

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

Hao Cheng^{*,**,*1}, Kadir Kizilkaya[†], Jian Zeng[‡], Dorian Garrick[§] and Rohan Fernando^{*}

^{*}Department of Animal Science, Iowa State University, Ames, Iowa, ^{**}Department of Statistics, Iowa State University, Ames, Iowa, [†]Department of Animal Science, Adnan Menderes University, Aydin, Turkey, [‡]Program in Complex Trait Genomics, Institute for Molecular Bioscience, University of Queensland, St Lucia, Brisbane, Australia, [§]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 π , 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 JWAS 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 π (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 π further treats π 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

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

where \mathbf{y}_i is a vector of phenotypes of t traits for individual i , $\boldsymbol{\mu}$ 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, α_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 \mathbf{R} , which in turn is assumed to have an inverse Wishart prior distribution, $W_t^{-1}(\mathbf{S}_e, \nu_e)$.

Multi-trait BayesCPII model

Priors for marker effects The prior for α_{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 σ_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., α_{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 α_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 δ_{jk} is an indicator variable indicating whether the marker effect of locus j for trait k is zero or non-zero, and β_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 BayesCPII 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 δ_j . For example, in a $t=2$ trait model, there are $2^2 = 4$ combinations of δ_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 δ_j , then the prior for δ_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 Π_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 $\boldsymbol{\Pi} = (\Pi_1, \Pi_2, \dots, \Pi_l)$. 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 BayesCPII Suppose the prior for δ_j is a categorical distribution whose support is for all 2^t possible outcomes of δ_j . For convenience, from now on let "1" denote trait k and "2" the other $t - 1$ traits. In our sampling scheme, β_{j1} and δ_{j1} are sampled from their joint full conditional distributions, which can be written as the product of the full conditional distribution of β_{j1} given δ_{j1} and the marginal full conditional distribution of δ_{j1} . Let $\boldsymbol{\theta}$ denote all other parameters except δ_{j1} and β_{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 β_{j1} , δ_{j1} , $\boldsymbol{\Pi}$, \mathbf{G} and \mathbf{R} for Gibbs sampler I, which were derived in the Appendix, are given below.

The full conditional distributions of β_{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

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

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

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

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

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

where n_i is the number of 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 β_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 with the i th row as $\boldsymbol{\beta}'_i$.

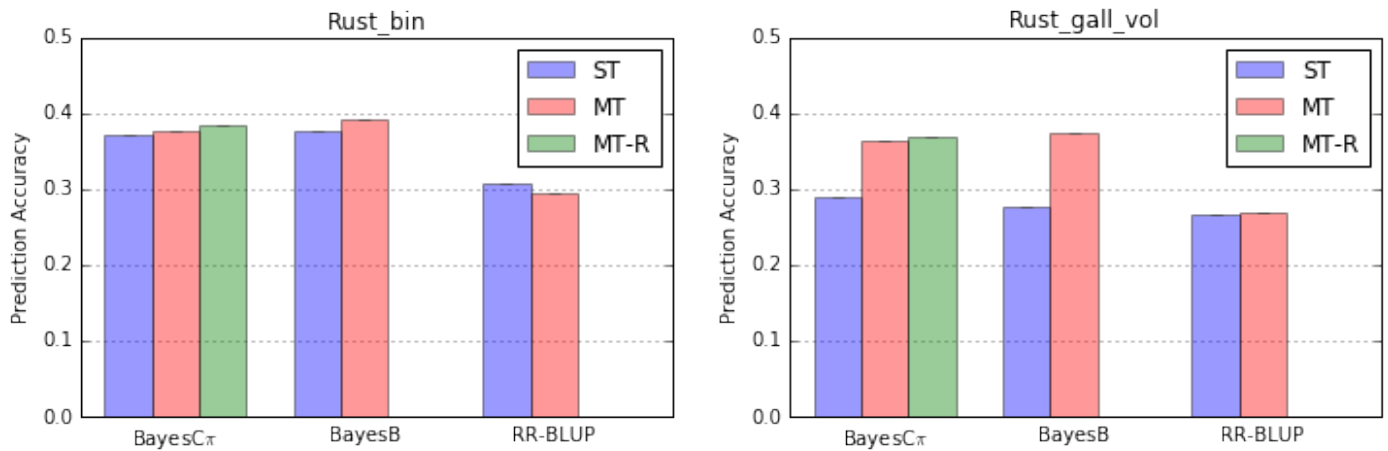


Figure 1 Comparison of single-trait and multi-trait methods for Rust_bin and Rust_gall_vol traits.

104 **Gibbs sampler II for multi-trait BayesCII** The Gibbs sampler
105 above requires that all 2^t outcomes for δ_j have non-zero prior
106 probabilities, i.e. none of Π_i can be zero. If some Π_i are zero,
107 the markov chain generated from Gibbs sampler I may not be
108 irreducible. Another more general Gibbs sampler that does not
109 require all Π_i to be non-zero is proposed below.

110 The full conditional distributions of β_j , δ_j , Π , \mathbf{G} , \mathbf{R} for Gibbs
111 sampler II, which were derived in the Appendix, are given below.
112

113 Let θ denote all other parameters except β_j and δ_j , then our
114 sampling scheme can be written as

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

115 The full conditional distribution of β_j is

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

116 where $\mathbf{C}_j = \mathbf{D}'_j \mathbf{R}^{-1} \mathbf{D}_j \sum_{i=1}^n m_{ij}^2 + \mathbf{G}^{-1}$ and $\mathbf{r}'_j =$
117 $(\sum_{i=1}^n \mathbf{w}'_i m_{ij}) \mathbf{R}^{-1} \mathbf{D}_j$.

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

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

119 where

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

120 This Gibbs sampler can accommodate the restricted multi-
121 trait BayesCII that was proposed by Jia et al. (Jia and Jannink
122 2012), which only allows δ_j to be a vector of all ones or a vector
123 of all zeros.

124 Multi-trait BayesB Model

125 The multi-trait BayesCII model proposed above can be modified
126 to accommodate the multi-trait BayesB model. Model equation
127 (1) can also be used for the multi-trait BayesB method. The
128 differences in multi-trait BayesB method is that the prior for
129 β_j is a multivariate normal distribution with null mean and
130 locus-specific covariance matrix \mathbf{G}_j . The locus-specific covari-
131 ance matrix \mathbf{G}_j is a priori assumed to follow an inverse Wishart
132 distribution, $W_t^{-1}(\mathbf{S}_\beta, \nu_\beta)$.

133 The derivations of the full conditional distributions of pa-
134 rameters of interest for Gibbs samplers are shown in the Ap-
135 pendix. In the multi-trait BayesB model, the full conditional
136 distributions for all parameters except \mathbf{G}_j are similar to the multi-
137 trait BayesCII model. The full conditional distribution for \mathbf{G}_j ,
138 the covariance matrix for β_j , is a inverse Wishart distribution,
139 $W_t^{-1}(\mathbf{S}_\beta + \beta_j \beta'_j, \nu_\beta + 1)$.

140 Data analyses

141 Published genotypic and deregressed phenotypic data for
142 Loblolly Pine (*Pinus Taeda* L.) were used (Resende et al. 2012).
143 Two disease traits, namely Rust_bin and Rust_gall_vol were ana-
144 lyzed. The reported heritability was 0.21 for Rust_bin and 0.12
145 for Rust_gall_vol. Loci with missing genotypes were imputed as
146 the mean of the observed genotype covariates at that locus and
147 loci with a missing rate >50% were excluded. After these quality
148 control edits, 4,828 SNPs on 807 individuals with phenotypes
149 and genotypes on both traits remained.

150 Prediction accuracy was calculated as the correlation be-
151 tween the vector of deregressed phenotypes and the vector of
152 estimated breeding values. Cross-validation using 10-folds
153 formed the basis for comparing our general multi-trait BayesCII
154 model (MT-BayesCII) to a similar model where the prior for
155 β_j is a multivariate normal rather than a mixture of multivari-
156 ate normals (MT-BayesC0), the restricted multi-trait BayesCII
157 proposed by Jia et al. (MT-BayesCII-R), multi-trait BayesB
158 with known Π (MT-BayesB) and the usual single trait formu-
159 lations of the mixture models (ST-BayesC0, ST-BayesCπ, ST-
160 BayesB). The constant Π used in BayesB were estimated using
161 BayesCII methods. All analyses were performed using JWAS
162 (Cheng et al. 2016), a publicly-available package for single-trait
163 and multi-trait whole-genome analyses written in the freely-

164 available Julia language. Since BayesC0 is equivalent to ran- 216
 165 dom regression best linear unbiased prediction (RR-BLUP), ST- 217
 166 BayesC0 and MT-BayesC0 are denoted as ST-RR-BLUP and MT- 218
 167 RR-BLUP below. The prior for the residual covariance matrix 219
 168 \mathbf{R} in all multi-trait methods was an inverse Wishart distribu- 220
 169 tion, $W^{-1}\left(\begin{bmatrix} 0.003 & 0 \\ 0 & 0.003 \end{bmatrix}, 6\right)$, for which the mean of R is 221
 170 $\begin{bmatrix} 0.001 & 0 \\ 0 & 0.001 \end{bmatrix}$. The prior for the marker effects covariance 222
 171 matrix \mathbf{G} in MT-BayesCII and MT-BayesCII-R was an inverse 223
 172 Wishart distribution, $W^{-1}\left(\begin{bmatrix} 0.003 & 0 \\ 0 & 0.003 \end{bmatrix}, 6\right)$, for which the 224
 173 mean of \mathbf{G} was $\begin{bmatrix} 0.001 & 0 \\ 0 & 0.001 \end{bmatrix}$. The priors for the residual vari- 225
 174 ance and marker effects variance in single-trait analyses were a 226
 175 scaled inverted chi-squared distribution with scale parameter 227
 176 $S^2 = 0.0005$ and degrees of freedom $\nu = 4$, for which the mean 228
 177 of the prior was also 0.001. Marker effect variances estimated 229
 178 from BayesCII were used to construct the priors for marker 230
 179 effect variances in the BayesB methods. 231

180 Results

181 The prediction accuracies from all methods for Rust_bin and 242
 182 Rust_gall_vol are in figure 1. The prediction accuracies from all 243
 183 single-trait analyses using JWAS are similar to those in (Resende 244
 184 *et al.* 2012). ST-BayesC π showed higher prediction accuracies 245
 185 than ST-RR-BLUP for both traits (Resende *et al.* 2012). The pre- 246
 186 diction accuracies from ST-BayesB were similar to those from 247
 187 ST-BayesC π , when both marker effect variances and π estimated 248
 188 from ST-BayesC π were used in ST-BayesB. 249

189 The analyses of Rust_bin exhibited no significant difference 250
 190 between multi-trait and single-trait analyses within each method 251
 191 (ST-RR-BLUP versus MT-RR-BLUP; ST-BayesC π versus MT- 252
 192 BayesCII; ST-BayesC π versus MT-BayesCII-R; ST-BayesB ver- 253
 193 sus MT-BayesB). 254

194 In contrast, analyses for the lower heritability Rust_gall_vol 255
 195 with MT-BayesCII showed significantly higher accuracies than 256
 196 ST-BayesC π . MT-BayesCII and MT-BayesCII-R showed similar 257
 197 prediction accuracies. The posterior means of Π for both meth- 258
 198 ods were shown in table 1. The performance of MT-BayesB were 259
 199 similar to MT-BayesCII, when both marker effect variances and 260
 200 Π estimated from MT-BayesCII were used. Similar prediction 261
 201 accuracies were observed in MT-RR-BLUP and ST-RR-BLUP for 262
 202 trait Rust_gall_vol.

203 Discussion

204 In the single trait analyses, accuracies from ST-BayesC π and 263
 205 ST-BayesB were higher than those from ST-RR-BLUP, suggesting 264
 206 that these two traits are influenced by a few QTL with large 265
 207 effects. The effect of genetic architecture on the performance 266
 208 of multi-trait analyses has been studied in previous simulation 267
 209 analyses (Jia and Jannink 2012). Using simulated data they found 268
 210 that multi-trait Bayesian variable selection methods outperform 269
 211 multi-trait RR-BLUP in the presence of major QTL. This observa-
 212 tion was confirmed in our real data analyses that MT-BayesCII
 213 and MT-BayesB outperformed MT-RR-BLUP for both traits.

214 Significant differences between multi-trait and single-trait
 215 analyses were only observed for Rust_gall_vol within BayesC π

216 and BayesB methods (MT-BayesCII versus ST-BayesC π ; MT-
 217 BayesB versus ST-BayesB). MT-BayesCII and MT-BayesCII-R
 218 outperformed ST-BayesC π for Rust_gall_vol, and the accuracy
 219 gain was 26% (from 0.287 to 0.364). The lower-heritability trait
 220 Rust_gall_vol may borrow information from the other corre-
 221 lated trait Rust_bin. Thus higher prediction accuracy from
 222 MT-BayesCII were observed in trait Rust_gall_vol instead of
 223 Rust_bin. Results in (Jia and Jannink 2012) showed no difference
 224 between MT-BayesCII-R and ST-BayesC π because a reduced
 225 marker panel (500 markers) was used. The performance of MT-
 226 BayesB was similar to MT-BayesCII, when both marker effect
 227 variances and Π estimated from MT-BayesCII were used. Fur-
 228 ther analyses may be required to study the effects of priors in
 229 MT-BayesB.

230 The fact that RR-BLUP showed no improvement in multi-trait
 231 analyses suggested that benefits from MT-BayesCII may caused
 232 by the estimation of hyper-parameter Π . In the MT-BayesCII,
 233 the mean of the posterior probability that a marker has a null
 234 effect on Rust_gall_vol was about 0.97, calculated as the summa-
 235 tion of posterior mean of Π for categories (0,0) and (1,0). The
 236 posterior mean of π , the probability that a marker has a null
 237 effect, in ST-BayesC π for Rust_gall_vol was 0.74, different from
 238 the equivalent value, 0.97, in MT-BayesCII showed above. Thus
 239 ST-BayesC π with constant π , equal to 0.97, were performed. Pre-
 240 diction accuracies from ST-BayesC π with constant $\pi = 0.97$ was
 241 0.361, which was similar to the accuracies from MT-BayesCII.
 242 This suggests that high-heritability traits may help with variable
 243 selection in correlated low-heritability traits.

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

255 In practice, genetic variances from previous conventional
 256 analyses are always used to construct priors for marker ef-
 257 fect variances. For single trait analyses, under some assump-
 258 tions, it can be shown that the marker effect variance $\sigma_{\alpha}^2 =$
 259 $\frac{\sigma_g^2}{(1-\pi)\sum 2p_j(1-p_j)}$, where σ_g^2 is the genetic variance, p_j is the al-
 260 lele frequency for locus j and π is the probability that a marker
 261 has a null effect. Following similar strategies, the marker effect
 262 covariance matrix \mathbf{G} in two-trait analyses can be obtained as

$$263 \mathbf{G} = \frac{1}{\sum 2p_j(1-p_j)} \begin{bmatrix} \frac{Q_{11}}{p(\delta=(1,1))+p(\delta=(1,0))} & \frac{Q_{12}}{p(\delta=(1,1))} \\ \frac{Q_{21}}{p(\delta=(1,1))} & \frac{Q_{22}}{p(\delta=(1,1))+p(\delta=(0,1))} \end{bmatrix}, \quad (2)$$

264 where $\mathbf{Q} = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix}$ is the genetic covariance matrix and
 265 $p(\delta = (0,1))$, $p(\delta = (1,0))$, $p(\delta = (1,1))$ are the probability a
 266 marker has null effects on the first trait but not the second trait,
 267 on the second trait but not the first trait and on no traits. Thus
 268 the probability that a marker has an effect on the first trait can
 269 be obtained as $p(\delta = (1,1)) + p(\delta = (1,0))$, which is the de-
 270 nominator of the upper left element in (2). This strategy relating

	Different Categories of δ			
	(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 π for alternative multi-trait BayesCII methods. Posterior mean of Π were given for different categories of δ . Different categories of δ 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.

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

275 Literature Cited

- 276 Calus, M. P. and R. F. Veerkamp, 2011 Accuracy of multi-trait
277 genomic selection using different methods. *Genetics Selection
278 Evolution* **43**: 26.
- 279 Cheng, H., D. J. Garrick, and R. L. Fernando, 2016 JWAS: Julia
280 implementation of whole-genome analyses software using
281 univariate and multivariate Bayesian mixed effects model.
282 available from <http://QTL.rocks>.
- 283 Cheng, H., L. Qu, D. J. Garrick, and R. L. Fernando, 2015 A
284 fast and efficient Gibbs sampler for BayesB in whole-genome
285 analyses. *Genetics Selection Evolution* **47**: 1819.
- 286 Habier, D., R. L. Fernando, K. Kizilkaya, and D. J. Garrick,
287 2011 Extension of the bayesian alphabet for genomic selec-
288 tion. *BMC bioinformatics* **12**: 186.
- 289 Jia, Y. and J.-L. Jannink, 2012 Multiple-Trait Genomic Selection
290 Methods Increase Genetic Value Prediction Accuracy. *Genetics*
291 **192**: 1513–1522.
- 292 Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard, 2001 Pre-
293 diction of Total Genetic Value Using Genome-Wide Dense
294 Marker Maps. *Genetics* **157**: 1819–1829.
- 295 Resende, M. F. R., P. Muñoz, M. D. V. Resende, D. J. Garrick, R. L.
296 Fernando, J. M. Davis, E. J. Jokela, T. A. Martin, G. F. Peter,
297 and M. Kirst, 2012 Accuracy of Genomic Selection Methods in
298 a Standard Data Set of Loblolly Pine (*Pinus taeda* L.). *Genetics*
299 **190**: 1503–1510.

300 **Appendix**

301 **Gibbs sampler algorithm for multi-trait BayesCII**

302 **Single-site Gibbs sampler for multi-trait BayesCII**

303 The full conditional distribution of β_{j1} can be written as

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

304 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 β_{j1} ,

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

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

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

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

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

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

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

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

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

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

308 The marginal full conditional distribution of δ_{j1} can be written as

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

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

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

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

$$\begin{aligned} H &= (C_{j,11}^1)^{-\frac{1}{2}} (G^{11})^{\frac{1}{2}} \exp\left(-\frac{1}{2} (\beta_{j1}^{\hat{0}2} G^{11} - \beta_{j1}^{\hat{1}2} C_{j,11}^1)\right) \\ &= \exp\left\{-\frac{1}{2} (\log C_{j,11}^1 - \beta_{j1}^{\hat{1}2} C_{j,11}^1) - \left(-\frac{1}{2} (\log G^{11} - \beta_{j1}^{\hat{0}2} G^{11})\right)\right\} \end{aligned}$$

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

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

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

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

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

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

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

Let θ denote all other parameters except β_j and δ_j , then our sampling scheme can be written as

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

The marginal full conditional distribution of δ_j can be written as

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

316 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, \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\} \\ &\propto |\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}$$

317 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 δ_j . Thus the marginal full conditional distribution of δ_j can be written as

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

where

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

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

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

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

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

For convenience, from now on let “1” denote trait k and “2” the other traits. Thus, β_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 β_{jk} and δ_{jk} is derived as below. In our sampling scheme, β_{j1} and δ_{j1} are sampled from their joint full conditional distributions, which can be written as the product of the full conditional distribution of β_{j1} given δ_{j1} and the marginal full conditional distribution of δ_j . Let θ denote all other parameters except δ_{j1} and β_{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 β_j can be written as

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

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 β_{j1} ,

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

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

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

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

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

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

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

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

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

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

324 The marginal full conditional distribution of δ_{j1} can be written as

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

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

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

326 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

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

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

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

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

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

330 Let θ denote all other parameters except β_j and δ_j , then our sampling scheme can be written as

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

331 The marginal full conditional distribution of δ_j can be written as

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

332 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, \theta) &\propto \int f(\mathbf{y} | \boldsymbol{\beta}, \mathbf{D}, \mathbf{R}) f(\boldsymbol{\beta}_j | \mathbf{G}_j) 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}_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 \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}$$

333 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 $\mathbf{r}_j' = \left(\sum_{i=1}^n \mathbf{w}_i' m_{ij} \right) \mathbf{R}^{-1} \mathbf{D}_j$.

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

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

where

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

334 The full conditional distribution of β_j is

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