

# Genomic Prediction from Multiple-trait Bayesian Regression Methods using Mixture Priors

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, University of California Davis, California, <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>School of Agriculture, Massey University, New Zealand, <sup>\*\*</sup>Department of Animal Science, Iowa State University, Ames, Iowa

**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 and real data. These methods have been extended to multi-trait analyses, but only under the restrictive assumption that a locus simultaneously affects all the traits or none of them. This assumption is not biologically meaningful, especially in multi-trait analyses involving many traits. In this paper, we develop and implement a more general multi-trait BayesCII and BayesB methods allowing a broader range of mixture priors. Our methods allow a locus to affect any combination of traits, e.g., in a 5-trait analysis, the "restrictive" model only allows two situations, whereas ours allow all 32 situations. Further, we compare our methods to single-trait methods and the "restrictive" multi-trait formulation using real and simulated data. In the real data analysis, higher prediction accuracies were observed from both our new broad-based multi-trait methods and the "restrictive" formulation. The broad-based and restrictive multi-trait methods showed similar prediction accuracies. In the simulated data analysis, higher prediction accuracies to the "restrictive" method were observed from our general multi-trait methods for intermediate training population size. The software tool JWAS offers open-source routines to perform these analyses.

**KEYWORDS** multi-trait; mixture priors; genomic prediction; Bayesian regression; pleiotropy

## Introduction

Genomic prediction was proposed by [Meuwissen et al. \(2001\)](#) to incorporate marker effects from 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, including various extensions to the BayesB method of [Meuwissen et al. \(2001\)](#). BayesB accommodates models 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  $1 - \pi$  ([Meuwissen et al. 2001](#); [Gianola et al. 2009](#); [Cheng et al. 2015b](#)). Another model, BayesC, assumes a mixture with a point mass at zero with probability  $\pi$  and a univariate normal distribution with probability  $1 - \pi$  for all marker effects, 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 "restrictive" multi-trait BayesCII presented by [Jia and Jannink \(2012\)](#) assumes any particular locus affects none of the traits or simultaneously affects all traits. This assumption of genetic architecture in that multi-trait BayesCII model is violated if some loci have no effect on at least one of the traits while having an effect on the remaining traits.

In this paper, we proposed a more general class of multi-trait

48 BayesCII and BayesB methods, where each locus can have an  
 49 effect on any combination of traits. For example, in a 5-trait  
 50 analysis, the restricted model only allows 2 situations, whereas  
 51 ours allows all 32 situations. The previous restrictive multi-trait  
 52 models are special cases of this general class of models. Further,  
 53 our model allows the use of a single-site Gibbs sampler that  
 54 requires less computing effort than some alternative Markov  
 55 chain Monte Carlo approaches, especially for analyses involving  
 56 many traits. Methodologies for the new models are compared to  
 57 single-trait methods and the previous multi-trait methods using  
 58 real and simulated data.

## 59 Materials and Methods

### 60 Multi-trait Marker Effects Model

For simplicity of our description, but without loss of general-  
 ity, we will assume individuals have all traits measured with a  
 general mean as the only fixed effect, and write the multi-trait  
 model for individual  $i$  from  $n$  genotyped individuals as

$$\mathbf{y}_i = \boldsymbol{\mu} + \sum_{j=1}^p m_{ij} \boldsymbol{\alpha}_j + \mathbf{e}_i,$$

61 where  $\mathbf{y}_i$  is a vector of phenotypes of  $t$  traits for individual  $i$ ,  $\boldsymbol{\mu}$   
 62 is a vector of overall means for  $t$  traits,  $m_{ij}$  is the genotype covariate  
 63 at locus  $j$  for individual  $i$  (coded as 0,1,2),  $p$  is the number of  
 64 genotyped loci,  $\boldsymbol{\alpha}_j$  is a vector of allele substitution effects or  
 65 marker effects of  $t$  traits for locus  $j$ , and  $\mathbf{e}_i$  is a vector of random  
 66 residuals of  $t$  traits for individual  $i$ . The fixed effects, or general  
 67 mean in this case, are assigned flat priors. The residuals,  $\mathbf{e}_i$ , are *a*  
 68 *priori* assumed to be independently and identically distributed  
 69 multivariate normal vectors with null mean and covariance  
 70 matrix  $\mathbf{R}$ , which in turn is assumed to have an inverse Wishart  
 71 prior distribution,  $W_t^{-1}(\mathbf{S}_e, \nu_e)$ .

72 We will show that, employing the concept of data augmenta-  
 73 tion, the vector of marker effects at a particular locus  $\boldsymbol{\alpha}_j$  can be  
 74 written as  $\boldsymbol{\alpha}_j = \mathbf{D}_j \boldsymbol{\beta}_j$ , where  $\mathbf{D}_j$  is a diagonal matrix whose  $k$ th  
 75 diagonal entry is an indicator variable indicating whether the  
 76 marker effect of locus  $j$  for trait  $k$  is zero or non-zero, and  $\boldsymbol{\beta}_j$  fol-  
 77 lows a multivariate normal distribution in multi-trait BayesCII  
 78 or a multivariate  $t$  distribution in multi-trait BayesB.

### 79 Multi-trait BayesCII 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

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

The vector of marker effects at a particular locus  $\boldsymbol{\alpha}_j$  is written  
 as  $\boldsymbol{\alpha}_j = \mathbf{D}_j \boldsymbol{\beta}_j$ , where  $\mathbf{D}_j$  is a diagonal matrix with elements  
 $\text{diag}(\mathbf{D}_j) = \boldsymbol{\delta}_j = (\delta_{j1}, \delta_{j2}, \delta_{j3}, \dots, \delta_{jt})$ , where  $\delta_{jk}$  is an indicator  
 variable indicating whether the marker effect of locus  $j$  for trait

$k$  is zero or non-zero, and the vector  $\boldsymbol{\beta}_j$  follows a multivariate  
 normal distribution with null mean and covariance matrix  $\mathbf{G} =$   

$$\begin{bmatrix} \sigma_1^2 & \cdots & \sigma_{1t} \\ \vdots & \ddots & \vdots \\ \sigma_{1t} & \cdots & \sigma_t^2 \end{bmatrix}.$$
 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 BayesCII model with data augmentation is written  
 as

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

In the most general case, any marker effect might be zero for any  
 possible combination of  $t$  traits resulting in  $2^t$  possible combina-  
 tions of  $\boldsymbol{\delta}_j$ . For example, in a  $t=2$  trait model, there are  $2^2 = 4$   
 combinations for  $\boldsymbol{\delta}_j$ : (0, 0), (0, 1), (1, 0), (1, 1). In the restrictive  
 special case of this model described by Jia and Jannink (2012),  
 only two combinations, i.e., (0, 0) and (1, 1), have non-zero  
 probability. Suppose in general we use numerical labels "1",  
 "2", ..., "l" for the  $2^t$  possible outcomes for  $\boldsymbol{\delta}_j$ , then the prior for  
 $\boldsymbol{\delta}_j$  is a categorical distribution

$$p(\boldsymbol{\delta}_j = "i") \\ = \Pi_1 I(\boldsymbol{\delta}_j = "1") + \Pi_2 I(\boldsymbol{\delta}_j = "2") + \dots + \Pi_l I(\boldsymbol{\delta}_j = "l"),$$

80 where  $\sum_{i=1}^l \Pi_i = 1$  with  $\Pi_i$  being the prior probability that the  
 81 vector  $\boldsymbol{\delta}_j$  corresponds to the vector labelled "i". A Dirichlet  
 82 distribution with all parameters equal to one, i.e., a uniform  
 83 distribution, can be used for a prior for  $\boldsymbol{\Pi} = (\Pi_1, \Pi_2, \dots, \Pi_l)$ .

84 As shown below, we consider two Gibbs samplers to draw  
 85 samples for all the parameters in this model. Gibbs sampler I is  
 86 a single-site sampler, where only one of the  $t$  indicator labels is  
 87 sampled at a time. Thus, in a 2-trait model, for example, this sam-  
 88 pler cannot move from (0, 0) to (1, 1) in a single step without  
 89 stepping through (1, 0) or (0, 1) for  $\boldsymbol{\delta}_j$ . Therefore, Gibbs sampler  
 90 I cannot be used for the restrictive model which excludes (1, 0)  
 91 and (0, 1) from the state space for  $\boldsymbol{\delta}_j$ . Gibbs sampler II, however,  
 92 samples all elements of  $\boldsymbol{\delta}_j$  jointly, and can move from (0, 0) to  
 93 (1, 1) in a single step. However, Gibbs sampler II is compu-  
 94 tationally more intensive because it requires drawing samples  
 95 from a multivariate normal distribution of order  $t$ , the number  
 96 of traits.

**Gibbs sampler I for multi-trait BayesCII** Suppose the prior for  
 $\boldsymbol{\delta}_j$  is a categorical distribution for which the support is the set  
 of  $2^t$  outcomes of  $\boldsymbol{\delta}_j$ . 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  $\beta_{j1}$ ,  $\delta_{j1}$ ,  $\boldsymbol{\Pi}$ ,  $\mathbf{G}$  and  $\mathbf{R}$  for  
 Gibbs sampler I, whose derivations are in the Appendix, are

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

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

with

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

where  $\mathbf{w}_i = \mathbf{y}_i - \boldsymbol{\mu} - \sum_{j' \neq j} m_{ij'} \mathbf{D}_{j'} \boldsymbol{\beta}_{j'}$ ,  $\mathbf{G}^{11}$  and  $\mathbf{G}^{12}$  are the partitions of  $\mathbf{G}^{-1}$  corresponding to trait  $k$  and covariances between trait  $k$  and other traits, respectively.  $\mathbf{R}^{11}$  and  $\mathbf{R}^{12}$  are the partitions of  $\mathbf{R}^{-1}$  corresponding to trait  $k$  and covariances between trait  $k$  and other traits, respectively.

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

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

where  $H =$

$$\exp \left\{ -\frac{1}{2} \left( \log \mathbf{C}_{j,11}^1 - \hat{\beta}_{j1}^1{}^2 \mathbf{C}_{j,11}^1 \right) - \left( -\frac{1}{2} \left( \log \mathbf{G}^{11} - \hat{\beta}_{j1}^0{}^2 \mathbf{G}^{11} \right) \right) \right\}$$

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

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

where  $n_i$  is the number of loci or markers for which  $\delta_j = "i"$ .

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 whose  $i$ th row is  $e'_i$ . The full conditional distribution for  $\mathbf{G}$ , the covariance matrix for  $\boldsymbol{\beta}_j$ , is an inverse Wishart distribution,  $W_t^{-1}(\mathbf{S}_\beta + \boldsymbol{\beta}'\boldsymbol{\beta}, \nu_\beta + p)$ , where  $\boldsymbol{\beta}$  is the  $p \times t$  matrix whose  $i$ th row is  $\boldsymbol{\beta}'_i$ .

**Gibbs sampler II for multi-trait BayesCII** The Gibbs sampler above, where only one of the  $t$  indicator labels is sampled at a time, cannot be used for the restrictive model assuming any particular locus affects all traits or none of them. Further, if some particular  $\Pi_i$  are near zero, the chain might exhibit mixing problems. Another more general but computationally intensive Gibbs sampler that samples all elements of  $\delta_j$  jointly and may exhibit improved mixing is proposed below.

The full conditional distributions of  $\boldsymbol{\beta}_j$ ,  $\delta_j$ ,  $\boldsymbol{\Pi}$ ,  $\mathbf{G}$ ,  $\mathbf{R}$  for Gibbs sampler II, whose derivations are in the Appendix, are given below.

Let  $\boldsymbol{\theta}$  denote all other parameters except  $\boldsymbol{\beta}_j$  and  $\delta_j$ , then our sampling scheme can be written as

$$f(\boldsymbol{\beta}_j, \delta_j | \boldsymbol{\theta}, \mathbf{y}) = f(\delta_j | \boldsymbol{\theta}, \mathbf{y}) f(\boldsymbol{\beta}_j | \delta_j, \boldsymbol{\theta}, \mathbf{y}).$$

The full conditional distribution of  $\boldsymbol{\beta}_j$  is

$$f(\boldsymbol{\beta}_j | \delta_j, \boldsymbol{\theta}, \mathbf{y}) \propto N(\mathbf{C}_j^{-1} \mathbf{r}'_j, \mathbf{C}_j^{-1}),$$

where  $\mathbf{C}_j = \mathbf{D}'_j \mathbf{R}^{-1} \mathbf{D}_j \sum_{i=1}^n m_{ij}^2 + \mathbf{G}^{-1}$  and  $\mathbf{r}'_j = (\sum_{i=1}^n \mathbf{w}'_i m_{ij}) \mathbf{R}^{-1} \mathbf{D}_j$ .

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

$$\begin{aligned} f(\delta_j = "i" | \boldsymbol{\theta}, \mathbf{y}) \\ = \frac{f(\mathbf{y} | \delta_j = "i", \boldsymbol{\theta}) f(\delta_j = "i" | \boldsymbol{\Pi})}{\sum_{"i" \in \{"1", "2", \dots, "l"\}} f(\mathbf{y} | \delta_j = "i", \boldsymbol{\theta}) f(\delta_j = "i" | \boldsymbol{\Pi})}, \end{aligned}$$

where

$$f(\mathbf{y} | \delta_j, \boldsymbol{\theta}) = |\mathbf{C}_j^{-1}|^{\frac{1}{2}} \exp \left\{ \frac{1}{2} \mathbf{r}'_j \mathbf{C}_j^{-1} \mathbf{r}_j \right\}.$$

This Gibbs sampler can accommodate the "restrictive" multi-trait BayesCII that was proposed by 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 BayesCII model proposed above can be modified to accommodate the general 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  $\boldsymbol{\beta}_j$  is a multivariate t distribution, rather than a multivariate normal distribution. This is equivalent to assuming  $\boldsymbol{\beta}_j$  has a multivariate normal distribution with null mean and locus-specific covariance matrix  $\mathbf{G}_j$ , which is assigned an inverse Wishart prior,  $W_t^{-1}(\mathbf{S}_\beta, \nu_\beta)$ .

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  $\mathbf{G}_j$  are similar to the multi-trait BayesCII model. The full conditional distribution for  $\mathbf{G}_j$ , the covariance matrix for  $\boldsymbol{\beta}_j$ , is an inverse Wishart distribution,  $W_t^{-1}(\mathbf{S}_\beta + \boldsymbol{\beta}_j \boldsymbol{\beta}'_j, \nu_\beta + 1)$ .

### Data analyses

**Real data** Published genotypic and deregressed breeding values based on phenotypic data for Loblolly Pine (*Pinus Taeda* L.) were used (Resende et al. 2012; Daetwyler et al. 2013). Two disease traits, namely presence or absence of rust (Rust\_bin) and gall volume (Rust\_gall\_vol) were analyzed. These are the two traits used in Jia and Jannink (2012). The reported heritabilities were 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 but loci with a missing rate >50% were excluded. After these quality control edits, 4,828 SNPs on 807 individuals with deregressed phenotypes and genotypes on both traits remained.

159 Prediction accuracy was calculated as the correlation between  
 160 the vector of deregressed phenotypes and the vector of estimated  
 161 breeding values. Cross-validation using 10 folds formed the  
 162 basis for comparing these methods. Paired t tests were used  
 163 for tests of significance of difference in prediction accuracies  
 164 between two methods, where prediction accuracies for two dif-  
 165 ferent methods from each validation fold were considered as  
 166 paired samples. The general multi-trait BayesCII model (MT-  
 167 BayesCII-G) was compared to a similar model where the prior  
 168 for  $\beta_j$  is a multivariate normal rather than a mixture of multi-  
 169 variate normals (MT-BayesC0), the more restricted multi-trait  
 170 BayesCII proposed by Jia and Jannink (2012) (MT-BayesCII-R)  
 171 and the usual single trait formulations of the mixture models  
 172 (ST-BayesC0, ST-BayesC $\pi$ ). Since BayesC0 is equivalent to ran-  
 173 dom regression best linear unbiased prediction (RR-BLUP), ST-  
 174 BayesC0 and MT-BayesC0 are denoted as ST-RR-BLUP and MT-  
 175 RR-BLUP below. The prior for the residual covariance matrix  
 176  $\mathbf{R}$  in all multi-trait methods was an inverse Wishart distribu-  
 177 tion,  $\mathbf{W}^{-1} = \left( \begin{bmatrix} 0.003 & 0 \\ 0 & 0.003 \end{bmatrix}, 6 \right)$ , for which the mean of  $\mathbf{R}$  is  
 178  $\begin{bmatrix} 0.001 & 0 \\ 0 & 0.001 \end{bmatrix}$ , the standard deviation of diagonal elements are  
 179  $1.4 \times 10^{-3}$ , and the standard deviation of off-diagonal elements  
 180 are 0. This same prior was used for the marker effects covari-  
 181 ance matrix  $\mathbf{G}$ . The priors for the residual variance and marker  
 182 effects variance in single-trait analyses were scaled inverted chi-  
 183 squared distributions with scale parameter  $S^2 = 0.0005$  and  
 184 degrees of freedom  $\nu = 4$ , for which the mean of the prior was  
 185 also 0.001. In the data analyses, multi-trait BayesB methods  
 186 provided similar results as multi-trait BayesCII methods. Thus,  
 187 only results from BayesCII analyses were presented below to  
 188 demonstrate the superiority of our multi-trait methods.

189 **Simulated data** Simulated data described below were used to  
 190 quantify the superiority of the general multi-trait Bayesian meth-  
 191 ods. Two scenarios were simulated. In scenario 1, as a known  
 192 ideal condition, the simulated genome consisted of 100 loci on  
 193 each of 2 chromosomes that were in Hardy-Weinberg and link-  
 194 age equilibria. All these loci were considered as QTL or causative  
 195 variants and used in the analyses. The QTL on the first chro-  
 196 some had effects only on trait 1 and those on the second  
 197 chromosome only on trait 2. The effects of these QTL were simu-  
 198 lated from a standard normal distribution and then were equally  
 199 scaled to provide unit genetic variance for each trait in the simu-  
 200 lated population of 8,000 unrelated individuals. The phenotypes  
 201 for these traits were obtained by adding independent residuals  
 202 to the genetic values. Two situations were simulated: 1) heri-  
 203 tabilities for both traits were 0.5; 2) heritability for trait 1 was  
 204 0.2 and for trait 2 was 0.8. The XSim package was used in the  
 205 simulation (Cheng et al. 2015a).

206 In scenario 2, both markers and QTL were simulated. The  
 207 simulated genome consisted of 100 evenly spaced loci on each of  
 208 3 chromosomes of length 10 cM. Ten loci were randomly selected  
 209 on each chromosome as QTL. Allele states were sampled from a  
 210 Bernoulli distribution with frequency 0.5 in the base population.  
 211 Starting from a base population of 500 males and 500 females,  
 212 random mating was simulated for 500 generations to generate  
 213 linkage disequilibrium. Random mating was continued for five  
 214 more generations to increase the population size to 4000 males  
 215 and 4000 females, which were used in the analyses. The effects of  
 216 QTL on the first two chromosomes were simulated following the

217 same strategy in scenario 1, i.e., the QTL on the first chromosome  
 218 had effects only on trait 1 and those on the second chromosome  
 219 only on trait 2. All QTL on the third chromosome had effects on  
 220 both traits. The effects of these QTL on the third chromosome  
 221 were simulated from a standard bivariate normal distribution  
 222 with correlation 0.5. The phenotypes for these traits were ob-  
 223 tained by adding independent residuals to the genetic values. In  
 224 total, 8,000 individuals were simulated with heritability 0.2 for  
 225 trait 1 and 0.8 for trait 2.

226 The same validation approaches were used for these two  
 227 simulation scenarios. A total of 500 individuals were used for  
 228 testing, and for each training population of size  $N$ , 100 replicates  
 229 of the training population were sampled from the remaining  
 230 individuals. The values considered for  $N$  were 50, 100, 200, 400,  
 231 1000, 2000, 4000 or 7000. The true genetic and residual vari-  
 232 ances were used to compute the scale parameters for the priors of the  
 233 variance components. The general multi-trait BayesCII model  
 234 (MT-BayesCII-G) was compared to the more restricted multi-  
 235 trait BayesCII (MT-BayesCII-R) using this dataset.

236 All analyses were performed using JWAS (Cheng et al.  
 237 2018), an open-source, publicly-available package for single-trait  
 238 and multi-trait whole-genome analyses written in the freely-  
 239 available Julia language.

#### 240 Data availability

241 The genotypic and phenotypic data used in the real data analysis  
 242 are publicly available (Resende et al. 2012). The scripts used to  
 243 generate the simulated data are provided as supplementary  
 244 information.

#### 245 Results

246 **Real data** The prediction accuracies from all methods for  
 247 Rust\_bin and Rust\_gall\_vol are in figure 1. The prediction accu-  
 248 racies from all single-trait analyses using JWAS are similar to  
 249 those in (Resende et al. 2012).

250 The predictions of Rust\_bin exhibited no significant differ-  
 251 ence in accuracy between multi-trait and single-trait analy-  
 252 ses within each method (ST-RR-BLUP versus MT-RR-BLUP;  
 253 ST-BayesC $\pi$  versus MT-BayesCII-R; ST-BayesC $\pi$  versus MT-  
 254 BayesCII-G).

255 In contrast, prediction accuracies for the lower heritability  
 256 Rust\_gall\_vol with MT-BayesCII-G were significantly higher  
 257 than those from ST-BayesC $\pi$ . MT-BayesCII-G and MT-  
 258 BayesCII-R showed similar prediction accuracies. The posterior  
 259 means of  $\Pi$  for both methods are in table 1. When RR-BLUP was  
 260 used for the analysis, however, the advantage of the multiple-  
 261 trait analysis (MT-RR-BLUP) over the single-trait analysis (ST-  
 262 RR-BLUP) for Rust\_gall\_vol was not observed.

263 **Simulated data** The prediction accuracies from MT-BayesCII-G  
 264 and MT-BayesCII-R methods were compared for varying size  
 265 ( $N$ ) of training populations under two simulation scenarios. In  
 266 simulation scenario 1, figure 2 shows the prediction accuracies  
 267 where heritabilities for both traits were 0.5. Figure 3 shows the  
 268 prediction accuracies where heritabilities for trait 1 and trait  
 269 2 were 0.2 and 0.8. When  $N = 50$ , both methods had similar  
 270 prediction accuracy. For both traits, as  $N$  increased, initially, MT-  
 271 BayesCII-G became superior to MT-BayesCII-R, but as expected,  
 272 the accuracies of these methods asymptotically converged (Kara-  
 273 man et al. 2016). In most cases, the differences in accuracies for  
 274 both traits were small. However, in figure 3, the differences in  
 275 accuracies for trait 1, for which the heritability was 0.2, were



substantial for intermediate values of  $N$ . Figure 4 shows the prediction accuracies for simulation scenario 2. The pattern observed is similar to figure 3 under simulation scenario 1. MT-BayesCPII-G was superior to MT-BayesCPII-R for intermediate training population, but as  $N$  increased, the accuracies of these methods asymptotically converged (Karaman *et al.* 2016).

## Discussion

**Real data** Significant differences between multi-trait and single-trait analyses were only observed for Rust\_gall\_vol within BayesC $\pi$  methods (MT-BayesCPII-G versus ST-BayesC $\pi$ ; MT-BayesCPII-R versus ST-BayesC $\pi$ ). MT-BayesCPII-G and MT-BayesCPII-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 benefited from information on the other correlated trait Rust\_bin. Thus higher prediction accuracy from MT-BayesCPII-G were observed in trait Rust\_gall\_vol but not for the high heritability Rust\_bin. Results in (Jia and Jannink 2012) showed no difference between MT-BayesCPII and ST-BayesC $\pi$  because a reduced marker panel (500 markers) was used.

The fact that RR-BLUP showed no improvement in multi-trait analyses suggested that benefits from MT-BayesCPII-G may be due to the estimation of the hyper-parameter  $\Pi$ . In the MT-BayesCPII-G, 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-BayesCPII-G shown above. Thus a ST-BayesC analysis with  $\pi = 0.97$  was undertaken. Prediction accuracy from this ST-BayesC $\pi$  analysis with  $\pi = 0.97$  was 0.361, which was similar to the accuracy from MT-BayesCPII-G. This shows that including an additional correlated trait, especially one with high heritability, will bring in more data into the analysis, helping variable selection in a low-heritability trait to become more effective and result in improved prediction accuracy.

The difference between MT-BayesCPII-G and MT-BayesCPII-R is that MT-BayesCPII-R assumes a locus has an effect on all traits or none of them. This assumption regarding genetic architecture is likely to be seldom true. MT-BayesCPII-G and MT-BayesCPII-R, however, showed similar prediction accuracies. This can be explained by the estimation of  $\Pi$  in MT-BayesCPII-G and MT-BayesCPII-R in table 1. The posterior probability means for (0,1) and (1,0) were almost zero in MT-BayesCPII-G and for (0,0) and (1,1) are similar in MT-BayesCPII-G and MT-BayesCPII-R, suggesting that the assumption of genetic architecture whereby the same loci affect both traits as explicit in MT-BayesCPII-R may be valid for these two disease traits. Note that the lack of difference between the methods may also result from the limited size of the training population.

**Simulated data** In scenario 1, we simulated bivariate data where each QTL had an effect on only one or the other of the traits. In MT-BayesCPII-R, if a locus has an effect on one of the traits, that locus is included in the model for all traits. So in the simulated data, MT-BayesCPII-R would need to include all loci in the model for both traits. Thus for the trait that had heritability 0.2, the contribution of noise to the prediction from loci on chromosome 2, which had no effect on this trait, is large relative to the real signal from QTL on chromosome 1. In contrast,

the general variable selection in MT-BayesCPII-G allows loci on chromosome 2, which have no effect on trait 1, to be excluded from the model for trait 1. Thus when sufficient data were available for variable selection to exclude loci on chromosome 2 for trait 1, MT-BayesCPII-G showed a substantial advantage over MT-BayesCPII-R. On the other hand, for the trait with heritability 0.8, the contribution of noise to the prediction from the loci on chromosome 1, which had no effect on this trait, is small relative to the signal from loci on chromosome 2. Thus MT-BayesCPII-G and MT-BayesCPII-R had similar accuracies. As the training population size increased, the contribution of noise to the prediction of a trait from loci which had no effect on this trait, vanished even when the heritability was low. This was observed for both traits as apparent in figures 2 and 3. Since only bivariate data with different heritabilities showed substantial differences in prediction accuracies, traits with different heritabilities were simulated in scenario 2. In scenario 2, both markers and QTL were simulated. As expected, MT-BayesCPII-G showed higher prediction accuracy to MT-BayesCPII-R for intermediate training population, but as  $N$  increased, the accuracies of these methods asymptotically converged (Karaman *et al.* 2016).

Further, in both real and simulated analyses, MT-BayesCPII-G gave equal or higher prediction accuracy than MT-BayesCPII-R. In addition, MT-BayesCPII-R requires drawing samples from a multivariate normal distribution of order  $t$ , whereas Gibbs sampler I, which can be used for MT-BayesCPII-G, requires sampling from a univariate normal. Thus, in addition to MT-BayesCPII-G giving equal or better performance than MT-BayesCPII-R, MT-BayesCPII-G can also be computationally more efficient.

**Priors** In practice, genetic variances from previous conventional analyses are usually used to construct priors for marker effect variances. For single trait analyses, under some assumptions, it can be shown that the marker effect variance  $\sigma_a^2$  can be obtained as

$$\sigma_a^2 = \frac{\sigma_g^2}{(1 - \pi) \sum 2p_j(1 - p_j)}, \quad (2)$$

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 (Habier *et al.* 2007; Gianola *et al.* 2009; Fernando and Garrick 2013). Following a similar strategy, the marker effect covariance matrix  $\mathbf{G}$  in a two-trait analysis can be obtained as

$$\mathbf{G} = \frac{1}{\sum 2p_j(1 - p_j)} \begin{bmatrix} \frac{Q_{11}}{p(\delta=(1,1)) + p(\delta=(1,0))} & \frac{Q_{12}}{p(\delta=(1,1))} \\ \frac{Q_{21}}{p(\delta=(1,1))} & \frac{Q_{22}}{p(\delta=(1,1)) + p(\delta=(0,1))} \end{bmatrix}, \quad (3)$$

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 probabilities a marker has null effects on the first trait but not the second trait, on the second trait but not the first trait or on neither 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 (3). This strategy relating marker effect covariance matrix to genetic covariance matrix can be readily extended to more than two traits. Note that positive definite matrix  $\mathbf{Q}$  may result in negative definite

375 matrix  $\mathbf{G}$  using (3), especially when the prior for the probability  
376 a marker has null effects is far from the real value. In that  
377 case, the diagonal elements of  $\mathbf{G}$ , which are the marker effect  
378 variances for different traits, can be obtained using (2), where  
379  $\pi$  may be estimated from previous single-trait analyses, and  
380 the off-diagonal elements of  $\mathbf{G}$  may be set to zero to guarantee  
381 positive definiteness of  $\mathbf{G}$ .

382 **Multi-trait variable selection** In regard to a single trait, a locus  
383 either has an effect, or it does not. Hence, the scalar parameter  
384  $\pi$  (and its complement  $1 - \pi$ ) completely defines this circum-  
385 stance. In a multi-trait setting, it is conceivable that loci that  
386 influence one trait, may or may not influence other traits. In  
387 that circumstance, a vector  $\Pi$  is required to define the genetic  
388 architecture. The number of parameters that constitute the vec-  
389 tor  $\Pi$  is  $2^t$  which grows rapidly with the number of traits. In  
390 most cases, the researcher will have little or no knowledge of the  
391 likely extent of pleiotropy of loci that influence two traits, other  
392 than knowing or having an estimate of the genetic covariance.  
393 There are two simple ways to reduce this complexity in priors.

394 First, one can assume as did Jia and Jannink (2012) that in  
395 the context of variable selection a locus should be selected for  
396 all of the traits or selected for none of the traits, reducing the  
397 required probabilities to being analogous to the single trait  $\pi$  and  
398 ( $1 - \pi$ ). This approach has the advantage of simplicity, but the  
399 disadvantage that many effects might need to be estimated for  
400 loci that have no effect on a trait, and this may erode the accuracy  
401 of prediction. This should not be a problem for asymptotically  
402 large datasets, as in that case the fitted locus effects should  
403 converge to zero for those traits not influenced by that locus.

404 A second simple way to accommodate the multiple trait circum-  
405 stance is to assume the  $2^t$  parameters can be derived from  $t$   
406 trait-specific parameters. However, when the probability that a  
407 single trait locus has an effect is small for each of two or more  
408 traits, the pair-wise probability that a locus affects all the traits  
409 will be the product of those small probabilities, making it very  
410 difficult for loci to enter the model for all traits simultaneously.

411 The better way to solve this problem is to use a hyper-  
412 parameter  $\Pi$  that completely defines the alternative models that  
413 are required to capture all the alternative forms of genetic archi-  
414 tecture. We have shown here how this can be done, with two  
415 alternative Gibbs sampling strategies. One involves single-site  
416 sampling for one locus and trait at a time. The other samples all  
417 the alternative combinations of effects for one locus considering  
418 all traits simultaneously. We have shown that both are practical  
419 with real data and can result in improved accuracies of predic-  
420 tion in certain circumstances in terms of genetic architecture and  
421 size of dataset.

## 422 Conclusions

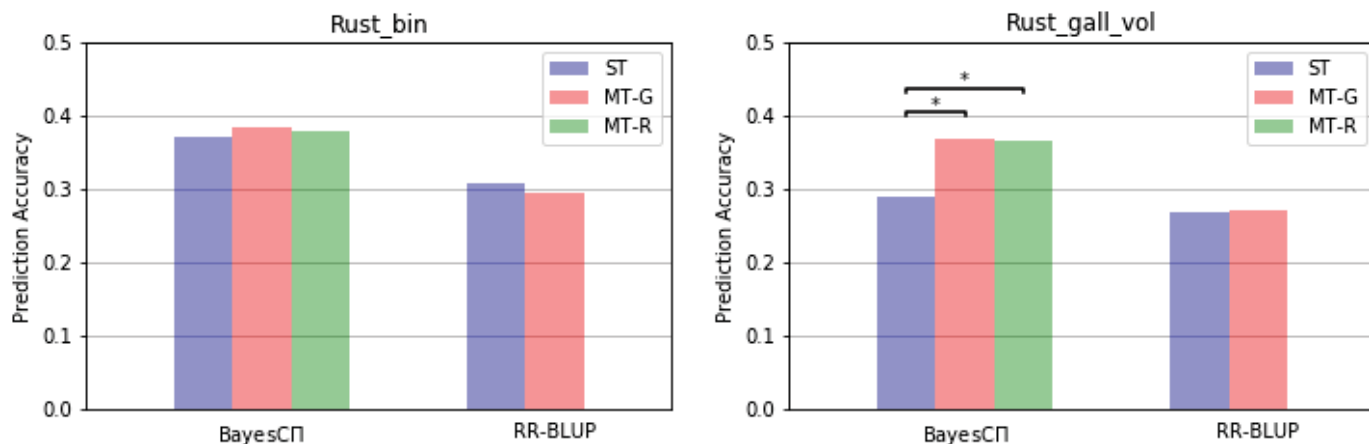
423 Many researchers are interested in genome-wide association  
424 studies and finding causal genes and variants. For those re-  
425 searchers, pleiotropy is of considerable interest, and they would  
426 want to know which loci affect which traits, from a purely biolog-  
427 ical perspective. Practitioners are often interested in "breaking"  
428 the genetic correlation, by selecting parents to give a favorable  
429 selection response in respect to multiple trait consequences. In  
430 either of these circumstances, with intermediate- rather than  
431 asymptotically-large datasets, we believe the methods described  
432 here and available in the open-source, freely-available JWAS  
433 package offer real promise.

## 434 acknowledgments

435 We thank bioRxiv for making this manuscript available early  
436 online as preprints. This work was supported by the US Depart-  
437 ment of Agriculture, Agriculture and Food Research Initiative  
438 National Institute of Food and Agriculture Competitive grant  
439 no. 2015-67015-22947.

## 440 Literature Cited

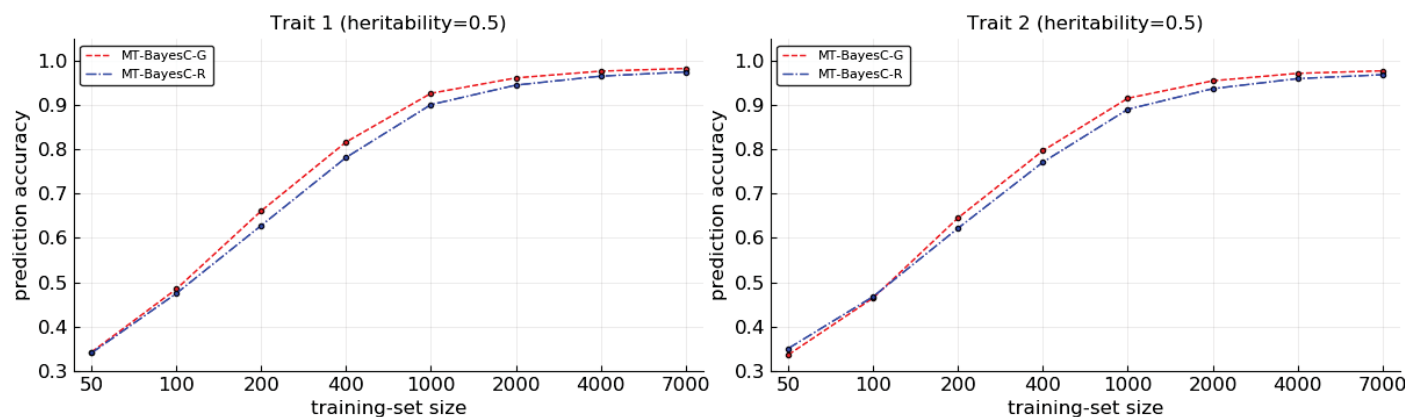
- 441 Calus, M. P. and R. F. Veerkamp, 2011 Accuracy of multi-trait  
442 genomic selection using different methods. *Genetics Selection  
443 Evolution* **43**: 26.
- 444 Cheng, H., R. L. Fernando, and D. J. Garrick, 2018 JWAS: Julia  
445 implementation of whole-genome analysis software. *Proceed-  
446 ings of the World Congress on Genetics Applied to Livestock  
447 Production*, 11.859 .
- 448 Cheng, H., D. Garrick, and R. Fernando, 2015a XSim: Simula-  
449 tion of Descendants from Ancestors with Sequence Data. *G3  
450 (Bethesda, Md.)* **5**: 1415–1417.
- 451 Cheng, H., L. Qu, D. J. Garrick, and R. L. Fernando, 2015b A  
452 fast and efficient Gibbs sampler for BayesB in whole-genome  
453 analyses. *Genetics Selection Evolution* **47**: 1819.
- 454 Daetwyler, H. D., M. P. L. Calus, R. Pong-Wong, G. de los Cam-  
455 pos, and J. M. Hickey, 2013 Genomic Prediction in Animals  
456 and Plants: Simulation of Data, Validation, Reporting, and  
457 Benchmarking. *Genetics* **193**: 347–365.
- 458 Fernando, R. L. and D. Garrick, 2013 Bayesian Methods Applied  
459 to GWAS. In *Genome-Wide Association Studies and Genomic Pre-  
460 diction*, pp. 237–274, Humana Press, Totowa, NJ.
- 461 Gianola, D., G. de los Campos, W. G. Hill, E. Manfredi,  
462 and R. Fernando, 2009 Additive genetic variability and the  
463 Bayesian alphabet. *Genetics* **183**: 347–363.
- 464 Habier, D., R. L. Fernando, and J. C. M. Dekkers, 2007 The Im-  
465 pact of Genetic Relationship Information on Genome-Assisted  
466 Breeding Values. *Genetics* **177**: 2389–2397.
- 467 Habier, D., R. L. Fernando, K. Kizilkaya, and D. J. Garrick,  
468 2011 Extension of the bayesian alphabet for genomic selec-  
469 tion. *BMC bioinformatics* **12**: 186.
- 470 Jia, Y. and J.-L. Jannink, 2012 Multiple-Trait Genomic Selection  
471 Methods Increase Genetic Value Prediction Accuracy. *Genetics*  
472 **192**: 1513–1522.
- 473 Karaman, E., H. Cheng, M. Z. Firat, D. J. Garrick, and R. L.  
474 Fernando, 2016 An Upper Bound for Accuracy of Prediction  
475 Using GBLUP. *PLoS one* **11**: e0161054.
- 476 Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard, 2001 Pre-  
477 diction of Total Genetic Value Using Genome-Wide Dense  
478 Marker Maps. *Genetics* **157**: 1819–1829.
- 479 Resende, M. F. R., P. Muñoz, M. D. V. Resende, D. J. Garrick, R. L.  
480 Fernando, J. M. Davis, E. J. Jokela, T. A. Martin, G. F. Peter,  
481 and M. Kirst, 2012 Accuracy of Genomic Selection Methods in  
482 a Standard Data Set of Loblolly Pine (*Pinus taeda* L.). *Genetics*  
483 **190**: 1503–1510.



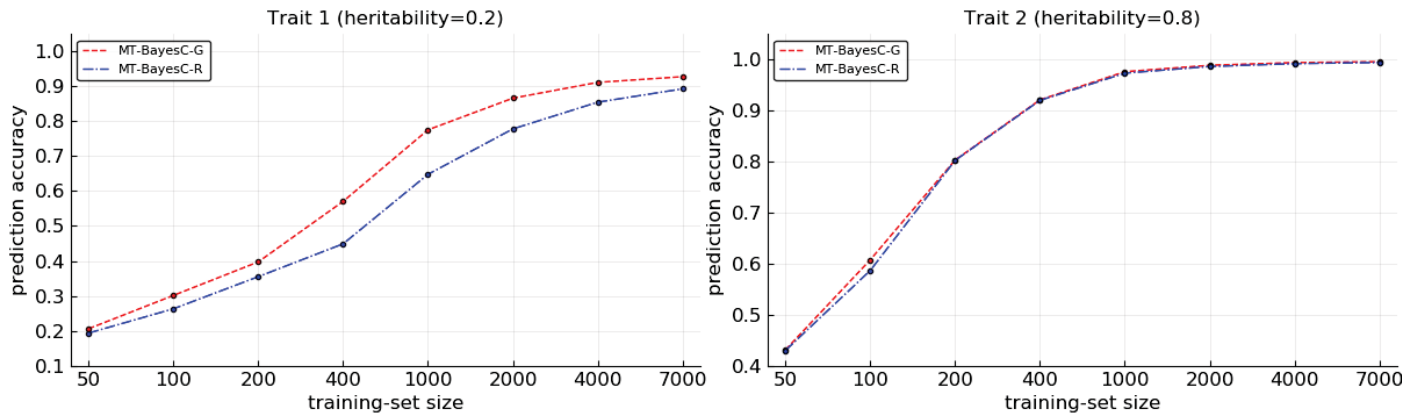
**Figure 1** Comparison of single-trait and multi-trait methods for Rust\_bin and Rust\_gall\_vol traits. ST, MT-G and MT-R indicate single-trait, our general multi-trait and restricted multi-trait analyses respectively. \*, indicates a statistically significant ( $P < 0.01$ ) difference between methods.

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

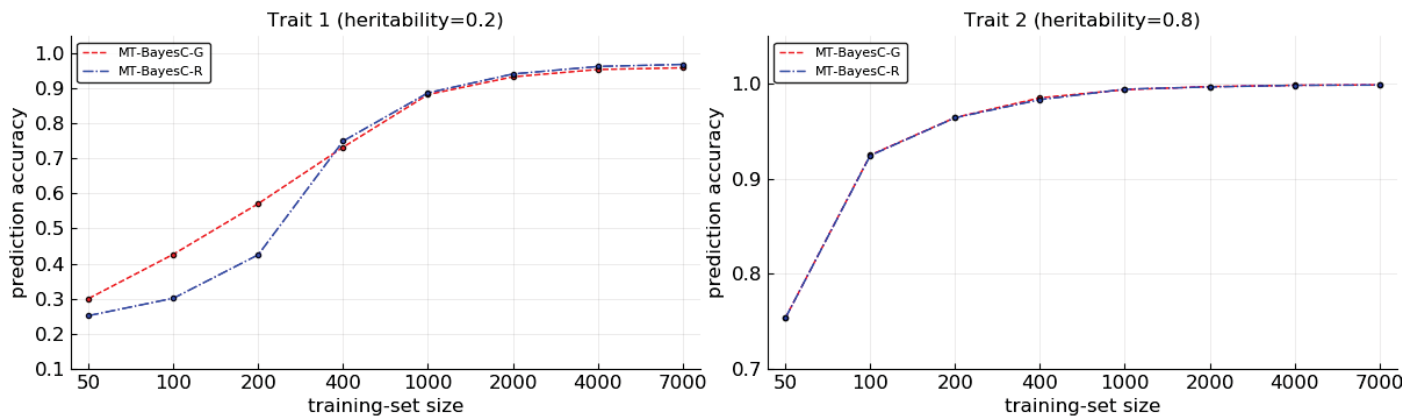
**Table 1** Estimation of  $\pi$  for alternative multi-trait BayesCPI methods. Posterior mean of  $\Pi$  were given for different categories of  $\delta$ . Different categories of  $\delta$  are denoted as  $(k_1, k_2)$ , where  $k_1 = 0$  if a marker has a null effect on Rust\_bin, otherwise  $k_1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



**Figure 2** Comparison of multi-trait BayesCPI methods for situation 1 under simulation scenario 1.



**Figure 3** Comparison of multi-trait BayesCII methods for situation 2 under simulation scenario 1.



**Figure 4** Comparison of multi-trait BayesCII methods under simulation scenario 2.



485 **Gibbs sampler algorithm for multi-trait BayesCII-G**486 **Single-site Gibbs sampler for multi-trait BayesCII-G**

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

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

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

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

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

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

$$\begin{aligned} \mathbf{C}_j &= \begin{bmatrix} \mathbf{C}_{j,11}^0 & \mathbf{C}_{j,12}^0 \\ \mathbf{C}_{j,21}^0 & \mathbf{C}_{j,22}^0 \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{G}^{11} & \mathbf{G}^{12} \\ \mathbf{G}^{21} & \mathbf{G}^{22} + \mathbf{D}'_{j2} \mathbf{R}^{22} \mathbf{D}_{j2} \sum_{i=1}^n m_{ij}^2 \end{bmatrix} \\ \mathbf{r}'_j &= \begin{bmatrix} r_{j1}^0 & r_{j2}^{0'} \end{bmatrix} \\ &= \begin{bmatrix} 0 & \left( \sum_{i=1}^n \mathbf{w}'_i m_{ij} \right) \begin{bmatrix} \mathbf{R}^{12} \\ \mathbf{R}^{22} \end{bmatrix} \mathbf{D}_{j2} \end{bmatrix} \end{aligned}$$

489 When  $\delta_{j1} = 1$ ,

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

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

$$f(\beta_{j1} \mid \delta_{j1} = 0, \boldsymbol{\beta}_{-j1}, \mathbf{D}_{-j1}, \mathbf{G}, \mathbf{R}, \mathbf{y}) \propto N\left(\hat{\beta}_{j1}^0, \left(\mathbf{C}_{j,11}^0\right)^{-1}\right) = N\left(-\left(\mathbf{G}^{11}\right)^{-1} \mathbf{G}^{12} \boldsymbol{\beta}_{j2}, \left(\mathbf{G}^{11}\right)^{-1}\right).$$

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

$$f(\beta_{j1} \mid \delta_{j1} = 1, \boldsymbol{\beta}_{-j1}, \mathbf{D}_{-j1}, \mathbf{G}, \mathbf{R}, \mathbf{y}) \propto N\left(\hat{\beta}_{j1}^1, \left(\mathbf{C}_{j,11}^1\right)^{-1}\right) = N\left(\left(\mathbf{C}_{j,11}^1\right)^{-1} \left(r_{j1} - \mathbf{C}_{j,12}^1 \boldsymbol{\beta}_{j2}\right), \left(\mathbf{C}_{j,11}^1\right)^{-1}\right).$$

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

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

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

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

492 Note that  $\sum_i \mathbf{w}_i' \mathbf{R}^{-1} \mathbf{w}_i$ ,  $\mathbf{r}'_{j2} \boldsymbol{\beta}_{j2}$ ,  $\boldsymbol{\beta}'_{j2} \mathbf{C}_{j,22} \boldsymbol{\beta}_{j2}$  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

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

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

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

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

495 The full conditional distribution for  $\boldsymbol{\Pi}$  can be written as

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

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

#### 497 **Joint Gibbs sampler for multi-trait BayesCPI-G**

Let  $\boldsymbol{\theta}$  denote all other parameters except  $\boldsymbol{\beta}_j$  and  $\delta_j$ , then our sampling scheme can be written as

$$f(\boldsymbol{\beta}_j, \delta_j | \boldsymbol{\theta}, \mathbf{y}) = f(\delta_j | \boldsymbol{\theta}, \mathbf{y}) f(\boldsymbol{\beta}_j | \delta_j, \boldsymbol{\theta}, \mathbf{y})$$

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

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

498 Denote  $\mathbf{w}_i = \mathbf{y}_i - \boldsymbol{\mu}_i - \sum_{j' \neq j} m_{ij'} \mathbf{D}_{j'} \boldsymbol{\beta}_{j'}$ , then

$$\begin{aligned}
f(\mathbf{y} | \delta_j, \boldsymbol{\theta}) &\propto \int f(\mathbf{y} | \boldsymbol{\beta}, \mathbf{D}, \mathbf{R}) f(\boldsymbol{\beta}_j | \mathbf{G}) d\boldsymbol{\beta}_j \\
&\propto \int \exp \left[ -\frac{1}{2} \sum_{i=1}^n (\mathbf{w}_i - m_{ij} \mathbf{D}_j \boldsymbol{\beta}_j)' \mathbf{R}^{-1} (\mathbf{w}_i - m_{ij} \mathbf{D}_j \boldsymbol{\beta}_j) \right] \exp \left( -\frac{1}{2} \boldsymbol{\beta}_j' \mathbf{G}^{-1} \boldsymbol{\beta}_j \right) d\boldsymbol{\beta}_j \\
&\propto \int \exp \left\{ -\frac{1}{2} \left[ \boldsymbol{\beta}_j' \left( \mathbf{D}_j' \mathbf{R}^{-1} \mathbf{D}_j \sum_{i=1}^n m_{ij}^2 + \mathbf{G}^{-1} \right) \boldsymbol{\beta}_j - 2 \sum_{i=1}^n \mathbf{w}_i' m_{ij} \mathbf{R}^{-1} \mathbf{D}_j \boldsymbol{\beta}_j + \sum_{i=1}^n \mathbf{w}_i' \mathbf{R}^{-1} \mathbf{w}_i \right] \right\} d\boldsymbol{\beta}_j \\
&\propto \int \exp \left\{ -\frac{1}{2} \left[ \boldsymbol{\beta}_j' \mathbf{C}_j \boldsymbol{\beta}_j - 2 \mathbf{r}_j' \boldsymbol{\beta}_j + \sum_{i=1}^n \mathbf{w}_i' \mathbf{R}^{-1} \mathbf{w}_i \right] \right\} d\boldsymbol{\beta}_j \\
&\propto \int \exp \left\{ -\frac{1}{2} \left[ (\boldsymbol{\beta}_j' - \mathbf{r}_j' \mathbf{C}_j^{-1}) \mathbf{C}_j (\boldsymbol{\beta}_j - \mathbf{C}_j^{-1} \mathbf{r}_j) + \sum_{i=1}^n \mathbf{w}_i' \mathbf{R}^{-1} \mathbf{w}_i - \mathbf{r}_j' \mathbf{C}_j^{-1} \mathbf{r}_j \right] \right\} d\boldsymbol{\beta}_j \\
&\propto \exp \left\{ -\frac{1}{2} \left[ \sum_{i=1}^n \mathbf{w}_i' \mathbf{R}^{-1} \mathbf{w}_i - \mathbf{r}_j' \mathbf{C}_j^{-1} \mathbf{r}_j \right] \right\} \\
&\times |\mathbf{C}_j^{-1}|^{\frac{1}{2}} \int |\mathbf{C}_j^{-1}|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (\boldsymbol{\beta}_j' - \mathbf{r}_j' \mathbf{C}_j^{-1}) \mathbf{C}_j (\boldsymbol{\beta}_j - \mathbf{C}_j^{-1} \mathbf{r}_j) \right] d\boldsymbol{\beta}_j \\
&\propto |\mathbf{C}_j^{-1}|^{\frac{1}{2}} \exp \left\{ -\frac{1}{2} \left[ \sum_{i=1}^n \mathbf{w}_i' \mathbf{R}^{-1} \mathbf{w}_i - \mathbf{r}_j' \mathbf{C}_j^{-1} \mathbf{r}_j \right] \right\},
\end{aligned}$$

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

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

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

where

$$f(\mathbf{y} | \delta_j, \boldsymbol{\theta}) \propto |\mathbf{C}_j^{-1}|^{\frac{1}{2}} \exp \left\{ \frac{1}{2} \mathbf{r}_j' \mathbf{C}_j^{-1} \mathbf{r}_j \right\}.$$

500 The full conditional distribution of  $\boldsymbol{\beta}_j$  is

$$\begin{aligned}
f(\boldsymbol{\beta}_j | \delta_j, \boldsymbol{\theta}, \mathbf{y}) &\propto \exp \left[ -\frac{1}{2} \sum_{i=1}^n (\mathbf{w}_i - m_{ij} \mathbf{D}_j \boldsymbol{\beta}_j)' \mathbf{R}^{-1} (\mathbf{w}_i - m_{ij} \mathbf{D}_j \boldsymbol{\beta}_j) \right] \exp \left( -\frac{1}{2} \boldsymbol{\beta}_j' \mathbf{G}^{-1} \boldsymbol{\beta}_j \right), \\
&\propto \exp \left\{ -\frac{1}{2} \left[ \boldsymbol{\beta}_j' \left( \mathbf{D}_j' \mathbf{R}^{-1} \mathbf{D}_j \sum_{i=1}^n m_{ij}^2 + \mathbf{G}^{-1} \right) \boldsymbol{\beta}_j - 2 \sum_{i=1}^n \mathbf{w}_i' m_{ij} \mathbf{R}^{-1} \mathbf{D}_j \boldsymbol{\beta}_j \right] \right\} \\
&\propto \exp \left\{ -\frac{1}{2} \left[ \boldsymbol{\beta}_j' \mathbf{C}_j \boldsymbol{\beta}_j - 2 \mathbf{r}_j' \boldsymbol{\beta}_j \right] \right\} \\
&\propto \exp \left\{ -\frac{1}{2} (\boldsymbol{\beta}_j' - \mathbf{r}_j' \mathbf{C}_j^{-1}) \mathbf{C}_j (\boldsymbol{\beta}_j - \mathbf{C}_j^{-1} \mathbf{r}_j) \right\} \\
&\propto N(\mathbf{C}_j^{-1} \mathbf{r}_j, \mathbf{C}_j^{-1})
\end{aligned}$$

501 **Gibbs sampler algorithm for multi-trait BayesB**

502 **Single-site Gibbs sampler for multi-trait BayesB**

For convenience, from now on let "1" denote trait  $k$  and "2" the other traits. Thus,  $\boldsymbol{\beta}_j$  can be denoted as  $\begin{bmatrix} \beta_{j1} \\ \beta_{j2} \end{bmatrix}$  and  $\mathbf{D}_j$  can be denoted

as  $\begin{bmatrix} \delta_{j1} & 0 \\ 0 & \mathbf{D}_{j2} \end{bmatrix}$ . 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  $\delta_j$ . Let  $\boldsymbol{\theta}$  denote all other parameters except  $\delta_{j1}$  and  $\beta_{j1}$ , then our sampling scheme can be written as

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



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

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

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

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

503 where  $\mathbf{C}_j = \mathbf{D}_j' \mathbf{R}^{-1} \mathbf{D}_j \sum_{i=1}^n m_{ij}^2 + \mathbf{G}_j^{-1}$  and  $\mathbf{r}_j' = (\sum_{i=1}^n \mathbf{w}_i' m_{ij}) \mathbf{R}^{-1} \mathbf{D}_j$ .

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

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

505 When  $\delta_{j1} = 1$ ,

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

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

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

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

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

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

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

508 Note that  $\sum_i \mathbf{w}_i' \mathbf{R}^{-1} \mathbf{w}_i, \mathbf{r}'_{j2} \boldsymbol{\beta}_{j2}, \boldsymbol{\beta}'_{j2} \mathbf{C}_{j,22} \boldsymbol{\beta}_{j2}$  are same when  $\delta_{j1} = 0$  or  $1$ . Thus the ratio  $\frac{f(\mathbf{y} \mid \delta_{j1}=1, \boldsymbol{\theta})}{f(\mathbf{y} \mid \delta_{j1}=0, \boldsymbol{\theta})}$  becomes

$$\begin{aligned}
H &= (\mathbf{C}_{j,11}^1)^{-\frac{1}{2}} (\mathbf{G}_j^{11})^{\frac{1}{2}} \exp \left( -\frac{1}{2} (\hat{\boldsymbol{\beta}}_{j1}^0{}^2 \mathbf{G}_j^{11} - \hat{\boldsymbol{\beta}}_{j1}^1{}^2 \mathbf{C}_{j,11}^1) \right) \\
&= \exp \left\{ -\frac{1}{2} (\log \mathbf{C}_{j,11}^1 - \hat{\boldsymbol{\beta}}_{j1}^1{}^2 \mathbf{C}_{j,11}^1) - \left( -\frac{1}{2} (\log \mathbf{G}_j^{11} - \hat{\boldsymbol{\beta}}_{j1}^0{}^2 \mathbf{G}_j^{11}) \right) \right\}
\end{aligned}$$

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

$$\left\{ 1 + \frac{f(\mathbf{y} \mid \delta_{j1} = 0, \boldsymbol{\theta}) f(\delta_{j1} = 0, \delta_{j2} \mid \boldsymbol{\Pi}_1, \boldsymbol{\Pi}_2, \dots)}{f(\mathbf{y} \mid \delta_{j1} = 1, \boldsymbol{\theta}) f(\delta_{j1} = 1, \delta_{j2} \mid \boldsymbol{\Pi}_1, \boldsymbol{\Pi}_2, \dots)} \right\}^{-1} = \left\{ 1 + \left( \frac{\boldsymbol{\Pi}_{j0}}{\boldsymbol{\Pi}_{j1}} H \right)^{-1} \right\}^{-1},$$

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

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

512 Let  $\boldsymbol{\theta}$  denote all other parameters except  $\boldsymbol{\beta}_j$  and  $\delta_j$ , then our sampling scheme can be written as

$$f(\boldsymbol{\beta}_j, \delta_j \mid \boldsymbol{\theta}, \mathbf{y}) = f(\delta_j \mid \boldsymbol{\theta}, \mathbf{y}) f(\boldsymbol{\beta}_j \mid \delta_j, \boldsymbol{\theta}, \mathbf{y})$$

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

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

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

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

515 where  $\mathbf{C}_j = \mathbf{D}_j' \mathbf{R}^{-1} \mathbf{D}_j \sum_{i=1}^n m_{ij}^2 + \mathbf{G}_j^{-1}$  and  $\mathbf{r}_j' = \left( \sum_{i=1}^n w_i' m_{ij} \right) \mathbf{R}^{-1} \mathbf{D}_j$ .

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

$$f(\delta_j \mid \boldsymbol{\theta}, \mathbf{y}) = \frac{f(\mathbf{y} \mid \delta_j, \boldsymbol{\theta}) f(\delta_j \mid \boldsymbol{\Pi})}{\sum_{\delta_j} f(\mathbf{y} \mid \delta_j, \boldsymbol{\theta}) f(\delta_j \mid \boldsymbol{\Pi})},$$

where

$$f(\mathbf{y} \mid \delta_j, \boldsymbol{\theta}) \propto |\mathbf{C}_j^{-1}|^{\frac{1}{2}} \exp \left\{ \frac{1}{2} \mathbf{r}_j' \mathbf{C}_j^{-1} \mathbf{r}_j \right\}.$$

516 The full conditional distribution of  $\boldsymbol{\beta}_j$  is

$$\begin{aligned}
f(\boldsymbol{\beta}_j \mid \delta_j, \boldsymbol{\theta}, \mathbf{y}) &\propto \exp \left[ -\frac{1}{2} \sum_{i=1}^n (w_i - m_{ij} D_j \beta_j)' \mathbf{R}^{-1} (w_i - m_{ij} D_j \beta_j) \right] \exp \left( -\frac{1}{2} \boldsymbol{\beta}_j' \mathbf{G}_j^{-1} \boldsymbol{\beta}_j \right), \\
&\propto \exp \left\{ -\frac{1}{2} \left[ \boldsymbol{\beta}_j' \left( \mathbf{D}_j' \mathbf{R}^{-1} \mathbf{D}_j \sum_{i=1}^n m_{ij}^2 + \mathbf{G}_j^{-1} \right) \boldsymbol{\beta}_j - 2 \sum_{i=1}^n w_i' m_{ij} \mathbf{R}^{-1} \mathbf{D}_j \boldsymbol{\beta}_j \right] \right\} \\
&\propto \exp \left\{ -\frac{1}{2} \left[ \boldsymbol{\beta}_j' \mathbf{C}_j \boldsymbol{\beta}_j - 2 \mathbf{r}_j' \boldsymbol{\beta}_j \right] \right\} \\
&\propto \exp \left\{ -\frac{1}{2} \left( \boldsymbol{\beta}_j' - \mathbf{r}_j' \mathbf{C}_j^{-1} \right) \mathbf{C}_j \left( \boldsymbol{\beta}_j - \mathbf{C}_j^{-1} \mathbf{r}_j \right) \right\} \\
&\propto N \left( \mathbf{C}_j^{-1} \mathbf{r}_j, \mathbf{C}_j^{-1} \right)
\end{aligned}$$