

UC Davis

UC Davis Previously Published Works

Title

Genomic Prediction from Multiple-Trait Bayesian Regression Methods Using Mixture Priors

Permalink

<https://escholarship.org/uc/item/01p957z6>

Journal

Genetics, 209(1)

ISSN

0016-6731

Authors

Cheng, Hao
Kizilkaya, Kadir
Zeng, Jian
et al.

Publication Date

2018-05-01

DOI

10.1534/genetics.118.300650

Peer reviewed

Genomic Prediction from Multiple-Trait Bayesian Regression Methods Using Mixture Priors

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

^{*}Department of Animal Science, University of California Davis, California 95616, [†]Department of Animal Science, Adnan Menderes University, Aydin, 9100 Turkey, [‡]Program in Complex Trait Genomics, Institute for Molecular Bioscience, University of Queensland, St Lucia, QLD 4072, Australia, [§]School of Agriculture, Massey University, Palmerston North 4442 New Zealand, and ^{**}Department of Animal Science, Iowa State University, Ames, Iowa 50011-1050

ORCID IDs: 0000-0001-5146-7231 (H.C.); 0000-0001-8640-5372 (D.G.); 0000-0001-5821-099X (R.F.)

ABSTRACT Bayesian multiple-regression methods incorporating different mixture priors for marker effects are used widely 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 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; GenPred; Shared data resources; Genomic Selection

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 used widely 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 π 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 π and a univariate normal distribution with probability $1 - \pi$ for all marker effects, 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 “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 propose a more general class of multi-trait BayesCII and BayesB methods, where each locus can have an

Copyright © 2018 by the Genetics Society of America

doi: <https://doi.org/10.1534/genetics.118.300650>

Manuscript received December 18, 2017; accepted for publication March 2, 2018; published Early Online March 7, 2018.

Supplemental material is available online at www.genetics.org/lookup/suppl/doi:10.1534/genetics.118.300650/-/DC1.

¹Corresponding author: 2139 Meyer Hall, Department of Animal Science, University of California Davis, 1 Shield Ave., Davis, CA 95616. E-mail: qtcheng@ucdavis.edu

effect on any combination of traits. For example, in a 5-trait analysis, the restricted model only allows two situations, whereas ours allows all 32 situations. The previous restrictive multi-trait models are special cases of this general class of models. Further, our model allows the use of a single-site Gibbs sampler that requires less computing effort than some alternative Markov chain Monte Carlo approaches, especially for analyses involving many traits. Methodologies for the new models are compared to single-trait methods and the previous multi-trait methods using real and simulated data.

Materials and Methods

Multi-trait marker effects model

For simplicity of our description, but without loss of generality, 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

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

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 (coded as 0, 1, and 2), p is the number of genotyped loci, $\boldsymbol{\alpha}_j$ is a vector of allele substitution effects or marker effects of t traits for locus j , and \mathbf{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, \mathbf{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, $W_t^{-1}(\mathbf{S}_e, \nu_e)$.

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

Multi-trait BayesCII 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 :

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

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

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

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 δ_{jk} is an indicator variable indicating whether the marker effect of locus j for trait k is zero or nonzero, 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 combinations 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 nonzero probability. Suppose, in general, we use numerical labels “1,” “2,” ..., “ t ” for the 2^t possible outcomes for $\boldsymbol{\delta}_j$, then the prior for $\boldsymbol{\delta}_j$ is a categorical distribution

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

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

As shown below, we consider two Gibbs samplers to draw samples for all the parameters in this model. Gibbs sampler I is a single-site sampler, where only one of the t indicator labels is sampled at a time. Thus, in a 2-trait model, for example, this sampler cannot move from (0, 0) to (1, 1) in a single step without stepping through (1, 0) or (0, 1) for $\boldsymbol{\delta}_j$. Therefore, Gibbs sampler I cannot be used for the restrictive model which excludes (1, 0) and (0, 1) from the state space for $\boldsymbol{\delta}_j$. Gibbs sampler II, however, samples all elements of $\boldsymbol{\delta}_j$ jointly, and can move from (0, 0) to (1, 1) in a single step. However, Gibbs sampler II is computationally more intensive because it requires drawing samples from a multivariate normal distribution of order t , the number 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, 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 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 distributions of β_{j1} , δ_{j1} , $\mathbf{\Pi}$, \mathbf{G} and \mathbf{R} for Gibbs sampler I, whose derivations are in the Appendix, are given below. The full conditional distributions of β_{j1} is

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

with

$$\hat{\beta}_{j1}^0 = - (G^{11})^{-1} \mathbf{G}^{12} \beta_{j2},$$

$$\hat{\beta}_{j1}^1 = (C_{j,11}^1)^{-1} (r_{j1} - \mathbf{C}_{j,12}^1 \beta_{j2}),$$

$$C_{j,11}^1 = G^{11} + R^{11} \sum_{i=1}^n m_{ij}^2$$

$$C_{j,12}^1 = G^{12} + R^{12} D_{j2} \sum_{i=1}^n m_{ij}^2,$$

$$r_{j1} = \left(\sum_{i=1}^n \mathbf{w}'_i m_{ij} \right) \begin{bmatrix} R^{11} \\ R^{21} \end{bmatrix},$$

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

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

where $H =$

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

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

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

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 \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 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 Π_i are near zero, the chain might exhibit mixing problems. Another, more general, but computationally intensive, Gibbs sampler that samples all elements of δ_j jointly and may exhibit improved mixing is proposed below.

The full conditional distributions of β_j , δ_j , $\mathbf{\Pi}$, \mathbf{G} , and \mathbf{R} for Gibbs sampler II, whose derivations are in the Appendix, are given below.

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

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

The full conditional distribution of β_j is

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

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

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

$$f(\delta_j = "i" | \theta, \mathbf{y}) = \frac{f(\mathbf{y} | \delta_j = "i", \theta) f(\delta_j = "i" | \mathbf{\Pi})}{\sum_{i' \in \{ "1", "2", \dots, "T" \}} f(\mathbf{y} | \delta_j = "i'", \theta) f(\delta_j = "i'" | \mathbf{\Pi})},$$

where

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

This Gibbs sampler can accommodate the "restrictive" multi-trait BayesCII that was proposed by Jia and Jannink (2012), which only allows δ_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 β_j is a multivariate t distribution, rather than a multivariate normal distribution. This is equivalent to assuming β_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 G_j are similar to the multi-trait BayesCII model. The full conditional distribution for G_j , the covariance matrix for β_j , is an inverse Wishart distribution, $W_t^{-1}(S_\beta + \beta_j\beta_j', \nu_\beta + 1)$.

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, 4828 SNPs on 807 individuals with deregressed phenotypes and genotypes on both traits remained.

Prediction accuracy was calculated as the correlation between the vector of deregressed phenotypes and the vector of estimated breeding values. Cross-validation using 10 folds formed the basis for comparing these methods. Paired *t*-tests were used for tests of significance of difference in prediction accuracies between two methods, where prediction accuracies for two different methods from each validation fold were considered as paired samples. The general multi-trait BayesCII model (MT-BayesCII-G) was compared to a similar model where the prior for α_j is a multivariate normal rather than a mixture of multivariate normals (MT-BayesC0), the more restricted multi-trait BayesCII proposed by Jia and Jannink (2012) (MT-BayesCII-R) and the usual single trait formulations of the mixture models (ST-BayesC0, ST-BayesC π). Since BayesC0 is equivalent to random regression best linear unbiased prediction (RR-BLUP), ST-BayesC0 and MT-BayesC0 are denoted as ST-RR-BLUP and MT-RR-BLUP below. The prior for the residual covariance matrix R in all multi-trait methods was an inverse Wishart distribution, $W^{-1} = \left(\begin{bmatrix} 0.003 & 0 \\ 0 & 0.003 \end{bmatrix}, 6 \right)$, for which the mean of R is $\begin{bmatrix} 0.001 & 0 \\ 0 & 0.001 \end{bmatrix}$, the SD of diagonal elements are 1.4×10^{-3} , and the SD of off-diagonal elements are 0. This same prior was used for the marker effects covariance matrix G . The priors for the residual variance and marker effects variance in single-trait analyses were scaled inverted chi-squared distributions with scale parameter $S^2 = 0.0005$ and degrees of freedom $\nu = 4$, for which the mean of the prior was also 0.001. In the data analyses, multi-trait BayesB methods provided similar results as multi-trait BayesCII methods. Thus, only results from BayesCII analyses were presented below to demonstrate the superiority of our multi-trait methods.

Simulated data: Simulated data described below were used to quantify the superiority of the general multi-trait Bayesian methods. Two scenarios were simulated. In scenario 1, as a known ideal condition, the simulated genome consisted of 100 loci on each of two chromosomes that were in Hardy-Weinberg and linkage equilibria. All these loci were considered as QTL or causative variants and used in the analyses. The QTL on the first chromosome had effects only on trait 1 and those on the second chromosome only on trait 2. The effects of these QTL were simulated from a standard normal distribution and then were equally scaled to provide unit genetic variance for each trait in the simulated population of 8000 unrelated individuals. The phenotypes for these traits were obtained by adding independent residuals to the genetic values. Two situations were simulated: (1) heritabilities for both traits were 0.5; (2) heritability for trait 1 was 0.2 and for trait 2 was 0.8. The XSim package was used in the simulation (Supplemental Material, File S1) (Cheng *et al.* 2015a).

In scenario 2, both markers and QTL were simulated. The simulated genome consisted of 100 evenly spaced loci on each of three chromosomes of length 10 cM. Ten loci were randomly selected on each chromosome as QTL. Allele states were sampled from a Bernoulli distribution with frequency 0.5 in the base population. Starting from a base population of 500 males and 500 females, random mating was simulated for 500 generations to generate linkage disequilibrium. Random mating was continued for five more generations to increase the population size to 4000 males and 4000 females, which were used in the analyses. The effects of QTL on the first two chromosomes were simulated following the same strategy in scenario 1, *i.e.*, the QTL on the first chromosome had effects only on trait 1, and those on the second chromosome only on trait 2. All QTL on the third chromosome had effects on both traits. The effects of these QTL on the third chromosome were simulated from a standard bivariate normal distribution with correlation 0.5. The phenotypes for these traits were obtained by adding independent residuals to the genetic values. In total, 8000 individuals were simulated with heritability 0.2 for trait 1 and 0.8 for trait 2.

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

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

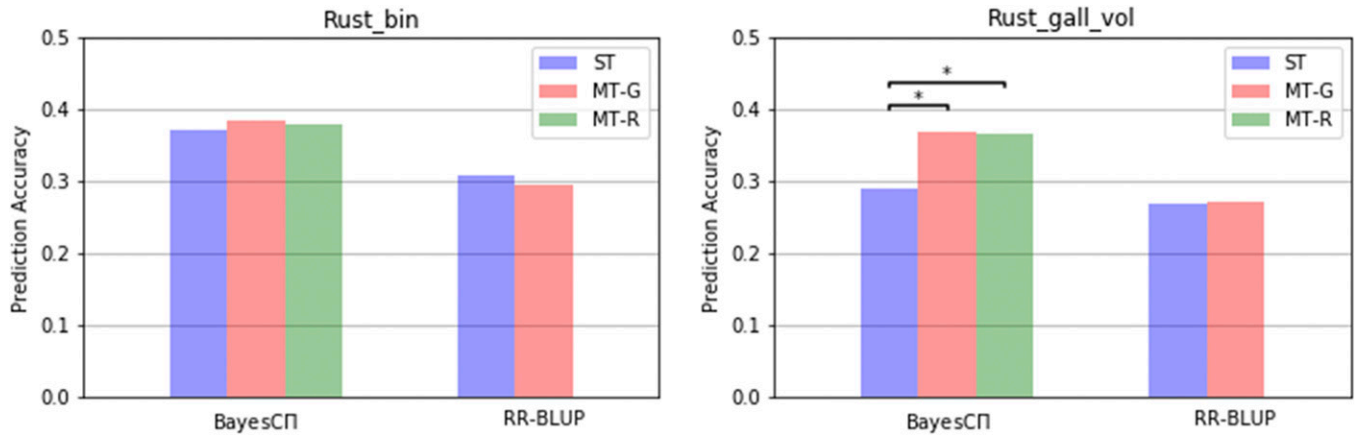


Figure 1 Comparison of single-trait and multi-trait methods for *Rust_bin* and *Rust_gall_vol* traits. ST, MT-G and MT-R indicate single-trait, our general multi-trait and restricted multi-trait analyses, respectively. * indicates a statistically significant ($P < 0.01$) difference between methods.

Data availability

The genotypic and phenotypic data used in the real data analysis are publicly available (Resende *et al.* 2012). The scripts used to generate the simulated data are provided as supplementary information. The authors state that all data necessary for confirming the conclusions presented in the article are represented fully within the article.

Results

Real data

The prediction accuracies from all methods for *Rust_bin* and *Rust_gall_vol* are in Figure 1. The prediction accuracies from all single-trait analyses using JWAS are similar to those in Resende *et al.* (2012).

The predictions of *Rust_bin* exhibited no significant difference in accuracy between multi-trait and single-trait analyses within each method (ST-RR-BLUP vs. MT-RR-BLUP; ST-BayesCπ vs. MT-BayesCπ-R; ST-BayesCπ vs. MT-BayesCπ-G).

In contrast, prediction accuracies for the lower heritability *Rust_gall_vol* with MT-BayesCπ-G were significantly higher than those from ST-BayesCπ. MT-BayesCπ-G and MT-BayesCπ-R showed similar prediction accuracies. The posterior means of Π for both methods are in Table 1. When RR-BLUP was used for the analysis, however, the advantage of the multiple-trait analysis (MT-RR-BLUP) over the single-trait analysis (ST-RR-BLUP) for *Rust_gall_vol* was not observed.

Simulated data

The prediction accuracies from MT-BayesCπ-G and MT-BayesCπ-R methods were compared for varying size (N) of training populations under two simulation scenarios. In simulation scenario 1, Figure 2 shows the prediction accuracies where heritabilities for both traits were 0.5. Figure 3 shows the prediction accuracies where heritabilities for trait 1 and trait 2 were 0.2 and 0.8, respectively. When $N = 50$, both methods had similar prediction accuracy. For both traits, as N increased, initially, MT-BayesCπ-G became superior to

MT-BayesCπ-R, but, as expected, the accuracies of these methods asymptotically converged (Karaman *et al.* 2016). In most cases, the differences in accuracies for both traits were small. However, in Figure 3, the differences in 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-BayesCπ-G was superior to MT-BayesCπ-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 BayesCπ methods (MT-BayesCπ-G vs. ST-BayesCπ; MT-BayesCπ-R vs. ST-BayesCπ). MT-BayesCπ-G and MT-BayesCπ-R outperformed ST-BayesCπ 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-BayesCπ-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-BayesCπ and ST-BayesCπ 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-BayesCπ-G may be due to the estimation of the hyper-parameter Π . In the MT-BayesCπ-G, the mean of the posterior probability that a marker has a null effect on *Rust_gall_vol* was ~ 0.97 , calculated as the summation of posterior mean of Π for categories (0, 0) and (1, 0). The posterior mean of π , the probability that a marker has a null effect, in ST-BayesCπ for *Rust_gall_vol* was 0.74, different from the equivalent value, 0.97, in MT-BayesCπ-G shown above. Thus a ST-BayesC

Table 1 Estimation of π for alternative multi-trait BayesCII methods

	Different categories of δ			
	(0, 0)	(1, 1)	(0, 1)	(1, 0)
MT-BayesCII-G	0.966	0.029	0.002	0.003
MT-BayesCII-R	0.971	0.029	NA ^a	NA ^a

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. ^a Combinations that do not exist in the restricted model.

analysis with $\pi = 0.97$ was undertaken. Prediction accuracy from this ST-BayesC π analysis with $\pi = 0.97$ was 0.361, which was similar to the accuracy from MT-BayesCII-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-BayesCII-G and MT-BayesCII-R is that MT-BayesCII-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-BayesCII-G and MT-BayesCII-R, however, showed similar prediction accuracies. This can be explained by the estimation of Π in MT-BayesCII-G and MT-BayesCII-R in Table 1. The posterior probability means for (0, 1) and (1, 0) were almost zero in MT-BayesCII-G and for (0, 0) and (1, 1) are similar in MT-BayesCII-G and MT-BayesCII-R, suggesting that the assumption of genetic architecture whereby the same loci affect both traits as explicit in MT-BayesCII-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-BayesCII-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-BayesCII-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-BayesCII-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-BayesCII-G showed a substantial advantage over MT-BayesCII-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-BayesCII-G and MT-BayesCII-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 Figure 2 and Figure 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-BayesCII-G showed higher prediction accuracy to MT-BayesCII-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-BayesCII-G gave equal or higher prediction accuracy than MT-BayesCII-R. In addition, MT-BayesCII-R requires drawing samples from a multivariate normal distribution of order t , whereas Gibbs sampler I, which can be used for MT-BayesCII-G, requires sampling from a univariate normal. Thus, in addition to MT-BayesCII-G giving equal or better performance than MT-BayesCII-R, MT-BayesCII-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,

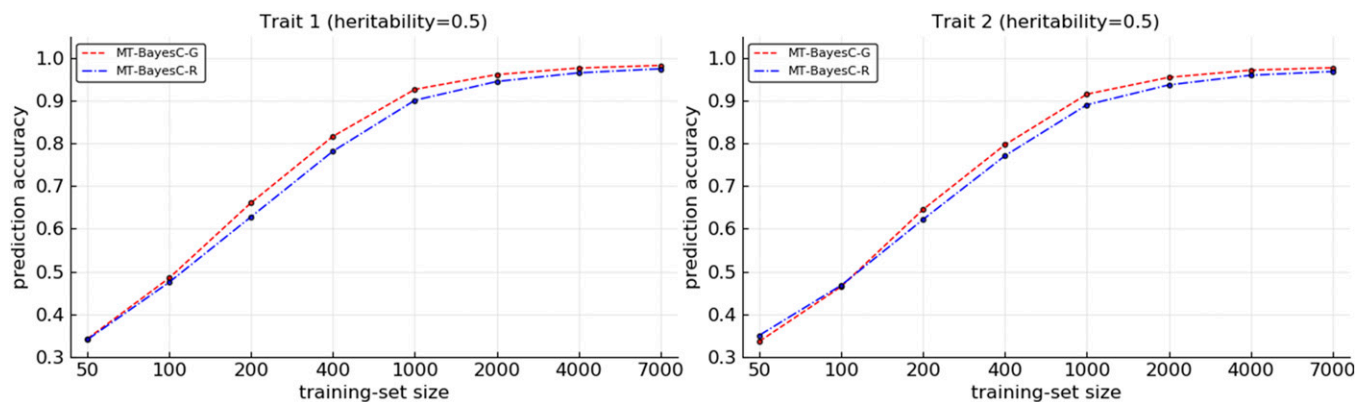


Figure 2 Comparison of multi-trait BayesCII methods for situation 1 under simulation scenario 1.

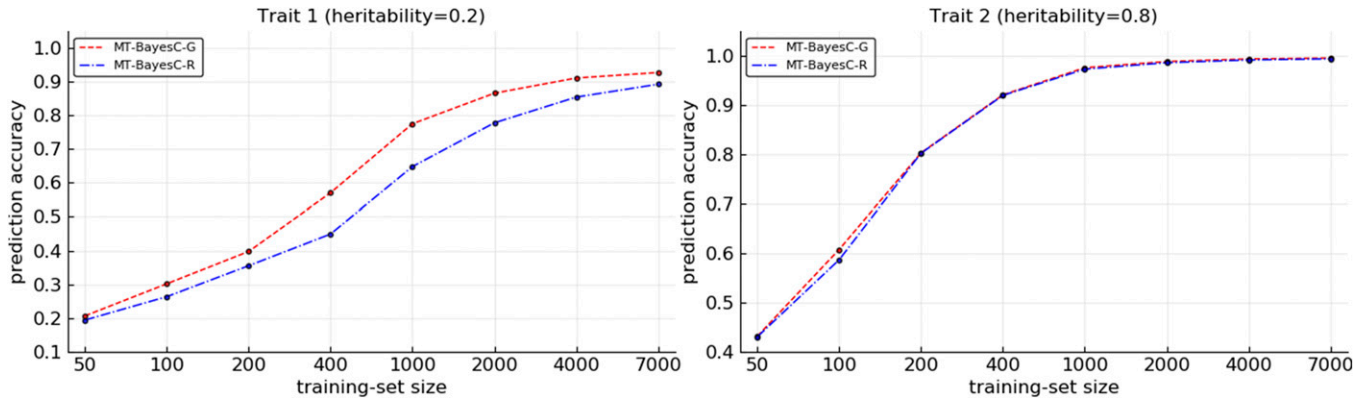


Figure 3 Comparison of multi-trait BayesCII methods for situation 2 under simulation scenario 1.

it can be shown that the marker effect variance σ_α^2 can be obtained as

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

where σ_g^2 is the genetic variance, p_j is the allele frequency for locus j , and π 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

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

where $\mathbf{Q} = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix}$ is the genetic covariance matrix and $p(\boldsymbol{\delta} = (0, 1))$, $p(\boldsymbol{\delta} = (1, 0))$, and $p(\boldsymbol{\delta} = (1, 1))$ 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 trait. Thus the probability that a marker has an effect on the first trait can be obtained as $p(\boldsymbol{\delta} = (1, 1)) + p(\boldsymbol{\delta} = (1, 0))$, 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 >2 traits. Note that positive definite matrix \mathbf{Q} may result in negative definite matrix \mathbf{G} using (3), especially when the prior for the probability a marker has null effects is far from the real value. In that case, the diagonal elements of \mathbf{G} , which are the marker effect variances for different traits, can be obtained using (2), where π may be estimated from previous single-trait analyses, and the off-diagonal elements of \mathbf{G} may be set to zero to guarantee positive definiteness of \mathbf{G} .

Multi-trait variable selection: In regard to a single trait, a locus either has an effect, or it does not. Hence, the scalar parameter π (and its complement $1 - \pi$) completely defines this circumstance. In a multi-trait setting, it is conceivable that loci that influence one trait, may or may not influence other traits. In that circumstance, a vector $\boldsymbol{\Pi}$ is required to define the genetic architecture. The number of parameters that constitute the vector $\boldsymbol{\Pi}$ is 2^t , which grows rapidly with the number of traits. In most cases, the researcher will have little or no knowledge of the likely extent of pleiotropy of loci that influence two traits, other than knowing or having an

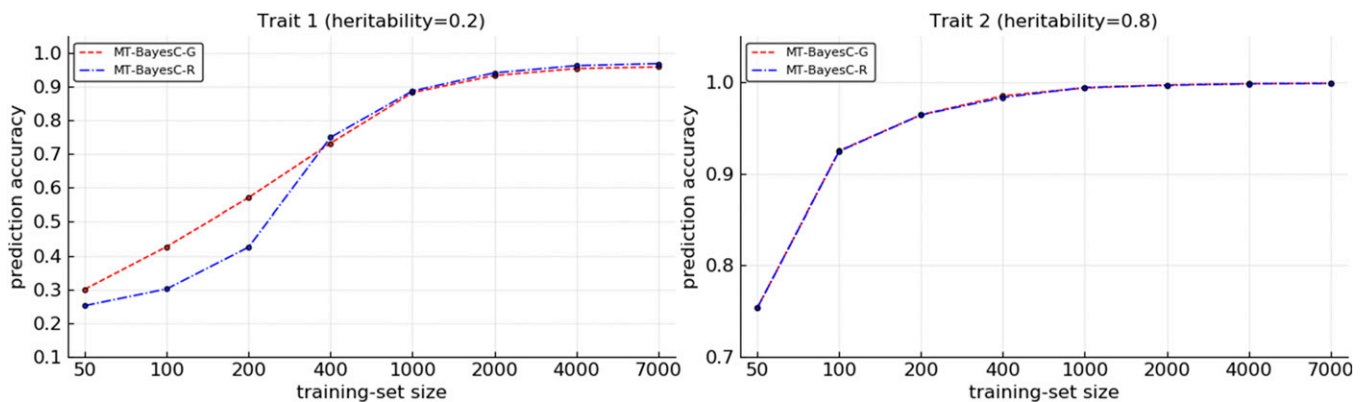


Figure 4 Comparison of multi-trait BayesCII methods under simulation scenario 2.

estimate of the genetic covariance. There are two simple ways to reduce this complexity in priors.

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

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

The better way to solve this problem is to use a hyperparameter Π that completely defines the alternative models that are required to capture all the alternative forms of genetic architecture. We have shown here how this can be done, with two alternative Gibbs sampling strategies. One involves single-site sampling for one locus and trait at a time. The other samples all the alternative combinations of effects for one locus considering all traits simultaneously. We have shown that both are practical with real data and can result in improved accuracies of prediction in certain circumstances in terms of genetic architecture and size of dataset.

Conclusions

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

Acknowledgments

We thank bioRxiv for making this manuscript available early online as preprints. This work was supported by the United

States Department of Agriculture, Agriculture and Food Research Initiative National Institute of Food and Agriculture Competitive grant no. 2015-67015-22947.

Literature Cited

- Calus, M. P., and R. F. Veerkamp, 2011 Accuracy of multi-trait genomic selection using different methods. *Genet. Sel. Evol.* 43: 26. <https://doi.org/10.1186/1297-9686-43-26>
- Cheng, H., D. Garrick, and R. Fernando, 2015a XSim: simulation of descendants from ancestors with sequence data. *G3* 5: 1415–1417. <https://doi.org/10.1534/g3.115.016683>
- Cheng, H., L. Qu, D. J. Garrick, and R. L. Fernando, 2015b A fast and efficient Gibbs sampler for BayesB in whole-genome analyses. *Genet. Sel. Evol.* 47: 80. <https://doi.org/10.1186/s12711-015-0157-x>
- Cheng, H., R. L. Fernando, and D. J. Garrick, 2018 JWAS: Julia implementation of whole-genome analysis software. *Proceedings of the World Congress on Genetics Applied to Livestock Production*, 11.859. Auckland, New Zealand.
- Daetwyler, H. D., M. P. L. Calus, R. Pong-Wong, G. de los Campos, and J. M. Hickey, 2013 Genomic prediction in animals and plants: simulation of data, validation, reporting, and benchmarking. *Genetics* 193: 347–365. <https://doi.org/10.1534/genetics.112.147983>
- Fernando, R. L., and D. Garrick, 2013 Bayesian methods applied to GWAS, pp. 237–274 in *Genome-Wide Association Studies and Genomic Prediction*. Humana Press, Totowa, NJ. https://doi.org/10.1007/978-1-62703-447-0_10
- Gianola, D., G. de los Campos, W. G. Hill, E. Manfredi, and R. Fernando, 2009 Additive genetic variability and the Bayesian alphabet. *Genetics* 183: 347–363. <https://doi.org/10.1534/genetics.109.103952>
- Habier, D., R. L. Fernando, and J. C. M. Dekkers, 2007 The impact of genetic relationship information on genome-assisted breeding values. *Genetics* 177: 2389–2397.
- Habier, D., R. L. Fernando, K. Kizilkaya, and D. J. Garrick, 2011 Extension of the Bayesian alphabet for genomic selection. *BMC Bioinformatics* 12: 186. <https://doi.org/10.1186/1471-2105-12-186>
- Jia, Y., and J.-L. Jannink, 2012 Multiple-trait genomic selection methods increase genetic value prediction accuracy. *Genetics* 192: 1513–1522. <https://doi.org/10.1534/genetics.112.144246>
- Karaman, E., H. Cheng, M. Z. Firat, D. J. Garrick, and R. L. Fernando, 2016 An upper bound for accuracy of prediction using GBLUP. *PLoS One* 11: e0161054. <https://doi.org/10.1371/journal.pone.0161054>
- Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard, 2001 Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157: 1819–1829.
- Resende, M. F. R., P. Muñoz, M. D. V. Resende, D. J. Garrick, R. L. Fernando *et al.*, 2012 Accuracy of genomic selection methods in a standard data set of Loblolly Pine (*Pinus taeda* L.). *Genetics* 190: 1503–1510. <https://doi.org/10.1534/genetics.111.137026>

Communicating editor: M. Calus

Appendix

Gibbs Sampler Algorithm for Multi-Trait BayesCII-G

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

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

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

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 β_{j1} ,

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

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 when $\delta_{j1} = 0$,

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

When $\delta_{j1} = 1$,

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

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

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

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

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

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

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

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

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

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 $f(\mathbf{y} | \delta_{j1} = 1, \boldsymbol{\theta}) / f(\mathbf{y} | \delta_{j1} = 0, \boldsymbol{\theta})$ becomes

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

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

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

where $\Pi_{j0} = Pr(\boldsymbol{\delta}_{j1} = 0, \boldsymbol{\delta}_{j2}|\boldsymbol{\Pi})$ and $\Pi_{j1} = Pr(\boldsymbol{\delta}_{j1} = 1, \boldsymbol{\delta}_{j2}|\boldsymbol{\Pi})$.
The full conditional distribution for $\boldsymbol{\Pi}$ can be written as

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

where n_i is the number of markers with $\boldsymbol{\delta}_j = "i"$.

Joint Gibbs sampler for multi-trait BayesCII-G

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

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

The marginal full conditional distribution of $\boldsymbol{\delta}_j$ can be written as

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

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}|\boldsymbol{\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 \left(\mathbf{w}_i - m_{ij} \mathbf{D}_j \boldsymbol{\beta}_j \right)' \mathbf{R}^{-1} \left(\mathbf{w}_i - m_{ij} \mathbf{D}_j \boldsymbol{\beta}_j \right) \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[\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) + \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} \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] 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}$$

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

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

where

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

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

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

Gibbs Sampler Algorithm for Multi-Trait BayesB

Single-site Gibbs sampler for multi-trait BayesB

For convenience, from now on let “1” denote trait k and “2” the other traits. Thus, $\boldsymbol{\beta}_j$ can be denoted as $\begin{bmatrix} \beta_{j1} \\ \boldsymbol{\beta}_{j2} \end{bmatrix}$ and \mathbf{D}_j can be denoted as $\begin{bmatrix} \delta_{j1} & 0 \\ 0 & \mathbf{D}_{j2} \end{bmatrix}$. The Gibbs sampler for 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 $\boldsymbol{\theta}$ denote all other parameters except δ_{j1} and β_{j1} , then our sampling scheme can be written as

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

The full conditional distribution of β_j can be written as

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

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 β_{j1} ,

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

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

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

When $\delta_{j1} = 1$,

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

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

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

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

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

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

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

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

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

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 $f(\mathbf{y} | \delta_{j1} = 1, \boldsymbol{\theta}) / f(\mathbf{y} | \delta_{j1} = 0, \boldsymbol{\theta})$ becomes

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

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

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

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

Joint Gibbs sampler for multi-trait BayesB

Let $\boldsymbol{\theta}$ denote all other parameters except $\boldsymbol{\beta}_j$ and δ_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 δ_j can be written as

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

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}|\boldsymbol{\delta}_j, \boldsymbol{\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}$$

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

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

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

where

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

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

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