The effect of genetic architecture on the performance of multi-trait analyses has been studied in previous simulation analyses (Jia and Jannink 2012). Using simulated data they found that multi-trait Bayesian variable selection methods outperform multi-trait RR-BLUP in the presence of major QTL. This observation was confirmed in our real data analyses that MT-BayesCII and MT-BayesB outperformed MT-RR-BLUP for both traits.

Significant differences between multi-trait and single-trait analyses were only observed for Rust\_gall\_vol within BayesC $\pi$

and BayesB methods (MT-BayesCII versus ST-BayesC $\pi$ ; MT-BayesB versus ST-BayesB). MT-BayesCII and MT-BayesCII-R outperformed ST-BayesC $\pi$  for Rust\_gall\_vol, and the accuracy gain was 26% (from 0.287 to 0.364). The lower-heritability trait Rust\_gall\_vol may borrow information from the other correlated trait Rust\_bin. Thus higher prediction accuracy from MT-BayesCII were observed in trait Rust\_gall\_vol instead of Rust\_bin. Results in (Jia and Jannink 2012) showed no difference between MT-BayesCII-R and ST-BayesC $\pi$  because a reduced marker panel (500 markers) was used. The performance of MT-BayesB was similar to MT-BayesCII, when both marker effect variances and  $\Pi$  estimated from MT-BayesCII were used. Further analyses may be required to study the effects of priors in MT-BayesB.

The fact that RR-BLUP showed no improvement in multi-trait analyses suggested that benefits from MT-BayesCII may caused by the estimation of hyper-parameter  $\Pi$ . In the MT-BayesCII, the mean of the posterior probability that a marker has a null effect on Rust\_gall\_vol was about 0.97, calculated as the summation of posterior mean of  $\Pi$  for categories (0,0) and (1,0). The posterior mean of  $\pi$ , the probability that a marker has a null effect, in ST-BayesC $\pi$  for Rust\_gall\_vol was 0.74, different from the equivalent value, 0.97, in MT-BayesCII showed above. Thus ST-BayesC $\pi$  with constant  $\pi$ , equal to 0.97, were performed. Prediction accuracies from ST-BayesC $\pi$  with constant  $\pi = 0.97$  was 0.361, which was similar to the accuracies from MT-BayesCII. This suggests that high-heritability traits may help with variable selection in correlated low-heritability traits.

The difference between MT-BayesCII and MT-BayesCII-R is that MT-BayesCII-R assumes a locus has an effect on all traits or none of them. This assumption of genetic architecture is always violated. MT-BayesCII and MT-BayesCII-R, however, showed similar prediction accuracies. This can be explained by the estimation of  $\Pi$  in MT-BayesCII and MT-BayesCII-R in table 1. The posterior probability means for (0,1) and (1,0) were almost zero in MT-BayesCII and for (0,0) and (1,1) are similar in MT-BayesCII and MT-BayesCII-R, suggesting that the assumption of genetic architecture for MT-BayesCII-R is valid for these two traits.

In practice, genetic variances from previous conventional analyses are always used to construct priors for marker effect variances. For single trait analyses, under some assumptions, it can be shown that the marker effect variance  $\sigma_{\alpha}^2 = \frac{\sigma_g^2}{(1-\pi) \sum 2p_j(1-p_j)}$ , where  $\sigma_g^2$  is the genetic variance,  $p_j$  is the allele frequency for locus  $j$  and  $\pi$  is the probability that a marker has a null effect. Following similar strategies, the marker effect covariance matrix  $\mathbf{G}$  in two-trait analyses can be obtained as

$$\mathbf{G} = \frac{1}{\sum 2p_j(1-p_j)} \begin{bmatrix} \frac{Q_{11}}{p(\delta=(1,1))+p(\delta=(1,0))} & \frac{Q_{12}}{p(\delta=(1,1))} \\ \frac{Q_{21}}{p(\delta=(1,1))} & \frac{Q_{22}}{p(\delta=(1,1))+p(\delta=(0,1))} \end{bmatrix}, \quad (2)$$

where  $\mathbf{Q} = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix}$  is the genetic covariance matrix and  $p(\delta=(0,1))$ ,  $p(\delta=(1,0))$ ,  $p(\delta=(1,1))$  are the probability a marker has null effects on the first trait but not the second trait, on the second trait but not the first trait and on no traits. Thus the probability that a marker has an effect on the first trait can be obtained as  $p(\delta=(1,1)) + p(\delta=(1,0))$ , which is the denominator of the upper left element in (2). This strategy relating

	Different Categories of $\delta$			
	(0,0)	(1,1)	(0,1)	(1,0)
MT-BayesCII	0.966	0.029	0.002	0.003
MT-BayesCII-R	0.971	0.029	NA	NA

**Table 1** Estimation of  $\pi$  for alternative multi-trait BayesCII methods. Posterior mean of  $\Pi$  were given for different categories of  $\delta$ . Different categories of  $\delta$  are denoted as  $(k_1, k_2)$ , where  $k_1 = 0$  if a marker has a null effect on Rust\_bin, otherwise  $k_1 = 1$ , and similarly for  $k_2$  representing sampled effects for Rust\_gall\_vol. Combinations listed as NA do not exist in the restricted model.

genetic covariance matrix to marker effect covariance matrix can also be used for analyses with more than two traits. Note that positive definite matrix **Q** may result in negative definite matrix **G** using (2), especially when the prior for the probability a marker has null effects violates the truth.

**Literature Cited**

Calus, M. P. and R. F. Veerkamp, 2011 Accuracy of multi-trait genomic selection using different methods. *Genetics Selection Evolution* **43**: 26.

Cheng, H., D. J. Garrick, and R. L. Fernando, 2016 JWAS: Julia implementation of whole-genome analyses software using univariate and multivariate Bayesian mixed effects model. available from <http://QTL.rocks>.

Cheng, H., L. Qu, D. J. Garrick, and R. L. Fernando, 2015 A fast and efficient Gibbs sampler for BayesB in whole-genome analyses. *Genetics Selection Evolution* **47**: 1819.

Habier, D., R. L. Fernando, K. Kizilkaya, and D. J. Garrick, 2011 Extension of the bayesian alphabet for genomic selection. *BMC bioinformatics* **12**: 186.

Jia, Y. and J.-L. Jannink, 2012 Multiple-Trait Genomic Selection Methods Increase Genetic Value Prediction Accuracy. *Genetics* **192**: 1513–1522.

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.

Resende, M. F. R., P. Muñoz, M. D. V. Resende, D. J. Garrick, R. L. Fernando, J. M. Davis, E. J. Jokela, T. A. Martin, G. F. Peter, and M. Kirst, 2012 Accuracy of Genomic Selection Methods in a Standard Data Set of Loblolly Pine (*Pinus taeda* L.). *Genetics* **190**: 1503–1510.



## Appendix

### Gibbs sampler algorithm for multi-trait BayesCII

#### Single-site Gibbs sampler for multi-trait BayesCII

The full conditional distribution of  $\beta_{j1}$  can be written as

$$\begin{aligned} f(\beta_{j1} \mid \delta_{j1}, \beta_{-j1}, D_{-j1}, G, R, y) &\propto f(y \mid \mu, \beta, D, G, R) f(\beta_{j1}, \beta_{j2} \mid G) \\ &\propto \exp \left[ -\frac{1}{2} \sum_{i=1}^n (w_i - m_{ij} D_j \beta_j)' R^{-1} (w_i - m_{ij} D_j \beta_j) \right] \exp \left( -\frac{1}{2} \beta_j' G^{-1} \beta_j \right), \end{aligned}$$

where  $w_i = y_i - \mu_i - \sum_{j' \neq j} m_{ij'} D_{j'}' \beta_{j'}$ . Further, by dropping factors that do not involve  $\beta_{j1}$ ,

$$\begin{aligned} f(\beta_{j1} \mid \delta_{j1}, \beta_{-j1}, D_{-j1}, G, R, y) &\propto \exp \left\{ -\frac{1}{2} \left[ \beta_j' \left( D_j' R^{-1} D_j \sum_{i=1}^n m_{ij}^2 + G^{-1} \right) \beta_j - 2 \sum_{i=1}^n w_i' m_{ij} R^{-1} D_j \beta_j \right] \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \left[ \beta_j' C_j \beta_j - 2 r_j' \beta_j \right] \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \begin{bmatrix} \beta_{j1} & \beta_{j2}' \end{bmatrix} \begin{bmatrix} C_{j,11} & C_{j,12} \\ C_{j,21} & C_{j,22} \end{bmatrix} \begin{bmatrix} \beta_{j1} \\ \beta_{j2} \end{bmatrix} - 2 \begin{bmatrix} r_{j1} & r_{j2}' \end{bmatrix} \begin{bmatrix} \beta_{j1} \\ \beta_{j2} \end{bmatrix} \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \left( C_{j,11} \beta_{j1}^2 + (2 C_{j,12} \beta_{j2} - 2 r_{j1}) \beta_{j1} \right) \right\} \\ &\propto \exp \left\{ -\frac{C_{j,11}}{2} \left( \beta_{j1} + (C_{j,12} \beta_{j2} - r_{j1}) C_{j,11}^{-1} \right)^2 \right\} \\ &\propto N \left( C_{j,11}^{-1} (r_{j1} - C_{j,12} \beta_{j2}), C_{j,11}^{-1} \right) \\ &\propto N \left( \hat{\beta}_{j1}, C_{j,11}^{-1} \right) \end{aligned}$$

where  $C_j = D_j' R^{-1} D_j \sum_{i=1}^n m_{ij}^2 + G^{-1}$  and  $r_j' = \left( \sum_{i=1}^n w_i' m_{ij} \right) R^{-1} D_j$ .

Note that when  $\delta_{j1} = 0$ ,

$$\begin{aligned} C_j &= \begin{bmatrix} C_{j,11}^0 & C_{j,12}^0 \\ C_{j,21}^0 & C_{j,22}^0 \end{bmatrix} \\ &= \begin{bmatrix} G^{11} & G^{12} \\ G^{21} & G^{22} + D_{j2}' R^{22} D_{j2} \sum_{i=1}^n m_{ij}^2 \end{bmatrix} \\ r_j' &= \begin{bmatrix} r_{j1}^0 & r_{j2}^{0'} \end{bmatrix} \\ &= \begin{bmatrix} 0 & \left( \sum_{i=1}^n w_i' m_{ij} \right) \begin{bmatrix} R^{12} \\ R^{22} \end{bmatrix} D_{j2} \end{bmatrix} \end{aligned}$$

When  $\delta_{j1} = 1$ ,

$$\begin{aligned} C_j &= \begin{bmatrix} C_{j,11}^1 & C_{j,12}^1 \\ C_{j,21}^1 & C_{j,22}^1 \end{bmatrix} \\ &= \begin{bmatrix} G^{11} + R^{11} \sum_{i=1}^n m_{ij}^2 & G^{12} + R^{12} D_{j2} \sum_{i=1}^n m_{ij}^2 \\ G^{21} + D_{j2}' R^{21} \sum_{i=1}^n m_{ij}^2 & G^{22} + D_{j2}' R^{22} D_{j2} \sum_{i=1}^n m_{ij}^2 \end{bmatrix} \\ r_j' &= \begin{bmatrix} r_{j1}^1 & r_{j2}^{1'} \end{bmatrix} \\ &= \begin{bmatrix} \left( \sum_{i=1}^n w_i' m_{ij} \right) \begin{bmatrix} R^{11} \\ R^{21} \end{bmatrix} & \left( \sum_{i=1}^n w_i' m_{ij} \right) \begin{bmatrix} R^{12} \\ R^{22} \end{bmatrix} D_{j2} \end{bmatrix} \end{aligned}$$

Thus when  $\delta_{j1} = 0$ , the full conditional distribution of  $\beta_{j1}$  is

$$f(\beta_{j1} | \delta_{j1} = 0, \beta_{-j1}, D_{-j1}, G, R, y) \propto N(\hat{\beta}_{j1}^0, (C_{j,11}^0)^{-1}) = N(- (G^{11})^{-1} G^{12} \beta_{j2}, (G^{11})^{-1}).$$

When  $\delta_{j1} = 1$ , the full conditional distribution of  $\beta_{j1}$  becomes

$$f(\beta_{j1} | \delta_{j1} = 1, \beta_{-j1}, D_{-j1}, G, R, y) \propto N(\hat{\beta}_{j1}^1, (C_{j,11}^1)^{-1}) = N((C_{j,11}^1)^{-1} (r_{j1} - C_{j,12}^1 \beta_{j2}), (C_{j,11}^1)^{-1}).$$

The marginal full conditional distribution of  $\delta_{j1}$  can be written as

$$\begin{aligned} f(\delta_{j1} = 1 | \theta, y) &= \frac{f(\delta_{j1} = 1, \theta, y)}{\sum_{\delta_{j1} \in (0,1)} f(\delta_{j1}, \theta, y)} \\ &= \frac{f(y | \delta_{j1} = 1, \theta) f(\delta_{j1} = 1, \delta_{j2} | \Pi)}{\sum_{\delta_{j1} \in (0,1)} f(y | \delta_{j1}, \theta) f(\delta_{j1} | \Pi)} \\ &= \left\{ 1 + \frac{f(y | \delta_{j1} = 0, \theta) f(\delta_{j1} = 0, \delta_{j2} | \Pi)}{f(y | \delta_{j1} = 1, \theta) f(\delta_{j1} = 1, \delta_{j2} | \Pi)} \right\}^{-1} \end{aligned}$$

The factor  $f(y | \delta_{j1}, \theta)$  can be written as

$$\begin{aligned} f(y | \delta_{j1}, \theta) &\propto \int f(y | \mu, \beta_{j1}, \beta_{-j1}, D, G, R) f(\beta_{j1}, \beta_{j2} | G) d\beta_{j1} \\ &\propto \int \exp \left[ -\frac{1}{2} \sum_{i=1}^n (w_i - m_{ij} D_j \beta_j)' R^{-1} (w_i - m_{ij} D_j \beta_j) \right] \exp \left( -\frac{1}{2} \beta_j' G^{-1} \beta_j \right) d\beta_{j1} \\ &\propto \exp \left\{ -\frac{1}{2} \left( \sum_i w_i' R^{-1} w_i - 2r_{j2}' \beta_{j2} + \beta_{j2}' C_{j,22} \beta_{j2} - (r_{j1} - C_{j,12} \beta_{j2})^2 C_{j,11}^{-1} \right) \right\} \\ &\times \int \exp \left[ -\frac{1}{2} (\beta_{j1} - \hat{\beta}_{j1})^2 C_{j,11} \right] d\beta_{j1} \\ &\propto (C_{j,11})^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \left( \sum_i w_i' R^{-1} w_i - 2r_{j2}' \beta_{j2} + \beta_{j2}' C_{j,22} \beta_{j2} - (r_{j1} - C_{j,12} \beta_{j2})^2 C_{j,11}^{-1} \right) \right\} \\ &\propto (C_{j,11})^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \left( \sum_i w_i' R^{-1} w_i - 2r_{j2}' \beta_{j2} + \beta_{j2}' C_{j,22} \beta_{j2} - \hat{\beta}_{j1}^2 C_{j,11} \right) \right\}. \end{aligned}$$

Note that  $\sum_i w_i' R^{-1} w_i, r_{j2}' \beta_{j2}, \beta_{j2}' C_{j,22} \beta_{j2}$  are same when  $\delta_{j1} = 0$  or 1. Thus the ratio  $\frac{f(y|\delta_{j1}=1,\theta)}{f(y|\delta_{j1}=0,\theta)}$  becomes

$$\begin{aligned} H &= (C_{j,11}^1)^{-\frac{1}{2}} (G^{11})^{\frac{1}{2}} \exp \left( -\frac{1}{2} (\hat{\beta}_{j1}^2 G^{11} - \hat{\beta}_{j1}^2 C_{j,11}^1) \right) \\ &= \exp \left\{ -\frac{1}{2} \left( \log C_{j,11}^1 - \hat{\beta}_{j1}^2 C_{j,11}^1 \right) - \left( -\frac{1}{2} (\log G^{11} - \hat{\beta}_{j1}^2 G^{11}) \right) \right\} \end{aligned}$$

Thus the conditional probability of  $\delta_{j1} = 1$  is

$$\left\{ 1 + \frac{f(y | \delta_{j1} = 0, \theta) f(\delta_{j1} = 0, \delta_{j2} | \Pi)}{f(y | \delta_{j1} = 1, \theta) f(\delta_{j1} = 1, \delta_{j2} | \Pi)} \right\}^{-1} = \left\{ 1 + \left( \frac{\Pi_{j0}}{\Pi_{j1}} H \right)^{-1} \right\}^{-1},$$

where  $\Pi_{j0} = Pr(\delta_{j1} = 0, \delta_{j2} | \Pi)$  and  $\Pi_{j1} = Pr(\delta_{j1} = 1, \delta_{j2} | \Pi)$ .

The full conditional distribution for  $\Pi$  can be written as

$$\begin{aligned} f(\Pi | \beta, D, G, R, y) &\propto f(\delta | \Pi) f(\Pi) \\ &\propto \Pi_1^{n_1} \Pi_2^{n_2} \dots \Pi_l^{n_l} \\ &\propto \text{Dirichlet}(n_1 + 1, n_2 + 1, \dots), \end{aligned}$$

where  $n_i$  is the number of markers with  $\delta_j = "i"$ .

### 315 **Joint Gibbs sampler for multi-trait BayesCII**

Let  $\theta$  denote all other parameters except  $\beta_j$  and  $\delta_j$ , then our sampling scheme can be written as

$$f(\beta_j, \delta_j | \theta, y) = f(\delta_j | \theta, y) f(\beta_j | \delta_j, \theta, y)$$

The marginal full conditional distribution of  $\delta_j$  can be written as

$$\begin{aligned} f(\delta_j | \theta, y) &= \frac{f(\delta_j, \theta, y)}{\sum_{\delta_j} f(\delta_j, \theta, y)} \\ &= \frac{f(y | \delta_j, \theta) f(\delta_j | \Pi)}{\sum_{\delta_j} f(y | \delta_j, \theta) f(\delta_j | \Pi)}. \end{aligned}$$

316 Denote  $w_i = y_i - \mu_i - \sum_{j \neq i} m_{ij} D_j \beta_j$ , then

$$\begin{aligned} f(y | \delta_j, \theta) &\propto \int f(y | \beta, D, R) f(\beta_j | G) d\beta_j \\ &\propto \int \exp \left[ -\frac{1}{2} \sum_{i=1}^n (w_i - m_{ij} D_j \beta_j)' R^{-1} (w_i - m_{ij} D_j \beta_j) \right] \exp \left( -\frac{1}{2} \beta_j' G^{-1} \beta_j \right) d\beta_j \\ &\propto \int \exp \left\{ -\frac{1}{2} \left[ \beta_j' \left( D_j' R^{-1} D_j \sum_{i=1}^n m_{ij}^2 + G^{-1} \right) \beta_j - 2 \sum_{i=1}^n w_i' m_{ij} R^{-1} D_j \beta_j + \sum_{i=1}^n w_i' R^{-1} w_i \right] \right\} d\beta_j \\ &\propto \int \exp \left\{ -\frac{1}{2} \left[ \beta_j' C_j \beta_j - 2 r_j' \beta_j + \sum_{i=1}^n w_i' R^{-1} w_i \right] \right\} d\beta_j \\ &\propto \int \exp \left\{ -\frac{1}{2} \left[ (\beta_j' - r_j' C_j^{-1}) C_j (\beta_j - C_j^{-1} r_j) + \sum_{i=1}^n w_i' R^{-1} w_i - r_j' C_j^{-1} r_j \right] \right\} d\beta_j \\ &\propto \exp \left\{ -\frac{1}{2} \left[ \sum_{i=1}^n w_i' R^{-1} w_i - r_j' C_j^{-1} r_j \right] \right\} \\ &\propto |C_j^{-1}|^{\frac{1}{2}} \int |C_j^{-1}|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (\beta_j' - r_j' C_j^{-1}) C_j (\beta_j - C_j^{-1} r_j) \right] d\beta_j \\ &\propto |C_j^{-1}|^{\frac{1}{2}} \exp \left\{ -\frac{1}{2} \left[ \sum_{i=1}^n w_i' R^{-1} w_i - r_j' C_j^{-1} r_j \right] \right\}, \end{aligned}$$

317 where  $C_j = D_j' R^{-1} D_j \sum_{i=1}^n m_{ij}^2 + G^{-1}$  and  $r_j' = \left( \sum_{i=1}^n w_i' m_{ij} \right) R^{-1} D_j$ .

Note that  $\sum_i w_i' R^{-1} w_i$  is same for different  $\delta_j$ . Thus the marginal full conditional distribution of  $\delta_j$  can be written as

$$f(\delta_j | \theta, y) = \frac{f(y | \delta_j, \theta) f(\delta_j | \Pi)}{\sum_{\delta_j} f(y | \delta_j, \theta) f(\delta_j | \Pi)},$$

where

$$f(y | \delta_j, \theta) \propto |C_j^{-1}|^{\frac{1}{2}} \exp \left\{ \frac{1}{2} r_j' C_j^{-1} r_j \right\}.$$

318 The full conditional distribution of  $\beta_j$  is

$$\begin{aligned} f(\beta_j | \delta_j, \theta, y) &\propto \exp \left[ -\frac{1}{2} \sum_{i=1}^n (w_i - m_{ij} D_j \beta_j)' R^{-1} (w_i - m_{ij} D_j \beta_j) \right] \exp \left( -\frac{1}{2} \beta_j' G^{-1} \beta_j \right), \\ &\propto \exp \left\{ -\frac{1}{2} \left[ \beta_j' \left( D_j' R^{-1} D_j \sum_{i=1}^n m_{ij}^2 + G^{-1} \right) \beta_j - 2 \sum_{i=1}^n w_i' m_{ij} R^{-1} D_j \beta_j \right] \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \left[ \beta_j' C_j \beta_j - 2 r_j' \beta_j \right] \right\} \\ &\propto \exp \left\{ -\frac{1}{2} (\beta_j' - r_j' C_j^{-1}) C_j (\beta_j - C_j^{-1} r_j) \right\} \\ &\propto N(C_j^{-1} r_j, C_j^{-1}) \end{aligned}$$



## Gibbs sampler algorithm for multi-trait BayesB

### Single-site Gibbs sampler for multi-trait BayesB

For convenience, from now on let “1” denote trait  $k$  and “2” the other traits. Thus,  $\beta_j$  can be denoted as  $\begin{bmatrix} \beta_{j1} \\ \beta_{j2} \end{bmatrix}$  and  $D_j$  can be denoted as  $\begin{bmatrix} \delta_{j1} & 0 \\ 0 & D_{j2} \end{bmatrix}$ . The Gibbs sampler for  $\beta_{jk}$  and  $\delta_{jk}$  is derived as below. In our sampling scheme,  $\beta_{j1}$  and  $\delta_{j1}$  are sampled from their joint full conditional distributions, which can be written as the product of the full conditional distribution of  $\beta_{j1}$  given  $\delta_{j1}$  and the marginal full conditional distribution of  $\delta_{j1}$ . Let  $\theta$  denote all other parameters except  $\delta_{j1}$  and  $\beta_{j1}$ , then our sampling scheme can be written as

$$f(\beta_{j1}, \delta_{j1} | \theta, \mathbf{y}) = f(\beta_{j1} | \delta_{j1}, \theta, \mathbf{y}) f(\delta_{j1} | \theta, \mathbf{y}).$$

The full conditional distribution of  $\beta_j$  can be written as

$$\begin{aligned} f(\beta_{j1} | \delta_{j1}, \beta_{-j1}, D_{-j1}, G_j, G_{-j}, R, \mathbf{y}) &\propto f(\mathbf{y} | \mu, \beta, D, G_j, G_{-j}, R) f(\beta_{j1}, \beta_{j2} | G_j) \\ &\propto \exp \left[ -\frac{1}{2} \sum_{i=1}^n (\mathbf{w}_i - m_{ij} D_j \beta_j)' R^{-1} (\mathbf{w}_i - m_{ij} D_j \beta_j) \right] \exp \left( -\frac{1}{2} \beta_j' G_j^{-1} \beta_j \right), \end{aligned}$$

where  $\mathbf{w}_i = \mathbf{y}_i - \mu_i - \sum_{j' \neq j} m_{ij'} D_{j'} \beta_{j'}$ . Further, by dropping factors that do not involve  $\beta_{j1}$ ,

$$\begin{aligned} f(\beta_{j1} | \delta_{j1}, \beta_{-j1}, D_{-j1}, G_j, G_{-j}, R, \mathbf{y}) &\propto \exp \left\{ -\frac{1}{2} \left[ \beta_j' \left( D_j' R^{-1} D_j \sum_{i=1}^n m_{ij}^2 + G_j^{-1} \right) \beta_j - 2 \sum_{i=1}^n \mathbf{w}_i' m_{ij} R^{-1} D_j \beta_j \right] \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \left[ \beta_j' C_j \beta_j - 2 \mathbf{r}_j' \beta_j \right] \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \left[ \begin{bmatrix} \beta_{j1} & \beta_{j2} \end{bmatrix} \begin{bmatrix} C_{j,11} & C_{j,12} \\ C_{j,21} & C_{j,22} \end{bmatrix} \begin{bmatrix} \beta_{j1} \\ \beta_{j2} \end{bmatrix} - 2 \begin{bmatrix} r_{j1} & \mathbf{r}_{j2}' \end{bmatrix} \begin{bmatrix} \beta_{j1} \\ \beta_{j2} \end{bmatrix} \right] \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \left( C_{j,11} \beta_{j1}^2 + (2C_{j,12} \beta_{j2} - 2r_{j1}) \beta_{j1} \right) \right\} \\ &\propto \exp \left\{ -\frac{C_{j,11}}{2} \left( \beta_{j1} + (C_{j,12} \beta_{j2} - r_{j1}) C_{j,11}^{-1} \right)^2 \right\} \\ &\propto N \left( C_{j,11}^{-1} (r_{j1} - C_{j,12} \beta_{j2}), C_{j,11}^{-1} \right) \\ &\propto N \left( \hat{\beta}_{j1}, C_{j,11}^{-1} \right) \end{aligned}$$

where  $C_j = D_j' R^{-1} D_j \sum_{i=1}^n m_{ij}^2 + G_j^{-1}$  and  $\mathbf{r}_j' = \left( \sum_{i=1}^n \mathbf{w}_i' m_{ij} \right) R^{-1} D_j$ .

Note that when  $\delta_{j1} = 0$ ,

$$\begin{aligned} C_j &= \begin{bmatrix} G_j^{11} & G_j^{12} \\ G_j^{21} & G_j^{22} + D_{j2}' R^{22} D_{j2} \sum_{i=1}^n m_{ij}^2 \end{bmatrix} \\ \mathbf{r}_j' &= \begin{bmatrix} 0 & \left( \sum_{i=1}^n \mathbf{w}_i' m_{ij} \right) \begin{bmatrix} R^{12} \\ R^{22} \end{bmatrix} D_{j2} \end{bmatrix} \end{aligned}$$

When  $\delta_{j1} = 1$ ,

$$\begin{aligned} C_j &= \begin{bmatrix} C_{j,11}^1 & C_{j,12}^1 \\ C_{j,21}^1 & C_{j,22}^1 \end{bmatrix} \\ &= \begin{bmatrix} G_j^{11} + R^{11} \sum_{i=1}^n m_{ij}^2 & G_j^{12} + R^{12} D_{j2} \sum_{i=1}^n m_{ij}^2 \\ G_j^{21} + D_{j2}' R^{21} \sum_{i=1}^n m_{ij}^2 & G_j^{22} + D_{j2}' R^{22} D_{j2} \sum_{i=1}^n m_{ij}^2 \end{bmatrix} \\ \mathbf{r}_j' &= \begin{bmatrix} r_{j1}^1 & \mathbf{r}_{j2}^{1'} \end{bmatrix} \\ &= \begin{bmatrix} \left( \sum_{i=1}^n \mathbf{w}_i' m_{ij} \right) \begin{bmatrix} R^{11} \\ R^{21} \end{bmatrix} & \left( \sum_{i=1}^n \mathbf{w}_i' m_{ij} \right) \begin{bmatrix} R^{12} \\ R^{22} \end{bmatrix} D_{j2} \end{bmatrix} \end{aligned}$$

Thus when  $\delta_{j1} = 0$ , the full conditional distribution of  $\beta_{j1}$  is

$$f(\beta_{j1} | \delta_{j1} = 0, \beta_{-j1}, D_{-j1}, G_j, G_{-j}, \mathbf{R}, \mathbf{y}) \propto N\left(-\left(G_j^{11}\right)^{-1} \mathbf{G}_j^{12} \beta_{j2}, \left(G_j^{11}\right)^{-1}\right).$$

When  $\delta_{j1} = 1$ , the full conditional distribution of  $\beta_{j1}$  becomes

$$f(\beta_{j1} | \delta_{j1} = 1, \beta_{-j1}, D_{-j1}, G_j, G_{-j}, \mathbf{R}, \mathbf{y}) \propto N\left(C_{j,11}^{-1}\left(r_{j1} - \mathbf{C}_{j,12}^1 \beta_{j2}\right), C_{j,11}^{-1}\right).$$

The marginal full conditional distribution of  $\delta_{j1}$  can be written as

$$\begin{aligned} f(\delta_{j1} = 1 | \boldsymbol{\theta}, \mathbf{y}) &= \frac{f(\delta_{j1}, \boldsymbol{\theta}, \mathbf{y})}{\sum_{\delta_{j1} \in (0,1)} f(\delta_{j1}, \boldsymbol{\theta}, \mathbf{y})} \\ &= \frac{f(\mathbf{y} | \delta_{j1} = 1, \boldsymbol{\theta}) f(\delta_{j1} = 1, \delta_{j2} | \boldsymbol{\Pi})}{\sum_{\delta_{j1} \in (0,1)} f(\mathbf{y} | \delta_{j1}, \boldsymbol{\theta}) f(\delta_j | \boldsymbol{\Pi})} \\ &= \left\{ 1 + \frac{f(\mathbf{y} | \delta_{j1} = 0, \boldsymbol{\theta}) f(\delta_{j1} = 0, \delta_{j2} | \boldsymbol{\Pi})}{f(\mathbf{y} | \delta_{j1} = 1, \boldsymbol{\theta}) f(\delta_{j1} = 1, \delta_{j2} | \boldsymbol{\Pi})} \right\}^{-1} \end{aligned}$$

The factor  $f(\mathbf{y} | \delta_{j1}, \boldsymbol{\theta})$  can be written as

$$\begin{aligned} f(\mathbf{y} | \delta_{j1}, \boldsymbol{\theta}) &\propto \int f(\mathbf{y} | \boldsymbol{\mu}, \beta_{j1}, \beta_{-j1}, D, \mathbf{G}, \mathbf{R}) f(\beta_{j1}, \beta_{j2} | \mathbf{G}_j) d\beta_{j1} \\ &\propto \int \exp\left[-\frac{1}{2} \sum_{i=1}^n (\mathbf{w}_i - m_{ij} D_j \beta_j)' R^{-1} (\mathbf{w}_i - m_{ij} D_j \beta_j)\right] \exp\left(-\frac{1}{2} \beta_j' \mathbf{G}_j^{-1} \beta_j\right) d\beta_{j1} \\ &\propto \exp\left\{-\frac{1}{2} \left(\sum_i \mathbf{w}_i' R^{-1} \mathbf{w}_i - 2r_{j2}' \beta_{j2} + \beta_{j2}' C_{j,22} \beta_{j2} - (r_{j1} - C_{j,12} \beta_{j2})^2 C_{j,11}^{-1}\right)\right\} \\ &\times \int \exp\left[-\frac{1}{2} (\beta_{j1} - \hat{\beta}_{j1})^2 C_{j,11}\right] d\beta_{j1} \\ &\propto (C_{j,11})^{-\frac{1}{2}} \exp\left\{-\frac{1}{2} \left(\sum_i \mathbf{w}_i' R^{-1} \mathbf{w}_i - 2r_{j2}' \beta_{j2} + \beta_{j2}' C_{j,22} \beta_{j2} - (r_{j1} - C_{j,12} \beta_{j2})^2 C_{j,11}^{-1}\right)\right\} \\ &\propto (C_{j,11})^{-\frac{1}{2}} \exp\left\{-\frac{1}{2} \left(\sum_i \mathbf{w}_i' R^{-1} \mathbf{w}_i - 2r_{j2}' \beta_{j2} + \beta_{j2}' C_{j,22} \beta_{j2} - \hat{\beta}_{j1}^2 C_{j,11}\right)\right\}. \end{aligned}$$

Note that  $\sum_i \mathbf{w}_i' R^{-1} \mathbf{w}_i, r_{j2}' \beta_{j2}, \beta_{j2}' C_{j,22} \beta_{j2}$  are same when  $\delta_{j1} = 0$  or 1. Thus the ratio  $\frac{f(\mathbf{y} | \delta_{j1}=1, \boldsymbol{\theta})}{f(\mathbf{y} | \delta_{j1}=0, \boldsymbol{\theta})}$  becomes

$$\begin{aligned} H &= (C_{j,11})^{-\frac{1}{2}} (G_j^{11})^{\frac{1}{2}} \exp\left(-\frac{1}{2} (\hat{\beta}_{j1}^2 G_j^{11} - \hat{\beta}_{j1}^2 C_{j,11}^1)\right) \\ &= \exp\left\{-\frac{1}{2} \left(\log C_{j,11} - \hat{\beta}_{j1}^2 C_{j,11}^1\right) - \left(-\frac{1}{2} (\log G_j^{11} - \hat{\beta}_{j1}^2 G_j^{11})\right)\right\} \end{aligned}$$

Thus the conditional probability of  $\delta_{j1} = 1$  is

$$\left\{ 1 + \frac{f(\mathbf{y} | \delta_{j1} = 0, \boldsymbol{\theta}) f(\delta_{j1} = 0, \delta_{j2} | \Pi_1, \Pi_{2...})}{f(\mathbf{y} | \delta_{j1} = 1, \boldsymbol{\theta}) f(\delta_{j1} = 1, \delta_{j2} | \Pi_1, \Pi_{2...})} \right\}^{-1} = \left\{ 1 + \left( \frac{\Pi_{j0}}{\Pi_{j1}} H \right)^{-1} \right\}^{-1},$$

where  $\Pi_{j0} = Pr(\delta_{j1} = 0, \delta_{j2} | \boldsymbol{\Pi})$  and  $\Pi_{j1} = Pr(\delta_{j1} = 1, \delta_{j2} | \boldsymbol{\Pi})$ .

### 329 **Joint Gibbs sampler for multi-trait BayesB**

330 Let  $\theta$  denote all other parameters except  $\beta_j$  and  $\delta_j$ , then our sampling scheme can be written as

$$f(\beta_j, \delta_j | \theta, y) = f(\delta_j | \theta, y) f(\beta_j | \delta_j, \theta, y)$$

331 The marginal full conditional distribution of  $\delta_j$  can be written as

$$\begin{aligned} f(\delta_j | \theta, y) &= \frac{f(\delta_j, \theta, y)}{\sum_{\delta_j} f(\delta_j, \theta, y)} \\ &= \frac{f(y | \delta_j, \theta) f(\delta_j | \Pi)}{\sum_{\delta_j} f(y | \delta_j, \theta) f(\delta_j | \Pi)}. \end{aligned}$$

332 Denote  $w_i = y_i - \mu_i - \sum_{j' \neq j} m_{ij'} D_{j'} \beta_{j'}$ , then

$$\begin{aligned} f(y | \delta_j, \theta) &\propto \int f(y | \beta, D, R) f(\beta_j | G_j) d\beta_j \\ &\propto \int \exp \left[ -\frac{1}{2} \sum_{i=1}^n (w_i - m_{ij} D_j \beta_j)' R^{-1} (w_i - m_{ij} D_j \beta_j) \right] \exp \left( -\frac{1}{2} \beta_j' G_j^{-1} \beta_j \right) d\beta_j \\ &\propto \int \exp \left\{ -\frac{1}{2} \left[ \beta_j' \left( D_j' R^{-1} D_j \sum_{i=1}^n m_{ij}^2 + G_j^{-1} \right) \beta_j - 2 \sum_{i=1}^n w_i' m_{ij} R^{-1} D_j \beta_j + \sum_{i=1}^n w_i' R^{-1} w_i \right] \right\} d\beta_j \\ &\propto \int \exp \left\{ -\frac{1}{2} \left[ \beta_j' C_j \beta_j - 2 r_j' \beta_j + \sum_{i=1}^n w_i' R^{-1} w_i \right] \right\} d\beta_j \\ &\propto \int \exp \left\{ -\frac{1}{2} \left[ (\beta_j' - r_j' C_j^{-1}) C_j (\beta_j - C_j^{-1} r_j) + \sum_{i=1}^n w_i' R^{-1} w_i - r_j' C_j^{-1} r_j \right] \right\} d\beta_j \\ &\propto \exp \left\{ -\frac{1}{2} \left[ \sum_{i=1}^n w_i' R^{-1} w_i - r_j' C_j^{-1} r_j \right] \right\} \\ &\propto |C_j^{-1}|^{\frac{1}{2}} \int |C_j^{-1}|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (\beta_j' - r_j' C_j^{-1}) C_j (\beta_j - C_j^{-1} r_j) \right] d\beta_j \\ &\propto |C_j^{-1}|^{\frac{1}{2}} \exp \left\{ -\frac{1}{2} \left[ \sum_{i=1}^n w_i' R^{-1} w_i - r_j' C_j^{-1} r_j \right] \right\}, \end{aligned}$$

333 where  $C_j = D_j' R^{-1} D_j \sum_{i=1}^n m_{ij}^2 + G_j^{-1}$  and  $r_j' = \left( \sum_{i=1}^n w_i' m_{ij} \right) R^{-1} D_j$ .

Note that  $\sum_i w_i' R^{-1} w_i$  is same for different  $\delta_j$ . Thus the marginal full conditional distribution of  $\delta_j$  can be written as

$$f(\delta_j | \theta, y) = \frac{f(y | \delta_j, \theta) f(\delta_j | \Pi)}{\sum_{\delta_j} f(y | \delta_j, \theta) f(\delta_j | \Pi)},$$

where

$$f(y | \delta_j, \theta) \propto |C_j^{-1}|^{\frac{1}{2}} \exp \left\{ \frac{1}{2} r_j' C_j^{-1} r_j \right\}.$$

334 The full conditional distribution of  $\beta_j$  is

$$\begin{aligned} f(\beta_j | \delta_j, \theta, y) &\propto \exp \left[ -\frac{1}{2} \sum_{i=1}^n (w_i - m_{ij} D_j \beta_j)' R^{-1} (w_i - m_{ij} D_j \beta_j) \right] \exp \left( -\frac{1}{2} \beta_j' G_j^{-1} \beta_j \right), \\ &\propto \exp \left\{ -\frac{1}{2} \left[ \beta_j' \left( D_j' R^{-1} D_j \sum_{i=1}^n m_{ij}^2 + G_j^{-1} \right) \beta_j - 2 \sum_{i=1}^n w_i' m_{ij} R^{-1} D_j \beta_j \right] \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \left[ \beta_j' C_j \beta_j - 2 r_j' \beta_j \right] \right\} \\ &\propto \exp \left\{ -\frac{1}{2} (\beta_j' - r_j' C_j^{-1}) C_j (\beta_j - C_j^{-1} r_j) \right\} \\ &\propto N(C_j^{-1} r_j, C_j^{-1}) \end{aligned}$$