

HIERARCHICAL EXTENSIONS OF BAYESIAN PARAMETRIC MODELS FOR
WHOLE GENOME PREDICTION

By

Wenzhao Yang

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Animal Science – Doctor of Philosophy
Quantitative Biology – Doctor of Philosophy

2014

ABSTRACT

HIERARCHICAL EXTENSIONS OF BAYESIAN PARAMETRIC MODELS FOR WHOLE GENOME PREDICTION

By

Wenzhao Yang

Whole genome prediction (WGP) is increasingly used to predict breeding values (BV) of plants and animals based on the use of single nucleotide polymorphism (SNP) marker panels. Two particularly popular WGP models, labeled BayesA and BayesB, are based on specifying all SNP-associated effects to be independent of each other. In this dissertation, we further extend these two models to allow for greater flexibility to infer upon BV and SNP effects in three different frameworks: 1) allowing for correlated SNP effects, 2) reaction norm modeling of genotype by environment interaction ($G \times E$) and 3) bivariate WGP models. We complement these efforts with focusing on strategies to infer upon key hyperparameters that anchor some of these specifications.

Based on a first order nonstationary antedependence specification, we extended BayesA and BayesB to account for spatial correlation between SNP effects due to the proximal QTL; we label the corresponding extensions as ante-BayesA and ante-BayesB respectively. Using simulation studies and application to the publicly available heterogeneous stock mice data and other provided benchmark data, we determined that antedependence models had significantly higher WGP accuracies compared to their conventional counterparts, especially at higher LD levels. Subsequently, we extended reaction norm (RN) and random regression (RR) models to account for $G \times E$. Several

specifications on the SNP-specific variance-covariance matrices (VCV) of intercept and slope effects were considered using independent inverted Wishart (IW) prior densities (IW-BayesA, IW-BayesB and IW-BayesC). Two potentially more flexible RR/RN models using square root free Cholesky decomposition (CD) were proposed (CD-BayesA and CD-BayesB). Based on a RN simulation study and a RR data analysis in pigs, RR/RN WGP models provided greater WGP accuracies compared to conventional WGP models although differences were not substantial between the competing IW- vs CD-based methods except with simpler genetic architectures (i.e., low number of QTL). We also developed bivariate WGP models based on more or less the same specifications for SNP-specific VCV in RR/RN models (i.e., IW-BayesA, CD-BayesA and CD-BayesB) comparing them to the more conventional bivariate genomic BLUP (bGBLUP) model. Using a LD simulation study, the three bivariate trait models generally demonstrated higher WGP accuracy than univariate BayesA or BayesB when the number of pleiotropic QTL was relatively large and the heritability of the trait was low. Furthermore, in an application to data from pine trees, CD-BayesB exhibited higher predictive ability compared to other competing models. Comparisons between competing WGP models require appropriate tuning of key hyperparameters. Hence we also studied three alternative Metropolis-Hastings (MH) sampling strategies to infer upon key hyperparameters in BayesA and BayesB. Both simulation studies and application to the heterogeneous stock mice data, strategies that were more heavily based on Metropolis Hastings sampling of key hyperparameters demonstrated significantly greater computational efficiencies compared to strategies that deferred to usage of Gibbs sampling.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Robert Tempelman, for taking me as his student and guiding me throughout my PhD research project with encouragement and patience. Dr. Robert Tempelman not only offered me a great opportunity to work with him but also helped me developing critical thinking. I greatly appreciate his guidance and mentorship he provided to me over the years.

I want to thank my guidance committee members for their constructive and thought-provoking suggestions on my research work. Dr. Juan Steibel, who was my second reader, provided me important data resource for my research project. Dr. Cathy Ernst helped me understanding genetics and guided me for my dual major. Dr. Yuehua Cui inspired me to learn quantitative genetics and Dr. Qing Lu gave me great input from an epigenetics perspective.

Furthermore, I would like to thank United States Department of Agriculture (USDA), Animal Science department and Quantitative Biology program for sponsoring my research project. I am also grateful to my colleagues Nora Bello, Chunyu Chen, Heng Wang, Lei Zhou, Igseo Choi, Yvonne Badke, Jose Luis Gualdron and Pablo Reeb for their feedback and friendship.

Last but not the least, I'd like to express my thanks to my husband and my parents for their unconditional love and support.

TABLE OF CONTENTS

| | |
|--|------|
| LIST OF TABLES | viii |
| LIST OF FIGURES | ix |
| Chapter 1 Introduction | 1 |
| Chapter 2 A Bayesian antedependence model for whole genome prediction | 5 |
| 2.1 Background | 5 |
| 2.2 Materials and Methods | 7 |
| 2.2.1 Conventional WGP model | 7 |
| 2.2.2 Antedependence extensions of WGP models | 9 |
| 2.2.3 Simulation study | 12 |
| 2.2.4 Application to Heterogeneous Stock Mice Dataset | 14 |
| 2.2.5 Application on Simulated Genomic Data from Hickey and Gorjanc | 16 |
| 2.2.6 Bayesian inference | 17 |
| 2.2.7 Prior specifications | 17 |
| 2.3 Results | 18 |
| 2.3.1 Simulation Study | 18 |
| 2.3.2 Application to Heterogeneous Stock Mice data | 28 |
| 2.3.3 Application to Hickey and Gorjanc Data | 28 |
| 2.4 Discussion and Conclusion | 30 |
| Chapter 3 Improving the computational efficiency of fully Bayes inference and assessing the effect of misspecification on hyperparameters in whole genome prediction model | 37 |
| 3.1 Introduction | 37 |
| 3.2 Materials and Methods | 40 |
| 3.2.1 WGP Model | 40 |
| 3.2.2 Univariate Metropolis Hastings sampling on ν and Gibbs update on s^2 (DFMH) | 41 |
| 3.2.3 Univariate Metropolis Hastings sampling for each of ν and s^2 (UNIMH) | 42 |
| 3.2.4 Bivariate Metropolis Hastings sampling on ν and s^2 (BIVMH) | 43 |
| 3.2.5 Simulation Study | 44 |
| 3.2.6 Data Application: Assessment of computational efficiency comparisons | 47 |
| 3.3 Results | 48 |
| 3.3.1 Simulation Study | 48 |
| 3.3.2 Application to Heterogeneous Stock Mice data | 60 |
| 3.4 Discussion | 60 |
| 3.5 Conclusions | 63 |

| | |
|---|-----|
| Chapter 4 Random regression and reaction norm extensions of whole genome prediction models to account for genotype by environment interaction | 64 |
| 4.1 Introduction | 64 |
| 4.2 Materials and Methods | 67 |
| 4.2.1 Random regression and reaction norm models..... | 67 |
| 4.2.2 Conventional BayesA and BayesB (BayesA\BayesB) | 69 |
| 4.2.3 Bivariate Normality (IW-BayesC)..... | 69 |
| 4.2.4 Bivariate Student t and Variable Selection (IW-BayesA\IW-BayesB) | 70 |
| 4.2.5 Cholesky decomposition specifications (CD-BayesA\CD-BayesB) | 71 |
| 4.2.6 Bayesian inference..... | 73 |
| 4.2.7 Simulation Study | 73 |
| 4.2.8 MSU Pig Resource Population data | 78 |
| 4.2.9 Priors used for data analyses | 81 |
| 4.3 Results | 82 |
| 4.3.1 Simulation Study | 82 |
| 4.3.2 MSU Pig Resource Population data | 87 |
| 4.4 Discussion | 94 |
| 4.5 Conclusions | 101 |
| Chapter 5 Exploring alternative specifications for bivariate trait whole genome prediction models | 102 |
| 5.1 Introduction | 102 |
| 5.2 Methods and Materials | 105 |
| 5.2.1 Whole genome prediction models | 105 |
| 5.2.2 Univariate BayesA and BayesB (uBayesA\uBayesB)..... | 106 |
| 5.2.3 Bivariate Ridge regression (bGBLUP)..... | 106 |
| 5.2.4 Bivariate Student- t (IWBayesA) | 107 |
| 5.2.5 Cholesky decomposition specifications (CDBayesA\CDBayesB)..... | 108 |
| 5.2.6 Bayesian inference..... | 111 |
| 5.2.7 Simulation studies..... | 111 |
| 5.2.8 Pine data analyses | 115 |
| 5.2.9 Priors used for data analyses | 117 |
| 5.3 Results | 118 |
| 5.3.1 Simulation Studies | 118 |
| 5.3.2 Pine data analyses | 121 |
| 5.4 Discussion | 124 |
| 5.5 Conclusions | 128 |
| Chapter 6 Discussion, Conclusions and Future Work..... | 129 |
| APPENDICES..... | 133 |
| APPENDIX A: Chapter 2 | 134 |
| APPENDIX B: Chapter 3..... | 164 |
| APPENDIX C: Chapter 4..... | 184 |
| APPENDIX D: Chapter 5 | 211 |

| | |
|--------------------|-----|
| BIBLIOGRAPHY | 242 |
|--------------------|-----|

LIST OF TABLES

| | |
|---|-----|
| Table 2.1: Summary statistics for 6 different marker densities in the simulation study over 20 replicates | 19 |
| Table 4.1: Summary of six scenarios in LD simulation..... | 74 |
| Table 5.1: Summary of two different populations compared in a LD simulation study. | 112 |

LIST OF FIGURES

| | |
|---|----|
| Figure 2.1: Average posterior means of s_g^2 (BayesA, BayesB) and s_δ^2 (ante-BayesA, ante-BayesB) across 20 replicates for six different levels of LD comparing BayesA and ante-BayesA in (A) and BayesB versus ante-BayesB in (B). Significant differences in posterior means between competing methods at each LD level are indicated by $*(P<0.01)$, $**(P<0.001)$, or $*** (P<0.0001)$ | 21 |
| Figure 2.2: Average posterior means of π_g (BayesB) versus π_δ (ante-BayesB) across 20 replicates as a function of six different LD levels. Significant differences in posterior means between competing methods at each LD level are indicated by $*(P<0.01)$, $**(P<0.001)$, or $*** (P<0.0001)$ | 22 |
| Figure 2.3: Average posterior means of v_g (BayesA, BayesB) and v_δ (ante-BayesA, ante-BayesB) across 20 replicates for six different levels of LD comparing BayesA and ante-BayesA in (A) and BayesB versus ante-BayesB in (B). No significant differences ($P>.01$) were determined between the two sets of competing procedures at each LD level. | 24 |
| Figure 2.4: Average accuracies of estimated breeding value across 20 replicates for analyses based on each of six LD levels. Differences in accuracy between BayesA and ante-BayesA (bottom symbols) and between BayesB with ante-BayesB (top symbols) indicated as significant by $*(P<.01)$ (*) or $**(P<.001)$ | 26 |
| Figure 2.5: Boxplots of average accuracies of estimated breeding value across 9 replicates for four traits in Generations 6, 8 and 10 for benchmark data from Hickey and Gorjanc (2012). Differences in accuracy between anteBayesB (black) and BayesB (dark gray) and between anteBayesA (light gray) with BayesA (white) indicated as significant by $*(0.05<P<.10)$, $**(0.01<P<.05)$ or $*** (P<.01)$ | 29 |
| Figure 3.1: Boxplots of effective sample size for v divided by total CPU time in seconds across 15 replicates for three different levels of LD comparing DFMH, UNIMH and BIVMH under BayesA model. Different letters indicate significant difference with $P<0.05$ | 50 |
| Figure 3.2: Boxplots of effective sample size for s^2 divided by total computational time in seconds across 15 replicates for three different levels of LD comparing DFMH, UNIMH and BIVMH under BayesA model. Different letters indicate significant difference with $P<0.05$ | 50 |

Figure 3.3: Boxplots of effective sample size for ν divided by total computational time in seconds across 15 replicates for three different levels of LD comparing DFMH, UNIMH and BIVMH under BayesB model. Different letters indicate significant difference with $P < 0.05$ 51

Figure 3.4: Boxplots of effective sample size for s^2 divided by total computational time in seconds across 15 replicates for three different levels of LD comparing DFMH, UNIMH and BIVMH under BayesB model. Different letters indicate significant difference with $P < 0.05$ 51

Figure 3.5: Boxplots of effective sample size for π divided by total computational time in seconds across 15 replicates for three different levels of LD comparing DFMH, UNIMH and BIVMH under BayesB model. Different letters indicate significant difference with $P < 0.05$ 52

Figure 3.6: Boxplots of accuracies of breeding value predictions under BayesA model across 15 replicates at medium marker density (pairwise $r^2=0.24$) for s^2 set equal to different multiples of average posterior mean of s^2 ($\bar{s}_{med}^2=7 \times 10^{-4}$) in a fully Bayes analysis. Significant differences with 1 \bar{s}_{med}^2 are indicated by $*(P < 0.01)$, $*** (P < 0.0001)$ 54

Figure 3.7: Boxplots of accuracies of breeding value predictions under BayesB model across 15 replicates at medium marker density (pairwise $r^2=0.24$) for s^2 set equal to different multiples of average posterior mean of s^2 ($\bar{s}_{med}^2=4 \times 10^{-2}$) in a fully Bayes analysis. Significant differences with 1 \bar{s}_{med}^2 are indicated by $*(P < 0.01)$, $*** (P < 0.0001)$ 54

Figure 3.8: Boxplots of posterior mean and median for ν under BayesA model across 15 replicates at medium marker density (pairwise $r^2=0.24$) for s^2 set equal to different multiples of average posterior mean of s^2 ($\bar{s}_{med}^2=7 \times 10^{-4}$) in a fully Bayes analysis. 56

Figure 3.9: Boxplots of posterior mean and median for ν under BayesB model across 15 replicates at medium marker density (pairwise $r^2=0.24$) for s^2 set equal to different multiples of average posterior mean of s^2 ($\bar{s}_{med}^2=4 \times 10^{-2}$) in a fully Bayes analysis. 57

Figure 3.10: Boxplots of posterior mean and median for π under BayesB model across 15 replicates at medium marker density (pairwise $r^2=0.24$) for s^2 set equal to different multiples of average posterior mean of s^2 ($\bar{s}_{med}^2=4 \times 10^{-2}$) in a fully Bayes analysis..... 57

Figure 3.11: Accuracy of breeding value predictions under BayesA model across 15 replicates at high and low marker densities (pairwise LD $r^2=0.32$ and 0.17) using DFMH (red), DFMH with fixed scale $s^2=7 \times 10^{-4}/4=1.75 \times 10^{-4}$ (green) at high marker density and DFMH with fixed scale $s^2=7 \times 10^{-4} \times 2.5=1.75 \times 10^{-3}$ (blue) at low marker density. Significant difference in accuracy between DFMH (red) and DFMH with $s^2=1.75 \times 10^{-4}$ (green) was found at $P < 0.05$ 59

Figure 3.12: Accuracy of breeding value predictions under BayesB model across 15 replicates at high and low marker densities (pairwise LD $r^2=0.32$ and 0.17) using DFMH (red), DFMH with fixed scale $s^2=0.04/4=0.01$ (green) at high marker density and DFMH with fixed scale $s^2=0.04 \times 2.5=0.1$ (blue) at low marker density. No significant differences in accuracy among each pair of methods were found. 59

Figure 4.1: Average accuracy of breeding value prediction for seven methods (BayesA\BayesB\IW-BayesA\IW-BayesB\IW-BayesC\CD-BayesA\CD-BayesB) at three environmental covariates in six scenarios, i.e. (1) $M_{\text{both}} = 100$, $M_{\text{int}} = 0$, $\bar{\rho}_{g_1g_2} = 0$, $h^2 = 0.5$; (2) $M_{\text{both}} = 100$, $M_{\text{int}} = 0$, $\bar{\rho}_{g_1g_2} = 0.5$, $h^2 = 0.5$; (3) $M_{\text{both}} = 100$, $M_{\text{int}} = 0$, $\bar{\rho}_{g_1g_2} = 0.8$, $h^2 = 0.5$; (4) $M_{\text{both}} = 50$, $M_{\text{int}} = 50$, $\bar{\rho}_{g_1g_2} = 0.5$, $h^2 = 0.5$; (5) $M_{\text{both}} = 50$, $M_{\text{int}} = 50$, $\bar{\rho}_{g_1g_2} = 0.5$, $h^2 = 0.2$; (6) $M_{\text{both}} = 20$, $M_{\text{int}} = 20$, $\bar{\rho}_{g_1g_2} = 0.5$, $h^2 = 0.5$; Different letters indicate significant difference at $P < 0.05$ 84

Figure 4.2: Accuracy of intercept (A) and slope (B) breeding value prediction for five RN methods (IW-BayesA\IW-BayesB\IW-BayesC\CD-BayesA\CD-BayesB) under six scenarios: i.e. (1) $M_{\text{both}} = 100$, $M_{\text{int}} = 0$, $\bar{\rho}_{g_1g_2} = 0$, $h^2 = 0.5$; (2) $M_{\text{both}} = 100$, $M_{\text{int}} = 0$, $\bar{\rho}_{g_1g_2} = 0.5$, $h^2 = 0.5$; (3) $M_{\text{both}} = 100$, $M_{\text{int}} = 0$, $\bar{\rho}_{g_1g_2} = 0.8$, $h^2 = 0.5$; (4) $M_{\text{both}} = 50$, $M_{\text{int}} = 50$, $\bar{\rho}_{g_1g_2} = 0.5$, $h^2 = 0.5$; (5) $M_{\text{both}} = 50$, $M_{\text{int}} = 50$, $\bar{\rho}_{g_1g_2} = 0.5$, $h^2 = 0.2$; (6) $M_{\text{both}} = 20$, $M_{\text{int}} = 20$, $\bar{\rho}_{g_1g_2} = 0.5$, $h^2 = 0.5$; Different letters indicate significant difference at $P < 0.05$ 86

Figure 4.3: Predictive ability for eight methods (BayesA\BayesB\RR-BLUP\IW-BayesA\IW-BayesB\IW-BayesC\CD-BayesA\CD-BayesB) from cross-validation analysis using back fat thickness in MSU Pig Resource Population data. Different letters indicate significant difference at $P < 0.0001$ 87

Figure 4.4: Estimated intercept breeding values from conventional BayesA, IW-BayesA and CD-BayesA using the complete final analyses data on back fat thickness in MSU Pig Resource Population. Reference line is $y=x$ 89

| | |
|--|-----|
| Figure 4.5: Estimated slope breeding values from IW-BayesA and CD-BayesA using the complete final analyses data on back fat thickness in MSU Pig Resource Population. Reference line is $y=x$ | 89 |
| Figure 4.6: Estimated SNP effects from conventional BayesA against marker position using the complete final analyses data on back fat thickness in MSU Pig Resource Population. | 91 |
| Figure 4.7: Estimated SNP effects from IW-BayesA against marker position using the complete final analyses data on back fat thickness in MSU Pig Resource Population when A) at week 10, B) at week 16, C) at week 22. | 92 |
| Figure 4.8: Estimated SNP effects from CD-BayesA against marker position using the complete final analyses data on back fat thickness in MSU Pig Resource Population when A) at week 10, B) at week 16, C) at week 22. | 93 |
| Figure 5.1: Accuracy of breeding value prediction for six methods (uBayesA\uBayesB\bGBLUP\IWBayesA\CDBayesA\CDBayesB) in two scenarios: A) number of QTLs for both traits = 10; B) number of QTLs for both traits = 30. Different letters indicate significant difference with $P<0.05$ | 120 |
| Figure 5.2: Average predictive ability from cross-validation using the loblolly pine data set for eight methods (uBayesA\uBayesB\bGBLUP\IWBayesA\CDBayesA1\CDBayesB1\CDBayesA2\CDBayesB2), where CDBayesA1 and CDBayesB1 were using RBIN as the first trait and RGV as the second trait; CDBayesA2 and CDBayesB2 were using RGV as the first trait and RBIN as the second trait; Different letters indicate means are different ($P<0.05$) from each other. | 122 |
| Figure 5.3: Estimated SNP effects for RGV from using 807 individuals and 2684 SNPs in Pine data set using three methods: A) uBayesA; B) IWBayesA; C) CDBayesB1. | 123 |

Chapter 1 Introduction

High throughput genotyping technologies such as single nucleotide polymorphisms (SNP) have slowly redefined genetic evaluation systems for livestock by augmenting phenotype and pedigree information with rich genomic information on economically important traits. This has been particularly true in the dairy industry (Wiggans, VanRaden et al. 2011). This process has come to be known as whole genome prediction (WGP) or genomic selection based on the seminal work of Meuwissen et al. (2001). The idea of using whole genome SNP markers in WGP has been introduced to predict phenotypes for complex traits in human genetics (Lee, van der Werf et al. 2008; de los Campos, Gianola et al. 2010). In a human skin cancer study, joint analyses of SNP markers in WGP have been demonstrated to improve prediction accuracy of genetic risk (Vazquez, de los Campos et al. 2012). Next generation sequencing technologies will only further make SNP genotyping more cost effective and hence more dense (De Donato, Peters et al. 2013), thereby creating more opportunities for researchers to identify and exploit the use of genes or quantitative trait loci (QTL) that are in potentially high linkage disequilibrium (LD) with SNP and meaningfully influence economically important traits and disease incidence rates in various populations.

There are several statistical inferential issues associated with WGP. Firstly, the number of available markers (m) is typically much greater than that of animals having phenotypic records (n) such that least-squares or likelihood-based approaches are not practical without first imposing some model selection strategy for deleting markers until $m < n$. After all, it is widely believed that only a small unknown proportion of markers

are associated with non-zero genetic effects (Meuwissen, Hayes et al. 2001). There have been a number of different modeling approaches to specify the distribution of these effects; thus far, almost all such approaches have been based on these effects being independent of each other. Gianola et al. (2003) anticipated that some adjacent SNP markers within a chromosome might have dependencies among their estimated effects. In addition, SNP marker effects might have nonstationary spatial correlations within a chromosome given their proximity to the major QTL.

The most common approach used in current WGP evaluations (VanRaden 2008; Daetwyler 2009; Misztal, Legarra et al. 2009; Misztal, Aggrey et al. 2013) is based on the use of Genomic Best Linear Unbiased Prediction or GBLUP (HENDERSON 1976) whereby all marker effects are assumed to be normally, identically and independently distributed (*NIID*) random draws. This specification helps resolve the $m \gg n$ issue since the specification of random effects facilitates a borrowing of information across markers. However, it has been speculated that the distributional assumptions in GBLUP may be too strong, depending upon the genetic architecture of the trait; i.e., the distribution of the QTL effects themselves, often believed to be non-normal (HAYES and GODDARD 2001) or the relative number of QTL to number of markers.

Meuwissen et al. (2001) introduced parametric Bayesian models labeled “BayesA” and “BayesB” to provide additional distributional flexibility, with both approaches often demonstrating better fit for WGP compared to GBLUP (Meuwissen, Hayes et al. 2001; Habier, Fernando et al. 2007; Hayes, Bowman et al. 2009). The “BayesA” model specifies marker specific genetic effects to be normally distributed with mean 0 and marker specific variances being independent random draws from a scaled inverted chi-

square distribution; in essence, the genetic effects are marginally specified to be *IID* Student t distributed (de los Campos, Hickey et al. 2012). The “BayesB” model uses this same distributional assumption as one component of a mixture distribution, the other component being a point spike at 0; i.e., no effects for those markers belonging to that component. Since then, several other “Bayesian alphabet” models have been developed as well (de los Campos, Naya et al. 2009; Verbyla, Hayes et al. 2009; Habier, Fernando et al. 2011; Wang, Ding et al. 2013); nevertheless, it has been duly noted that these developments and any such comparisons involving new models might be tainted by mis-specification or inappropriate tuning of key hyperparameters that anchor their corresponding distributional specifications (GIANOLA 2013).

SNP effects have been jointly analyzed under a multivariate WGP framework across heterogeneous environments (Burgueno, de los Campos et al. 2012) or multiple traits (CALUS and VEERKAMP 2011). For heterogeneous environments in WGP, genotype by environment interaction ($G \times E$) can be detected by modeling SNP specific intercept and slope effects of environmental covariates (Lillehammer, Hayes et al. 2009) in random regression (RR) and reaction norm (RN) models (Berry, Buckley et al. 2003; Cardoso and Tempelman 2012). For multiple traits in WGP, pleiotropic regions on the genome can be detected by modeling SNP specific pleiotropic effects in multivariate trait models (van Binsbergen, Veerkamp et al. 2012). In RR/RN models and bivariate trait models, the same prior densities can be specified on the genetic variance-covariance matrices (VCV) of the SNP specific effects. Calus and Veerkamp (2011) proposed a multiple trait BayesA model with a conjugate inverted Wishart (IW) prior on VCV. It has potential inflexibility since uncertainty in all elements of a VCV is based on a single degrees of freedom

parameter (MUNILLA and CANTET 2012). Bello et al. (2010) suggested that the square root free Cholesky decomposition (CD) of the VCV in bivariate mixed models might allow greater flexibility as uncertainty can be differentially expressed on each element of a VCV using such a parameterization.

There are three overarching goals in this dissertation as it pertains to meaningfully developing WGP models with improved accuracies. Firstly, one objective was to help potentially improve WGP accuracy by extending existing models to account for potential spatially-induced correlations between SNP effects due to the proximity of QTL (Chapter 2) as originally anticipated by Gianola et al. (2003). A second objective was to investigate computational strategies that one might use to be reasonably able to infer upon all key hyperparameters that underlie these and other more conventional WGP models in Chapter 3, including an assessment of the implications of hyperparameter misspecification. Finally, I deemed it imperative to provide greater flexibility in bivariate SNP effects modeling than currently developed for WGP models, whether for inferring upon genotype by environment interactions from a reaction norm perspective (Chapter 4) or to provide for bivariate trait WGP models (Chapter 5). I conclude this dissertation with some further concluding thoughts and areas for future research in Chapter 6.

Chapter 2 A Bayesian antedependence model for whole genome prediction

2.1 Background

Whole genome prediction (WGP) using commercially available medium to high density ($>50,000$) single nucleotide polymorphism (SNP) panels have transformed livestock and plant breeding. Typically, the allelic substitution effects of all SNP markers are jointly estimated in WGP evaluation models assuming additive inheritance and summed to predict breeding values of each individual animal based on its SNP genotypes (Meuwissen, Hayes et al. 2001). This technology will not only lead to dramatically increased rates of genetic improvement for economically important traits such as meat and milk production in livestock (Wiggans, VanRaden et al. 2011) or crop production (Lorenz, Chao et al. 2011), but would also improve predictions of genetic predisposition to human diseases for personalized medicine (de los Campos, Gianola et al. 2010).

Currently, the number (m) of available SNP markers is typically much greater than the number (n) of animals having phenotypic records. Hence, hierarchical mixed model or Bayesian approaches have been generally adopted in WGP to efficiently borrow information across these many markers by specifying their corresponding effects to be random. Following MEUWISSEN et al. (2001), these effects are typically specified to be either Gaussian or Student t -distributed (BayesA), or a mixture of either of these two densities with a point mass on zero (BayesB). When these effects are specified to be Gaussian, then best linear unbiased prediction of these effects is typically pursued because of computational tractability (VanRaden 2008; Hayes, Bowman et al. 2009); applied to WGP, this procedure is often known as GBLUP. Thus far, the distributional

specifications for these various hierarchical modeling approaches have been based on a prior assumption of independence between all such effects.

GIANOLA et al. (2003) anticipated that some of these effects might be spatially correlated within chromosomes such that greater inference efficiency might be provided by modeling these effects as correlated. Their proposed specifications required either equally spaced markers and/or within-chromosome correlations depending strictly on physical/linkage map distance between markers. However, the equally spaced assumption is rather tenuous for most currently available SNP marker panels. Even more importantly, the inferred correlation structure is likely to be nonstationary given that it should be primarily driven by the proximity of SNP markers to quantitative trait loci (QTL) of major effects. In other words, we anticipate that the correlation between the inferred effects of adjacent SNPs distal to major QTL would be substantially smaller than those proximal to these QTL.

Antedependence models have been increasingly advocated for the analysis of repeated measures data (ZIMMERMAN and NÚÑEZ-ANTÓN 2010) to parsimoniously account for nonstationary correlations between repeated measurements over time. In this paper, we develop first-order antedependence counterparts to BayesA and BayesB. Through simulation, a cross-validation study involving the publicly available heterogeneous stock mixture data (Valdar, Solberg et al. 2006; Valdar, Solberg et al. 2006) and journal-provided reference data (HICKEY and GORJANC 2012) to benchmark our proposed methods against others, we demonstrate that, compared to their conventional counterparts, these antedependence-based WGP models improve the

accuracy of genomic merit prediction as well as potentially increase the sensitivity of QTL detection, which is the key objective of genome wide association studies (GWAS).

2.2 Materials and Methods

2.2.1 Conventional WGP model

The base linear mixed model used for WGP is generally written as follows:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{g} + \mathbf{W}\mathbf{u} + \mathbf{e} \quad [1]$$

Here, $\mathbf{y} = \{y_i\}_{i=1}^n$ is a $n \times 1$ vector of phenotypes, $\boldsymbol{\beta}$ is an $p \times 1$ unknown vector of fixed effects connected to \mathbf{y} via a known $n \times p$ incidence or covariate matrix \mathbf{X} (e.g., environmental effects), $\mathbf{g} = \{g_j\}_{j=1}^m$ is an $m \times 1$ vector of random SNP effects connected to \mathbf{y} via a known $n \times m$ matrix of SNP genotypes coded as 0, 1, or 2 copies of the minor allele for each SNP (column) and animal (row) in \mathbf{Z} . Furthermore, $\mathbf{u} = \{u_k\}_{k=1}^q$ is a $q \times 1$ vector of random polygenic effects connected to \mathbf{y} via a known $n \times q$ incidence matrix \mathbf{W} , and $\mathbf{e} = \{e_i\}_{i=1}^n$ is the residual vector. We assume that $\mathbf{u} \sim \mathbf{N}(0, \mathbf{A}\sigma_u^2)$, where \mathbf{A} denotes the pedigree-derived numerator relationship matrix (HENDERSON 1976), and is often included in WGP models due to insufficient genome coverage by \mathbf{Z} (CALUS and VEERKAMP 2007). Furthermore, we specify $\mathbf{g} \sim N(0, \boldsymbol{\Sigma}_g)$ where $\boldsymbol{\Sigma}_g = \text{diag}(\sigma_{g_j}^2)$ and $\mathbf{e} \sim N(0, \mathbf{I}\sigma_e^2)$. From a Bayesian perspective, a subjective prior may be also specified on $\boldsymbol{\beta}$ using $\boldsymbol{\beta} \sim N(\boldsymbol{\beta}_0, \mathbf{V}_\beta)$ with $\boldsymbol{\beta}_0$ and \mathbf{V}_β taken as known (SORENSEN and GIANOLA 2002).

Now the distinction between GBLUP, BayesA, and BayesB in MEUWISSEN et al. (2001) depends upon the characterization of Σ_g . If $\Sigma_g = \mathbf{I}\sigma_g^2$ (i.e., $\sigma_{g_j}^2 = \sigma_g^2 \forall j$), then the model is defined to be GBLUP. If, instead, the diagonal elements of Σ_g are independent random draws from an scaled inverted chi-square distribution, i.e., $\sigma_{g_j}^2 \sim \chi^{-2}(\nu_g, \nu_g s_g^2)$ such that $E(\sigma_{g_j}^2) = \frac{\nu_g s_g^2}{\nu_g - 2}$, then the model is said to be BayesA such that marginally g_j is a random draw from a Student t distribution with mean 0, degrees of freedom ν_g and scale parameter s_g^2 (de los Campos, Naya et al. 2009; Gianola, de los Campos et al. 2009). Now BayesB further extends BayesA by including a two-component mixture with one component being $\chi^{-2}(\nu_g, \nu_g s_g^2)$ and the other component being a spike or point mass at 0; i.e.,

$$\sigma_{g_j}^2 | \nu_g, s_g^2 \begin{cases} = 0 & \text{with probability } \pi_g \\ \sim \chi^{-2}(\nu_g, \nu_g s_g^2) & \text{with probability } (1 - \pi_g) \end{cases} \quad [2]$$

That is, π_g ($0 < \pi_g < 1$) represents the proportion of SNP markers having no associated genetic effects on the trait of interest.

Clear warnings have been provided on how sensitive inferences using BayesA or BayesB may be to specification of the hyperparameters (de los Campos, Naya et al. 2009; Gianola, de los Campos et al. 2009). It has not been widely appreciated that ν_g and s_g^2 are estimable; this recognition is critical as both hyperparameters help define the genetic architecture in BayesA and BayesB. That is, ν_g characterizes the variability of $\sigma_{g_j}^2$

about a typical variance component of s_g^2 . Details on how to estimate ν_g and s_g^2 in the context of BayesA were previously provided by YI and XU (2008). Furthermore, π_g is estimable in BayesB. For both BayesA and BayesB, we specify the prior distribution $\nu_g \sim p(\nu_g) \propto (\nu_g + 1)^{-2}$, similar to what we have previously adopted in other applications (Kizilkaya and Tempelman 2005; Bello, Steibel et al. 2010). Furthermore, we specify $\pi_g \sim p(\pi_g | \alpha_\pi, \beta_\pi) = \text{Beta}(\alpha_\pi, \beta_\pi)$ for BayesB, with values of α_π and β_π chosen to reflect prior uncertainty on π_g . We also specify a proper conjugate prior on s_g^2 in BayesB; i.e., $s_g^2 \sim p(s_g^2 | \alpha_g, \beta_g) = \text{Gamma}(\alpha_g, \beta_g)$ recognizing that the specifications on α_g and β_g become increasingly influential as $\pi_g \rightarrow 1$.

Finally, we specify noninformative priors $\sigma_e^2 \sim \chi^{-2}(-1, 0)$ and $\sigma_u^2 \sim \chi^{-2}(-1, 0)$ which are congruent with specifying uniform priors on σ_e and σ_u , respectively, and in line with recommendations for variance components by GELMAN (2006). We similarly and confidently specify $s_g^2 \sim \chi^{-2}(-1, 0)$ in BayesA, given that m is generally large enough for stable inference on s_g^2 without the need for more informative priors.

2.2.2 Antedependence extensions of WGP models

We propose a nonstationary first-order antedependence correlation structure for \mathbf{g} based on the relative physical location of SNP markers along the chromosome(s):

$$g_j = \begin{cases} \delta_1 & \text{if } j = 1 \\ t_{j,j-1}g_{j-1} + \delta_j & \text{if } 2 \leq j \leq m \end{cases} \quad [3]$$

Here $\delta_j \sim NID(0, \sigma_{\delta_j}^2)$, $j = 1, \dots, m$ whereas $t_{j,j-1}$ is the marker interval specific

antependence parameter (ZIMMERMAN and NÚÑEZ-ANTÓN 2010) of g_j on g_{j-1} in the specified order. We can rewrite the recursive expression in [3] in matrix notation:

$$\mathbf{g} = \mathbf{T}\mathbf{g} + \boldsymbol{\delta} \quad [4]$$

where $\boldsymbol{\delta} = \{\delta_j\}_{j=1}^m = (\mathbf{I} - \mathbf{T})\mathbf{g}$ for \mathbf{I} being a $m \times m$ identity matrix, and \mathbf{T} having all null values except for elements $t_{j,j-1}$ at the corresponding subscript addresses. It can be readily seen using Equation [4] that $\text{var}(\mathbf{g}) = \boldsymbol{\Sigma}_g = (\mathbf{I} - \mathbf{T})^{-1} \boldsymbol{\Lambda} (\mathbf{I} - \mathbf{T})^{-1}$, where $(\mathbf{I} - \mathbf{T})^{-1}$ is a lower triangular matrix with diagonal elements equal to 1 and $\boldsymbol{\Lambda} = \text{diag}\{\sigma_{\delta_j}^2\}_{j=1}^m$. As further illustrated in File S1 from the Supporting Information, $\boldsymbol{\Sigma}_g^{-1} = (\mathbf{I} - \mathbf{T})' \boldsymbol{\Lambda}^{-1} (\mathbf{I} - \mathbf{T})$ is a readily determined tri-diagonal matrix (ZIMMERMAN and NÚÑEZ-ANTÓN 2010) which is important as it facilitates inference on \mathbf{g} .

Some of the other developments closely follow the BayesA and BayesB models of MEUWISSEN et al. (2001). That is, we specify $\sigma_{\delta_j}^2 \sim \chi^{-2}(\nu_{\delta}, \nu_{\delta} s_{\delta}^2)$ in a model which we label *ante-BayesA*. Similarly, we propose an *ante-BayesB* model whereby we specify a mixture similar to Equation [2] except that it is specified on $\sigma_{\delta_j}^2$; i.e., a mixture of point mass on zero with probability π_{δ} and scaled inverted chi-square prior $\chi^{-2}(\nu_{\delta}, \nu_{\delta} s_{\delta}^2)$ with probability $(1 - \pi_{\delta})$. As we suggested earlier for π_g , we believe that π_{δ} is estimable such

that ante-BayesA is merely a special case of ante-BayesB. In turn, BayesA is merely a special case of ante-BayesA, as is BayesB of ante-BayesB, when $\mathbf{T} = \mathbf{0}$; i.e., $t_{j,j-1} = 0 \forall j$.

These antedependence extensions, nevertheless, do require inference on the unknown $m-1$ non-zero elements $\{t_{j,j-1}\}_{j=2}^m$ of \mathbf{T} . Borrowing from DANIELS and POURAHMADI (2002) and BELLO et al. (2010), we specify $t_{j,j-1} \sim N(\mu_t, \sigma_t^2)$ as a conjugate prior in both ante-BayesA and ante-BayesB, thereby allowing flexible inference on the nonstationary correlation structure in Σ_g . However, it should be further noted that if interval $j,j-1$ specifies that between the last SNP of one particular linkage group or chromosome and the first SNP in the arbitrarily subsequent linkage group, then we set the corresponding $t_{j,j-1} = 0$. The remaining prior specifications are specified on the hyperparameters that essentially characterize the hypothesized genetic architecture of the trait and are virtually identical to those previously prescribed for BayesA and BayesB; i.e., $\nu_\delta \propto (\nu_\delta + 1)^{-2}$, $s_\delta^2 \sim \text{Gamma}(\alpha_s, \beta_s)$, $\pi_\delta \sim \text{Beta}(\alpha_\pi, \beta_\pi)$ with $\alpha_\pi, \beta_\pi, \alpha_s$ and β_s again all specified as known. Similarly, we also estimate μ_t and σ_t^2 by placing subjective priors, $\mu_t \sim N(\mu_{t0}, s_{t0}^2)$ and $\sigma_t^2 \sim \chi^{-2}(\nu_t, \nu_t s_t^2)$ on these key hyperparameters, where $\mu_{t0}, s_{t0}^2, \nu_t$ and s_t^2 are specified to be known.

As in MEUWISSEN et al. (2001) and subsequent work, our implementation strategy is based on the use of Markov Chain Monte Carlo (MCMC) methods; however, we also additionally infer key hyperparameters; i.e., $\nu_g(\nu_\delta)$, $s_g^2(s_\delta^2)$, and $\pi_g(\pi_\delta)$ that characterize the genetic architecture of the trait, as alluded to earlier. Further details on the full

conditional densities and any necessary Metropolis Hastings strategies to sample from the joint posterior density of all unknown parameters using MCMC are provided in Appendix A1.

2.2.3 Simulation study

We compare the performance of BayesA and BayesB with their antedependence counterparts, ante-BayesA and ante-BayesB, in a simulation study. Twenty replicated datasets were each generated from a base population containing 50 unrelated males and 50 unrelated females. Each dataset underwent random mating while maintaining constant population size for 6001 generations beyond the base population. The entire genome was composed of one chromosome of length 1 Morgan. All of 20,001 potential SNP markers were equally spaced on this genome with a potential QTL placed directly in the middle of each interval of adjacent markers. In the base population, all 20,000 QTL and 20,001 SNP marker alleles were coded as monomorphic. The number of simulated crossover events per meiosis was generated from a Poisson (mean = 1) distribution with the location of the crossover events uniformly distributed throughout the chromosome in accordance with the Haldane mapping function. The mutation rate for both QTL and SNP markers was specified to be 10^{-4} per locus per generation and to be recurrent, that is, switching between one of two alternative allelic states 0 and 1 whenever mutation occurred so as to ensure biallelic loci (Coster, Bastiaansen et al. 2010; Daetwyler, Pong-Wong et al. 2010).

In Generation 6001, all SNP markers and QTL with a minor allele frequency (MAF) less than 0.05 were discarded. We then randomly selected only 30 of the

remaining QTL and their corresponding allelic substitution effects. For each of these $k=1,2,\dots,30$ QTL, an allelic substitution effect (α_k) was drawn from a reflected gamma distribution with shape parameter 0.4 and scale parameter 1.66, with a positive or negative sign on α_k sampled with equal probability. The genetic variance at QTL k was determined to be $2p_k(1-p_k)\alpha_k^2$, where p_k is the MAF at QTL k . The total genetic variance was subsequently determined to be the summation of these terms across the 30 selected QTLs; i.e., as $2\sum_{k=1}^{30} p_k(1-p_k)\alpha_k^2$. Now the true breeding values (TBV) were defined to be a genotype-based linear function of the 30 generated QTL effects which, because these QTL were located between various SNP, are not subsets of \mathbf{g} . These TBV were further scaled such that the total genetic variance was 1 as per MEUWISSEN and GODDARD (2010). Residual effects were, in turn, sampled from a standard normal distribution, such that the heritability was 0.50. That is, each phenotypic record was generated by adding the TBV for that animal plus its corresponding residual. Hence, 100 animals with known phenotypes and genotypes in Generation 6001 were simulated for inferring upon the SNP effects, using each of the competing methods. Genotypes and the TBV for each of 100 offspring were also generated in Generation 6002, based on randomly mating animals in Generation 6001.

For each of the 20 replicated datasets, the effect of 6 different marker densities on the comparison between the competing methods were investigated by selecting every 1, 4, 7, 10, 15, and 20 SNP markers from those with $\text{MAF} > 0.05$. That is, the datasets were used as a blocking factor in comparing different marker densities for the accuracy of predicting genetic merit in Generation 6002, using each of the four different methods:

BayesA, BayesB, ante-BayesA and ante-BayesB. Accuracy was defined as the correlation between estimated breeding values (EBV) for Generation 6002, using just Generation 6001 phenotypes and genotypes data, and the corresponding TBV of Generation 6002. These EBV are based on the posterior mean ($\bar{\mathbf{g}}$) of \mathbf{g} ; i.e., EBV are elements of $\mathbf{Z}\bar{\mathbf{g}}$. Comparisons were also drawn between the BayesA/BayesB procedures and their antedependence counterparts for inference on the key hyperparameters that characterize genetic architecture. This was conducted using multifactorial ANOVA on the posterior means using replicate as the blocking factor for assessing the importance of model and marker density and their interaction across the 20 replicates. Furthermore, an assessment of the relative ability of ante-BayesB compared to BayesB to identify the top QTL by genetic variance was based on the difference in posterior probabilities of δ_j and g_j , respectively, of adjacent SNP markers being non-zero. As QTLs were placed between SNP markers and never on top of SNP markers, we calculated this *probability of association* by determining the proportion of MCMC cycles that either or both of the two markers adjacent to the known QTL were chosen to be non-zero within each analysis.

All comparisons were based on the linear mixed model in Equation [2] with \mathbf{X} being a column vector of ones, except that polygenic effects (\mathbf{u}) were ignored for simplicity and computational tractability.

2.2.4 Application to Heterogeneous Stock Mice Dataset

We used a dataset publicly available from the Wellcome Trust (<http://gscan.well.ox.ac.uk/>) which includes phenotypic records on 2,296 mice, each genotyped for 12,147 SNP markers. This data resource, which also includes pedigree

information, was based on an advanced intercross mating among 8 inbred strains after 50 generations of random mating (Valdar, Solberg et al. 2006). The average linkage disequilibrium (LD), as measured by r^2 between adjacent markers is 0.62 (Legarra, Robert-Granie et al. 2008), which is high compared to commonly used SNP panels available for livestock populations. For example, the average r^2 between adjacent markers in most commercially available livestock SNP panels ranges from 0.10 to 0.37 for markers that are generally around 100kb apart (Du, Clutter et al. 2007; De Roos, Hayes et al. 2008; Abasht, Sandford et al. 2009; Jarmila, Sargolzaei et al. 2010).

Given this high pairwise LD, we considered only a random subset of all markers from this dataset to ensure adjacent marker LD levels that are representative of livestock populations. We first excluded SNP markers if the percentage of missing genotypes across samples was greater than 10% or if the MAF was less than 2.5%. We also discarded animals having greater than 20% missing SNP genotypes. We then randomly selected 50 SNP markers from each of the 19 autosomes, leading to an average LD of r^2 of 0.35 between adjacent markers. The resulting dataset then involved records on 1917 animals with genotypes on 950 SNPs.

As in LEGARRA et al. (2008), we also added the random effect of cage in the WGP model of [2]; i.e., $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{g} + \mathbf{W}\mathbf{u} + \mathbf{S}\mathbf{c} + \mathbf{e}$ where $\mathbf{c} \sim N(\mathbf{0}, \mathbf{I}\sigma_c^2)$ and \mathbf{S} is the corresponding incidence matrix with all other terms defined as before. Furthermore, we specified GELMAN's prior $\sigma_c^2 \sim \chi^{-2}(-1, 0)$ on σ_c^2 in addition to all previously provided prior specifications. Also, as per LEGARRA et al. (2008), we chose to use the data provided on body weight at 6 weeks that was already pre-corrected for fixed effects such

that \mathbf{X} was a column vector of ones and $\boldsymbol{\beta}$ consisted of just an overall mean. Missing SNP genotypes were simply imputed from binary distributions based on their corresponding allelic frequencies in the dataset following LEGARRA et al. (2008). We adopted the same within-family cross validation technique as described in LEGARRA et al. (2008) by randomly partitioning each family into two. This partitioning was replicated 20 times to obtain 20 different nearly equal sized partitions of training and validation data subsets. Also, as in LEGARRA et al. (2008), we compared the various methods using predictive abilities, defined as the correlation between phenotypes in the validation subset and their corresponding predictions based on their inferences from the training data subset.

2.2.5 Application on Simulated Genomic Data from Hickey and Gorjanc

To provide a benchmark comparison of our proposed methods with competing methods in other papers in this issue, we analyze simulated datasets provided by and described in detail by Hickey and Gorjanc (2012). They generated 10 replicated datasets for each of four different traits whereby 9000 QTL effects were generated for Trait 1 and 900 QTL effects were generated for Trait 2. Traits 3 and 4 mirrored Traits 1 and 2, respectively, with the further requirement that the MAF for these QTL was less than 0.30. Since we were permitted to simultaneously run 144 jobs on the High Performance Computing Cluster at MSU (hpcc.msu.edu), we chose to compare the four methods for each of the four traits on each of the first nine datasets ($4 \times 4 \times 9 = 144$). For all analyses, training data were based on 2000 animals in Generations 4 and 5 whereas TBV were provided on 500 animals within each Generation 6, 8 and 10. To facilitate computing tractability, we saved every tenth SNP marker that had a $\text{MAF} > 0.20$. This led to a

range of 2884 to 2952 SNP markers and an average LD between adjacent markers from 0.16 to 0.17 across the nine replicates. All four models also included polygenic effects. Antedependence methods were directly compared with their classical counterparts for accuracy (correlation of EBV with TBV) and bias (deviation of slope from 1 from regressing TBV on EBV) in these latter validation generations using a Wilcoxon signed rank test.

2.2.6 Bayesian inference

For each of the four methods, BayesA, BayesB, ante-BayesA and ante-BayesB in both our simulation study and the heterogeneous stock mice application, we ran MCMC for 50,000 cycles of burn-in followed by an additional 300,000 cycles; for the benchmark data from Hickey and Gorjanc (2012), the corresponding numbers were 80,000 and 1,000,000, respectively. Every tenth MCMC cycle was subsequently saved for inference post burn-in. We monitored MCMC convergence via inspection of trace plots and determined the effective sample size (ESS) for number of random draws from the joint posterior density for all key hyperparameters using the R package CODA (Plummer, Best et al. 2006). The larger number of MCMC cycles for the Hickey-Gorjanc data were based on ensuring that ESS for all hyperparameters at least exceeded 100. Inferences were primarily based on the posterior means and posterior standard deviations for key parameters, including those hyperparameters that characterize genetic architecture.

2.2.7 Prior specifications

For all analyses in this paper, we chose $\alpha_{\pi} = 10$ and $\beta_{\pi} = 1$ in both BayesB and ante-BayesB to reflect the prior belief that most of the markers will not be associated

with any genetic effects; however, the dispersion of this corresponding beta distribution is still large enough such that values of π_g (π_δ) close to 0.70 are plausible. Based on preliminary runs, we also found that this prior specification led to superior mixing properties of the MCMC chains over a naïve Uniform(0,1) prior, yet facilitated domination of data over prior information since $\alpha_\pi + \beta_\pi \ll m$.

For BayesB and ante-BayesB, we always specified $\alpha_s = \beta_s = 0.1$ for the Gamma prior on s_g^2 (s_δ^2). For the antedependence based models, we specified $\mu_{t0} = 0$, $s_{t0}^2 = 1$, $\nu_t = -1$, and $s_t^2 = 0$; i.e., a standard normal prior on μ_t and GELMAN's prior on σ_t^2 . We also always specified a flat prior on β by defining $\mathbf{V}_\beta^{-1} = \mathbf{0}$. Prior specifications for all other parameters (e.g., variance components) were based on those previously recommended in this paper.

2.3 Results

2.3.1 Simulation Study

For the six different marker densities, the average distances between adjacent markers ranged from 0.046 to 0.918 cM over the 20 replicates whereas the average LD between adjacent markers, measured by r^2 values, ranged from 0.15 to 0.31, as shown in Table 2.1. Among the 30 chosen QTLs within each of the 20 replicates, anywhere between 6 to 11 of the QTLs had variances greater than 2% of the total genetic variance.

Table 2.1: Summary statistics for 6 different marker densities in the simulation study over 20 replicates

| Marker density level [†] | Average number of markers per replicate | Average distance between adjacent marker loci (cM) per replicate | Average r^2 between adjacent marker loci per replicate |
|-----------------------------------|---|--|--|
| 1 | 108 | 0.918 | 0.15 |
| 2 | 145 | 0.689 | 0.18 |
| 3 | 217 | 0.459 | 0.21 |
| 4 | 311 | 0.321 | 0.24 |
| 5 | 545 | 0.184 | 0.27 |
| 6 | 2182 | 0.046 | 0.31 |

[†]Marker density levels 1 through 6 pertain to saving every 20th, 15th, 10th, 7th, 4th, and every single SNP marker from a single 1M chromosome within each data replicate.

It is important to recognize that none of the modeling assumptions behind BayesA, BayesB, ante-BayesA, or ante-BayesB truly match the data generation model based on thousands of generations of LD created between markers and QTLs, even for simulated data. This goes beyond the fact that the QTL effects were drawn from reflected Gamma distributions in our simulation study as typically done (e.g., Meuwissen, Hayes et al. 2001; Meuwissen and Goddard 2010). That is, the process of recombination over thousands of generations in terms of how it generates LD between QTL and SNP markers is not explicitly captured in any known WGP model, including any of the competing models, especially when the effects of neighboring SNP markers rather than the causal

QTL effects are being estimated. Hence, there is no way to surmise the “true” values of key hyperparameters, whether for s_g^2 , ν_g , or π_g in BayesA or BayesB or for s_δ^2, ν_δ , π_δ , μ_t or σ_t^2 in ante-BayesA or ante-BayesB. However, one should anticipate that estimates of s_g^2 or s_δ^2 should be inversely related to marker density, since they closely represent the mean value of the variance components $\left\{\sigma_{g_j}^2\right\}_{j=1}^m$ or $\left\{\sigma_{\delta_j}^2\right\}_{j=1}^m$, respectively, accounted for by each SNP. Indeed, we observe this phenomena in the comparison between BayesA and ante-BayesA in Figure 2.1A. We also note a similar comparison between s_δ^2 and s_g^2 for BayesB versus ante-BayesB in Figure 2.1B, but further recognize that the corresponding estimates of s_g^2 and s_δ^2 are roughly one order of magnitude greater than those seen in Figure 2.1A. That is, s_g^2 and s_δ^2 specify a typical value for $\sigma_{g_i}^2$ and $\sigma_{\delta_i}^2$, respectively, over many more loci in (ante)BayesA than their (ante)BayesB counterparts. In spite of the lower values observed in Figure 2.1A, however, there was a significant difference ($P < 0.01$) between s_g^2 and s_δ^2 when $r^2 \geq 0.21$.

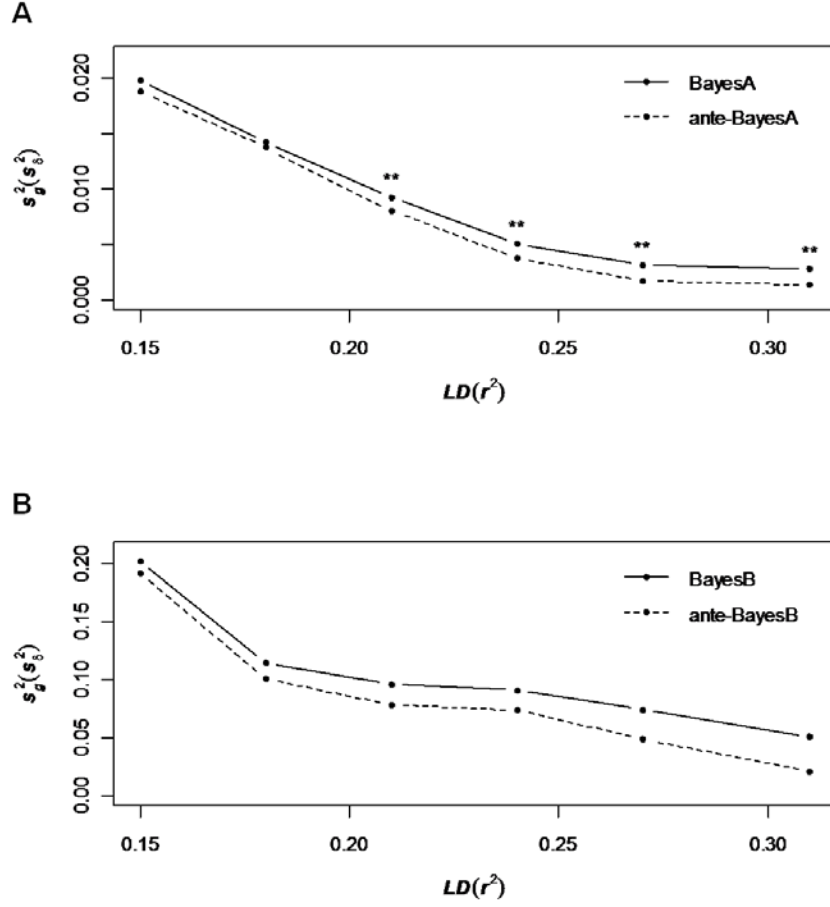


Figure 2.1: Average posterior means of s_g^2 (BayesA, BayesB) and s_d^2 (ante-BayesA, ante-BayesB) across 20 replicates for six different levels of LD comparing BayesA and ante-BayesA in (A) and BayesB versus ante-BayesB in (B). Significant differences in posterior means between competing methods at each LD level are indicated by *($P < 0.01$), **($P < 0.001$), or ***($P < 0.0001$).

As marker density increased, we also expected that the estimates of π_g or π_δ should increase as well; that is, it becomes increasingly unlikely that individual SNP markers become associated with a particular QTL with greater marker density. Indeed we observed this in Figure 2.2. It was particularly interesting that the posterior means of π_δ were generally lower than that of π_g , with differences widening with increasing marker density (i.e., LD level) such that the differences were significant beyond $r^2=0.24$ ($P<0.01$). Note the subtle difference in interpretation between π_g and π_δ as π_g pertains to the probability of non-association for the corresponding SNP whereas π_δ pertains to the probability of non-association conditional on a neighboring SNP.

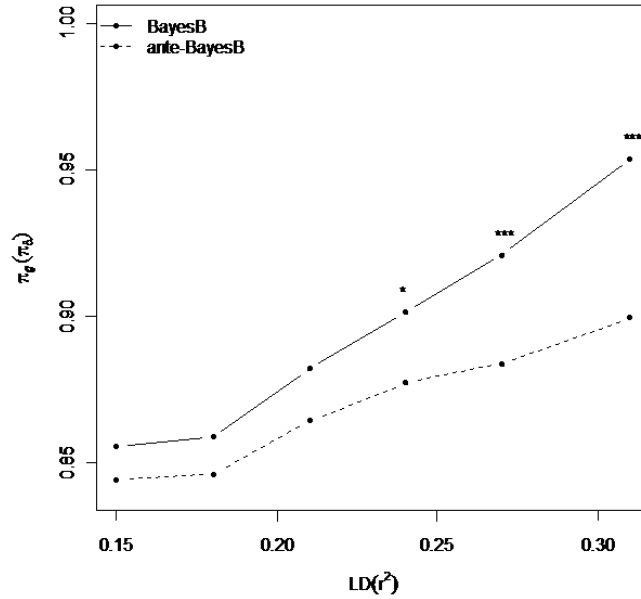


Figure 2.2: Average posterior means of π_g (BayesB) versus π_δ (ante-BayesB) across 20 replicates as a function of six different LD levels. Significant differences in posterior means between competing methods at each LD level are indicated by $^*(P<0.01)$, $^{**}(P<0.001)$, or $^{***}(P<0.0001)$.

The estimates of ν_g and ν_δ also changed as a function of marker density for ante-BayesA versus BayesA in Figure 2.3A and for ante-BayesB versus BayesB in Figure 2.3B. Specifically, the posterior means of ν_g and particularly of ν_δ both decrease with increasing marker intensity. Since these parameters, respectively, characterize the heterogeneity of $\sigma_{g_j}^2$ and $\sigma_{\delta_j}^2$ across SNP or, alternatively, the heaviness of the tails for the resulting marginal Student t distribution on g_j and δ_j across SNP, our results imply that these hierarchical methods, and particularly those based on nonstationary first order antedependence correlation structures, identify SNP with large effects as being more outlying relative to a normal distribution when marker density increases. However, these differences between ν_g and ν_δ were not seen to be statistically significant at any marker density.

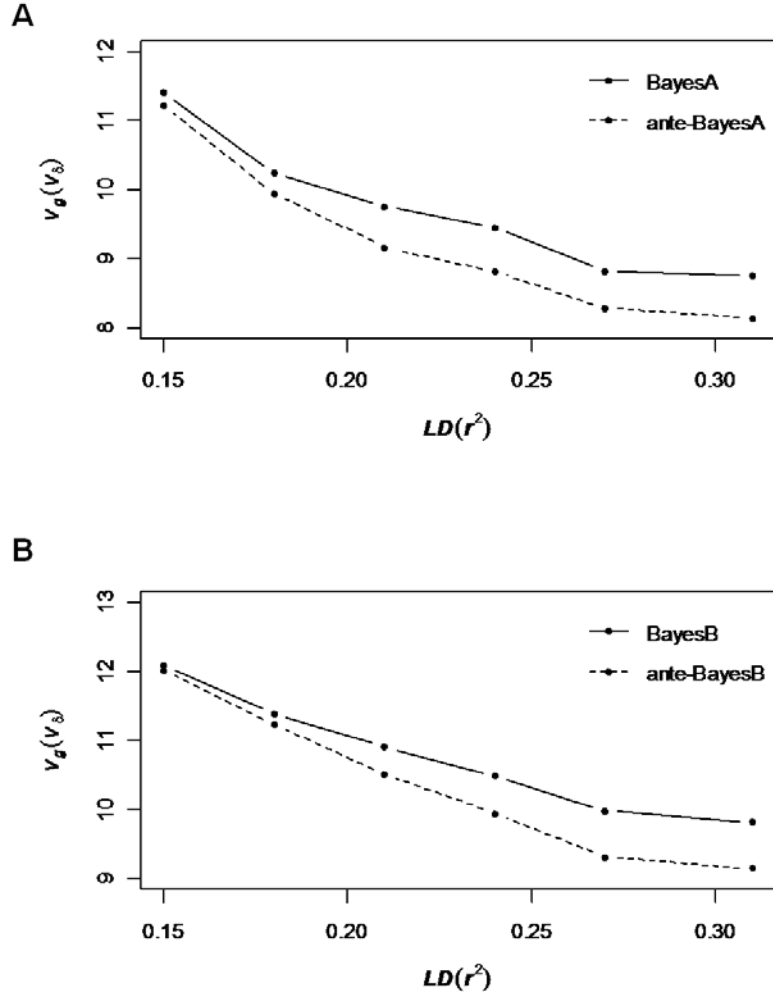


Figure 2.3: Average posterior means of v_g (BayesA, BayesB) and v_δ (ante-BayesA, ante-BayesB) across 20 replicates for six different levels of LD comparing BayesA and ante-BayesA in (A) and BayesB versus ante-BayesB in (B). No significant differences ($P > .01$) were determined between the two sets of competing procedures at each LD level.

Figure A2.1 and A2.2 (see Appendix A2) show, respectively, the average posterior means for μ_i and σ_i^2 against LD level across the 20 replicates under both ante-BayesA and ante-BayesB. There was no evidence ($P>0.01$) across these 20 replicates that the posterior means of μ_i were different from zero at any LD levels; however, at higher LD levels, the posterior means tended to converge to zero as anticipated. Similarly, Figure A2.2 showed that the posterior estimates for σ_i^2 were also lower at higher LD levels. Again, this was somewhat anticipated since there should be less disparity in different values of the antedependence parameters ($t_{j,j-1}$) between adjacent markers with increasing marker intensity.

The average accuracies of the EBV over the 20 replicated datasets are plotted as a function of the average r^2 (i.e., the different marker densities) between adjacent markers for the four different methods in Figure 2.4. As anticipated, given the simulated genetic architecture of few QTL, the accuracies for the BayesB methods were consistently greater than their corresponding BayesA counterparts at all marker densities. Also, ante-BayesA and ante-BayesB outperformed their classical counterparts with differences increasing with LD level. Specifically, ante-BayesA had significantly greater accuracies compared to conventional BayesA, as did ante-BayesB compared to BayesB ($P<0.01$), when average LD levels exceeded $r^2 = 0.24$.

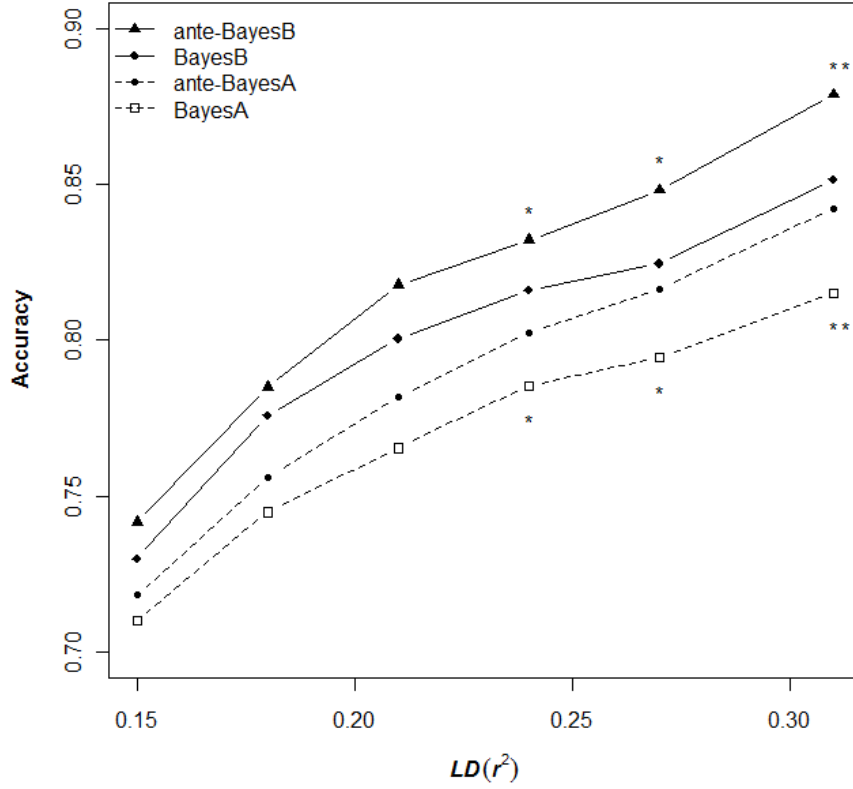


Figure 2.4: Average accuracies of estimated breeding value across 20 replicates for analyses based on each of six LD levels. Differences in accuracy between BayesA and ante-BayesA (bottom symbols) and between BayesB with ante-BayesB (top symbols) indicated as significant by $*$ ($P < .01$) ($*$) or $**$ ($P < .001$)

We anticipated that the antedependence parameters $t_{j,j-1}$'s would have greater importance at higher marker densities. To demonstrate this, we standardized the posterior means of these parameters as a ratio over their posterior standard deviations, i.e.,

$$\tilde{t}_{j,j-1} = \frac{E(t_{j,j-1} | \mathbf{y})}{\sqrt{\text{var}(t_{j,j-1} | \mathbf{y})}}, \text{ for each analysis. We then determined the proportion of these}$$

$\tilde{t}_{j,j-1}$ whose absolute value exceeded an arbitrary value of 2 for each data replicate and

marker density analysis to indicate the relative importance of these antedependence parameters. We present boxplots of these proportions across the 20 replicates for ante-BayesA and for ante-BayesB in Figure A2.3 (see Appendix A2). We anticipated and noted that a higher proportion of $\tilde{t}_{j,j-1}$ exceeded 2 in datasets characterized by higher marker densities, thereby indicating that, in general, nonstationary serial correlation between adjacent markers becomes increasingly more important with higher levels of LD. We believe this phenomenon is responsible for driving the differences in accuracies between ante-BayesA (ante-BayesB) and BayesA (BayesB) with increasing LD levels as seen earlier in Figure 2.4.

Hierarchical methods that are similar to BayesB, in that they jointly infer upon all SNP effects, have been increasingly advocated as tools for GWAS (Hoggart, Whittaker et al. 2008; Lee, van der Werf et al. 2008; Logsdon, Hoffman et al. 2010). Figure A2.4 (see Appendix A2) shows the average (across 20 replicates) posterior mean probabilities of identifying the largest QTL by genetic variance within each replicate using BayesB and ante-BayesB, respectively. These estimated posterior probabilities increased with LD level for both models but were significantly greater for ante-BayesB than for BayesB with statistical significance also increasing with LD or marker density. That is, the precision for detecting QTL was increasingly greater for ante-BayesB compared to BayesB at higher LD levels. We observed this consistently across data replicates with the ability of ante-BayesB to better track causal variants increasing with marker density (see Appendix A2, Figure A2.5).

2.3.2 Application to Heterogeneous Stock Mice data

We summarize posterior inferences of key parameters using BayesA and BayesB in Table A2.1 and for their antedependence counterparts in Table A2.2 (see Appendix A2) for the heterogeneous stock mice data. Inferences on σ_u^2 , σ_c^2 , and σ_e^2 were consistent with results previously reported by LEGARRA et al. (2008). As expected from our simulation study, the estimates for ν_g (ν_δ) and s_g^2 (s_δ^2) were substantially greater for BayesB (ante-BayesB) than for BayesA (ante-BayesA). Although the posterior mean for π_g of 0.81 (BayesB) was only slightly larger than $\pi_\delta = 0.80$ (ante-BayesB), the posterior mean of s_δ^2 was substantially larger in ante-BayesB compared to s_g^2 in BayesB. The average estimates \pm empirical standard errors of predictive ability correlations over the 20 cross validation partitions of training and validation data subsets were 0.57 ± 0.01 , 0.62 ± 0.01 , 0.60 ± 0.01 and 0.66 ± 0.01 for BayesA, BayesB, ante-BayesA and ante-BayesB, respectively. The differences between BayesA with ante-BayesA and BayesB with ante-BayesB were both determined to be statistically significant ($P < 0.005$), indicating the relative advantage of the antedependence methods. Furthermore, BayesB and ante-BayesB had significantly greater predictive abilities than BayesA and ante-BayesA, respectively ($P < .001$).

2.3.3 Application to Hickey and Gorjanc Data

Average posterior means for key hyperparameters for each of the four methods across the nine replicates are provided in Table A2.3, whereas the corresponding average ESS are provided in Table A2.4 (see Appendix A2). Estimates of π_g (π_δ) and s_g^2 (s_δ^2)

were lower whereas estimates of ν_g (ν_δ) are higher for traits with higher numbers of QTL (Traits 1 and 3) compared to those with lower numbers of QTL (Traits 2 and 4), relative to the same number of markers.

A side-by-side comparison of the accuracies of the four methods across the validation generations (6, 8, and 10) is provided in Figure 2.5. It is remarkable to note that ante-BayesA had generally significantly greater accuracies than BayesA for Traits 1 and 3 (larger numbers of QTL) that was still maintained until Generation 10, whereas ante-BayesB had generally significantly greater accuracies than BayesB for Traits 2 and 4 (lower numbers of QTL) but only in Generations 6 and 8. An assessment of bias of the four procedures based on regressing TBV on EBV is provided in Figure A2.6 (see Appendix A2). For all traits, all four methods had some significant bias in Generation 6 but not in later generations.

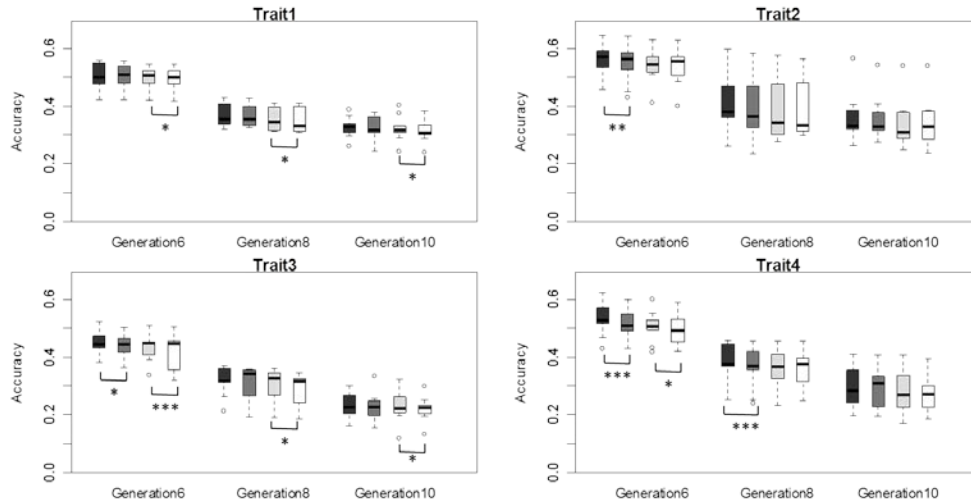


Figure 2.5: Boxplots of average accuracies of estimated breeding value across 9 replicates for four traits in Generations 6, 8 and 10 for benchmark data from Hickey and Gorjanc (2012). Differences in accuracy between anteBayesB (black) and BayesB (dark gray) and between anteBayesA (light gray) with BayesA (white) indicated as significant by * (0.05 < P < 0.10), ** (0.01 < P < 0.05) or *** (P < 0.01).

2.4 Discussion and Conclusion

In this paper, we extend two very popular Bayesian methods, BayesA and BayesB, for WGP to model potential nonstationary correlations between SNP effects in close proximity to QTL. We demonstrated using a small-scale simulation study that the accuracies of our proposed first order antedependence extensions, labeled ante-BayesA and ante-BayesB, were greater than their classical counterparts with differences increasing with marker density. This result was anticipated given that the magnitude and importance of the antedependence parameters $\{t_{j,j-1}\}_{j=2}^m$ in \mathbf{T} should increase as marker densities increase. To further illustrate the importance of modeling nonstationary correlations, rather than basing correlations between SNP effects purely as a function of distance (Gianola, Perez-Enciso et al. 2003), we observed the magnitude of the posterior means of $\{t_{j,j-1}\}_{j=2}^m$ at the various locations through the chromosome for the first four replicates using ante-BayesA and ante-BayesB (see Figure A2.7). These posterior means tended to be rather large in absolute value in the general vicinity of the major QTL. This result was anticipated since each QTL is likely tracked by several SNP, each in incomplete LD with the QTL (GODDARD and HAYES 2009). Interestingly, there appeared to be a greater spread in these posterior means around a greater number of QTL using ante-BayesB compared to ante-BayesA.

We realize that the order in which Equation [3] is specified for the antedependence methods is rather arbitrary; i.e., one might specify Equation [3] from the end of the p-arm to the end of the q-arm of a chromosome or vice-versa. For instance, instead of specifying Equation [3] from $j = 1, 2, \dots, m$, we might have also modeled

antedependence in the opposite direction; i.e., from $j = m, m-1, \dots, 1$. It has been demonstrated by ZIMMERMAN and NÚÑEZ-ANTÓN (2010) in the context of longitudinal data analysis that the (co)variances based on a first order antedependence model are invariant to directionality as long as the relative order is correctly specified. To illustrate this, we re-analyzed the first four replicates based on the highest average marker density ($r^2 = 0.31$), again using ante-BayesA and ante-BayesB but this time specifying Equation [3] in the opposite direction from what was used previously. We plotted the posterior means ($\bar{\mathbf{g}}$) of \mathbf{g} based on the analysis in the original direction against the same estimates based on the analysis in the opposite direction (Figure A2.8), further demonstrating that inferences on EBV are invariant to direction. Noting that the EBV are linear functions of $\bar{\mathbf{g}}$, i.e., $\mathbf{Z}\bar{\mathbf{g}}$, we noted even greater consistency for EBV between the two directions for these same four replicates in Figure A2.9 (see Appendix A2).

Given that the accuracy of EBV was greater using the antedependence-based procedures compared to their classical counterparts, we examined the elements of $\bar{\mathbf{g}}$ for each marker between the two different classes of models within each of the first four replicates and the highest average marker density (Figure A2.10). It was interesting to note that these elements were more shrunk to zero using the antedependence-based procedures compared to the conventional counterparts. Given the specification of nonstationary correlations between effects of adjacent SNP markers, the effective number of SNP may be considered to be somewhat lower using the antedependence based procedures, particularly in regions containing a major QTL. That is, larger elements of $\bar{\mathbf{g}}$ using the antedependence-based models will tend to be more highly correlated and

hence shrunk closer to zero compared to elements of $\bar{\mathbf{g}}$ derived from their conventional counterparts. GODDARD et al. (2009) have previously described the optimal statistical properties of $\bar{\mathbf{g}}$ when \mathbf{g} is specified as a set of “random effects” having an exchangeable distribution, such as the normal or Student t . However, these properties are better realized when a more appropriate correlation structure is specified; as would be true, for example, modeling polygenic effects with a classical animal model (HENDERSON 1984). The more optimal properties of $\bar{\mathbf{g}}$ using the antedependence based models was also partly reflected in the earlier EBV comparisons with their classical counterparts.

In spite of the limited scale of our simulation studies, it has been demonstrated that inferences on accuracy based on 100 individuals and a genome length of 1 M is roughly equivalent to inferences derived from 3000 individuals and a genome length of 30 M (MEUWISSEN and GODDARD 2010), the latter of which might depict a more common scenario in livestock populations. We also based all of our simulation work on a heritability of 50%; for situations with lower or higher heritabilities, we would naturally expect the accuracies to respectively, decrease or increase accordingly in concert with previous simulation results (CALUS and VEERKAMP 2007) and/or analytical derivations (MEUWISSEN and GODDARD 2010); however, we believe that there is no reason to believe that the antedependence methods would not outperform the conventional Bayesian WGP methods in these situations as well; this was further substantiated by our analyses of the data from Hickey and Gorjanc (2012) which was based on a heritability of 25%. In our own simulation study, there were around 2200 markers per the single 1M chromosome using the highest average LD level of $r^2 = 0.31$, whereas there were around 100 markers per the single chromosome with the lowest average LD level of $r^2 = 0.15$. Using

Meuwissen and Goddard (2010), these two specifications are, respectively, analogous to a panel of 60,000 SNP markers and to a panel of 3,000 SNP markers for a 30 M genome; commercially developed panels having roughly these same numbers of SNP markers are now widely available for cattle (Wiggans, VanRaden et al. 2011). Based on the results of our work, we anticipate that the antedependence-based methods, compared to their classical counterparts, would lead to even greater accuracies with higher density SNP marker panels ($m > 500,000$) that are being developed for livestock or for situations where there is sequence data (MEUWISSEN and GODDARD 2010). Along those lines, we anticipate that these methods would also perform better in populations where LD is greater between markers due to other phenomena; e.g., selection history.

Our simulation studies were also based on a particular genetic architecture; i.e. 30 QTL that were randomly distributed throughout a 1 M chromosome (or equivalently, 900 QTL for a 30 M genome). Although this is not the focus of our paper, we realize that genetic architecture (i.e., number of QTL, average QTL substitution effect, marker density, etc.) can impact the relative merit of BayesA, BayesB, and GBLUP based on other studies where key hyperparameters such as π_g , ν_g and s_g^2 are arbitrarily specified to be known (Daetwyler, Pong-Wong et al. 2010; Meuwissen and Goddard 2010). That is, the greater the number of QTL, each with small effects, relative to the number of SNP markers, the more likely the genetic architecture reflects the GBLUP assumptions ($\pi_g = 1$, $\nu_g \rightarrow \infty$ such that $\sigma_{g_j}^2 = s_g^2 \forall j$). Conversely, BayesB would be favored in the situation where SNP marker density is high relative to the number of QTL ($\pi_g < 1$). However, we believe that formal comparisons in data fit between BayesA, BayesB, and GBLUP, along

with ante-BayesA and ante-BayesB, are not entirely necessary since ante-BayesB represents the most general model. As previously noted, ante-BayesA is a special case of BayesA, as is ante-BayesB of BayesB, when $\mathbf{T} = \mathbf{0}$ such that then $\pi_{\delta} = \pi_g$, $\nu_{\delta} = \nu_g$, and $s_{\delta}^2 = s_g^2$. Furthermore, BayesB becomes BayesA as $\pi_g \rightarrow 1$, whereas BayesA becomes GBLUP as $\nu_g \rightarrow \infty$. Nevertheless, our claim that one only needs to fit ante-BayesB, rather than any of the other three competing submodels, vitally depends upon reliable inferences being provided on these key hyperparameters defining genetic architecture, rather than arbitrarily specifying them (Daetwyler, Pong-Wong et al. 2010; Meuwissen and Goddard 2010) or estimating a subset thereof (Habier, Fernando et al. 2011). We provide details on MCMC inference strategies on these and other unknown parameters in Appendix A1. We are currently pursuing more suitable inferential strategies for variable selection (O'HARA and SILLANPAA 2009) when inferring upon π_g or π_{δ} . Also, although our proposed antedependence methods seem to work well under additive genetic model assumptions, it is not clear how well they may perform in the presence, for example, of extensive non-additive gene action where nonparametric approaches may be warranted (Gianola, Wu et al. 2010). Nevertheless, even in the extensive presence of such phenomena, genetic variance is still considered to be primarily additive (Hill, Goddard et al. 2008).

Although the scope of this work was focused on the potential merit of these antedependence models for WGP, we suggested earlier that there may be also merit for using these models in GWAS in both livestock and human populations. It has become increasingly recognized that GWAS procedures based on joint analyses of all SNP

markers are more powerful than the conventional series of single SNP analyses. Our results suggest that modeling nonstationary correlations between SNP effects will further augment this power. At any rate, we recognize that for reasonably accurate GWAS, that a greater marker density (m) per chromosome and sample size (n) should be considered (e.g., MEUWISSEN and GODDARD 2010) than those studied in this paper; i.e., most of the posterior probabilities reported in Appendix Figure A2.4 and Figure A2.5 are too low to be of practical benefit in current applications.

We also acknowledge that our ante-BayesA and ante-BayesB models increase the computational load relative to their conventional counterparts. Since m is typically large, the computing time for the proposed antedependence models is bottlenecked primarily by the m elements of δ , the m diagonal elements of Δ and the $m-1$ non-zero elements of \mathbf{T} . Similarly, computing time for the two conventional methods, BayesA and BayesB, is primarily restricted by the dimension of δ and Δ ; i.e., roughly 2/3 as many variables for the antedependence-based models, ignoring the remaining parameters such as variance components and hyperparameters. Hence, the computing time for the antedependence based procedures should be somewhat less than 1/3 greater than for their conventional counterparts. Indeed, we discovered from our simulation study that computing time for all four competing models were linear in m with the antedependence based models taking less than 30% greater computing time compared to the conventional counterparts for the wide range of values of m considered in this paper. We recognize for much larger number of SNP markers, than those pursued in this study, that alternative algorithmic adaptations already developed for models similar to conventional BayesA or BayesB,

such as those based on the EM algorithm (Shepherd, Meuwissen et al. 2010) or variational Bayes (Logsdon, Hoffman et al. 2010), would be worth exploring.

We believe the proposed antedependence models provide opportunities for further study and extension. For example, it has been previously recognized that basing inferences on allelic effects on the use of multiple marker haplotypes rather than single markers increases accuracy of WGP (Calus, Meuwissen et al. 2008; Villumsen, Janss et al. 2008) or GWAS (Grapes, Dekkers et al. 2004). Given the difficulty in how to appropriately specify these haplotypes, we believe our antedependence-based methods may help bridge these two different strategies as the effects of adjacent SNP markers connected by large values of $t_{j,j-1}$ may somewhat determine “effective haplotype” effects. We also think that our antedependence specifications might facilitate multiple breed inference if, for example, genomic effect differences between breeds is primarily due to differences in SNP associations with QTL, as partly manifested in **T**

Chapter 3 Improving the computational efficiency of fully Bayes inference and assessing the effect of misspecification on hyperparameters in whole genome prediction model

3.1 Introduction

Genomic predictions based on high density single nucleotide polymorphisms (SNP) markers distributed over the whole genome have become increasingly adopted for animal and plant breeding. Parametric Bayesian methods have been particularly popular, most notably BayesA and BayesB as first presented by MEUWISSEN et al. (2001). BayesB specifies a mixture prior on the SNP specific effects having point mass at zero with probability π or randomly drawn, with probability $(1 - \pi)$, from a Student t distribution with degrees of freedom ν and scale parameter s^2 ; BayesA is BayesB with $\pi = 0$. Hence π is typically believed to be the proportion of SNPs that are not associated or in linkage disequilibrium (LD) with causal variants although this interpretation is somewhat complicated by the existence of LD. These hyperparameters (ν , s^2 and π) are relevant in that they partly determine the genetic architecture of traits and can be further shown to depend upon SNP marker densities used in the analyses (YANG and TEMPELMAN 2012).

Now inference in BayesA/B like models is conducted using either Markov Chain Monte Carlo (MCMC) methods for fully Bayesian inference or faster albeit approximate methods based on the use of the expectation maximization (EM) algorithm or its various derivatives (Shepherd, Meuwissen et al. 2010). Unfortunately, it has not been readily established how to properly infer upon these hyperparameters in the EM based methods such that they are often arbitrarily “tuned” or specified (KARKKAINEN and SILLANPAA 2012). Furthermore, although it is possible to infer upon these same hyperparameters

using MCMC, the poor efficiency and speed of these implementations have seemingly discouraged this practice (de los Campos, Hickey et al. 2013). In particular, it has been noted that the correlation between v and s^2 across MCMC cycles is generally so large that these two hyperparameters are nearly non-identifiable from each other (Habier, Fernando et al. 2011; de los Campos, Hickey et al. 2013). This particular MCMC analysis was based on a strategy first presented by Yi and Xu (2008) that invokes a Gibbs update for the full conditional density (FCD) on s^2 , as it is conditionally conjugate with a Gamma prior, whereas a Metropolis Hastings (MH) update was used on sampling from the FCD of v since it is not recognizable (YI and XU 2008). We label this particular algorithm as DFMH (i.e., sampling v using MH) and it is the control or reference strategy for this paper.

Now computational efficiency in MCMC schemes is related to the degree of mixing or autocorrelation between subsequent samples of the same parameter. The most popular metric for inferring the degree of mixing or autocorrelation for a fixed number of MCMC cycles is the effective sample size (ESS), which can be readily computed using software packages like CODA (Plummer, Best et al. 2006). The ESS determines the effective number of independent draws such that a greater degree of autocorrelation between subsequent samples for the same parameter would lead to a smaller ESS and hence poorer computational efficiency. Now, although there are clear exceptions, MCMC sampling strategies that lead to a greater ESS for a certain total number of MCMC cycles tend to have greater computational cost per cycle. This realization is reflected in other recent quantitative genetics applications (Shariati and Sorensen 2008;

Waagepetersen, Ibanez-Escriche et al. 2008) who derived various metrics to integrate together these two components of computational efficiency.

We surmised that there may be a number of strategies that could improve the computational efficiency of inferring upon key hyperparameters in a BayesA/B WGP model compared to DFMH. Furthermore, the efficiency of any such strategy could markedly depend on the use of an appropriate scale. For example, a highly nonlinear relationship between two variables can be rendered somewhat linear by transforming either one or both of the corresponding parameters. When ν and s^2 are both log-transformed, the resulting scatterplot of the transformed variables against each other tends to demonstrate a more linear relationship. Hence this change of variable might facilitate potentially more efficient MCMC sampling strategies based on multivariate proposal densities, for example.

Specification of key hyperparameters in a BayesA/B WGP model has been treated arbitrarily in a wide selection of genomic selection topics. Meuwissen et al. (2001) chose 4.2 as ν and 0.04 as s^2 for BayesB model based on population genetics arguments in a simulation study. Daetwyler et al. (2010) set both ν and s^2 to 1 under their BayesB model analyses across all simulation scenarios. In various simulation scenarios, it seems reasonable to choose different specifications on the key hyperparameters as their estimates were dependent on many factors such as marker density in the simulation study (YANG and TEMPELMAN 2012).

There were two primary objectives in this study. First, we wanted to explore alternative strategies to improve the computational efficiency of estimating

hyperparameters in BayesA/B WGP models. Secondly given the prevalent practice of specifying rather than estimating these hyperparameters, we wanted to assess the impact of misspecifying these hyperparameters on accuracy of breeding value prediction.

3.2 Materials and Methods

3.2.1 WGP Model

The WGP model used for comparison of the various computational strategies and/or hyperparameter specifications can be denoted as follows:

$$y_i = \mathbf{x}_i' \boldsymbol{\beta} + \sum_{j=1}^m z_{ij} g_j + e_i \quad [1]$$

Here y_i is the phenotype for i th animal ($i=1,2,\dots,n$), $\boldsymbol{\beta}$ is a vector of fixed effects such that \mathbf{x}_i' is the known incidence row vector connecting y_i to $\boldsymbol{\beta}$, z_{ij} is the genotype covariate for SNP j on animal i coded as either 0, 1, or 2 copies of a reference allele for SNP j on animal i , g_j is the random effect for SNP j , and e_i is the residual. The WGP model in matrix algebra notation can be written as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{g} + \mathbf{e} \quad [2]$$

Where $\mathbf{X} = \{\mathbf{x}_i'\}_{i=1}^n$, $\mathbf{Z} = \{z_{ij}\}$, $\mathbf{g} = \{g_j\}_{j=1}^m \sim N(\mathbf{0}, \boldsymbol{\Sigma}_g)$ with variance-covariance matrix

$\mathbf{G} = \text{diag}(\sigma_{g_j}^2)$ and residual vector $\mathbf{e} = \{e_i\}_{i=1}^n \sim N(\mathbf{0}, \mathbf{I}\sigma_e^2)$.

We compared three sampling strategies under BayesA and BayesB specifications (Meuwissen, Hayes et al. 2001) on $\sigma_{g_j}^2$ in WGP model. In BayesB, $\sigma_{g_j}^2$ has a mixture

prior of two components: a scaled inverted chi-square distribution with $\sigma_{g_j}^2 \sim \chi^{-2}(\nu, \nu s^2)$ with probability π and a spike at 0 with probability $(1 - \pi)$. Here π loosely represents the proportion of SNP markers having associated genetic effects on the phenotype. BayesA is a special case of BayesB when $\pi = 1$. Following Yang and Tempelman (2012), we specify the following prior distributions on the hyperparameters as: $\nu \sim p(\nu) \propto (\nu + 1)^{-2}$ and the Gelman prior $s^2 \sim \chi^{-2}(-1, 0)$ (GELMAN 2006) for BayesA, a proper conjugate prior $s^2 \sim \text{Gamma}(0.1, 0.1)$ and $\pi \sim p(\pi | \alpha_\pi, \beta_\pi) = \text{Beta}(\alpha_\pi = 1, \beta_\pi = 8)$ for BayesB. For all three computational strategies that we subsequently describe, we adapt the same commonly used MCMC strategies for sampling from all parameters/random variables, other than ν , s^2 , and π , as outlined by Meuwissen et al. (2001) for example. We now describe each of the three computational strategies in turn.

3.2.2 Univariate Metropolis Hastings sampling on ν and Gibbs update on s^2

(DFMH)

This strategy, which we designate as DFMH, closely follows Yi and Xu (2008). The FCD of ν does not have a recognizable form; hence sampling from this FCD requires a strategy other than a Gibbs step. Here, we used the MH algorithm to sample from the FCD of ν drawing from our experiences in various other applications (Kizilkaya, Carnier et al. 2003; Kizilkaya and Tempelman 2005; Bello, Steibel et al. 2010; Yang and Tempelman 2012). More specifically, we generate from the FCD of $\xi = \log(\nu)$, ensuring that the FCD of ξ takes into account the Jacobian of the transformation from ν to ξ (see Appendix B1.1). Since ξ can conceptually be defined

anywhere on the continuous real line, we believe this transformation better justifies the use of a Gaussian proposal density centered on the value of ξ from the previous MCMC cycle; i.e., a random walk MH step (CHIB and GREENBERG 1995); alternatively, a heavier-tailed Student t proposal density (CHIB and GREENBERG 1995) could be used as well. During the first half of burn-in, we adaptively tune the variance of this proposal density such that the MH acceptance ratios are intermediate (i.e., 25-75%) adapting the strategy described by Muller (1991) and in accordance with standard recommendations (Gelman, Carlin et al. 2003; Carlin and Louis 2008). This proposal density variance was then fixed for the last half of burn-in in order to ensure a proper convergent MCMC algorithm. Yi and Xu (2008) demonstrated that the FCD of s^2 is Gamma, provided that a conditionally conjugate Gamma or noninformative prior is used. Using the Gelman prior (GELMAN 2006) for s^2 as we have previously advocated for BayesA (YANG and TEMPELMAN 2012), the FCD of s^2 can be shown to be Gamma with shape $0.5(mv + 1)$ and scale $0.5\nu \sum_{j=1}^m \sigma_{sj}^2$. Hence, for DFMH, we sampled ν using the described MH update and s^2 with a Gibbs update. In DFMH, we sampled 10 MH samples per MCMC cycle for ν .

3.2.3 Univariate Metropolis Hastings sampling for each of ν and s^2 (UNIMH)

Metropolis Hastings sampling, if properly tuned with good proposal densities and intermediate acceptance rates, can often lead to faster mixing and hence greater MCMC efficiency relative to Gibbs sampling. This is because MH sampling typically proposes bigger jumps throughout the posterior density compared to the use of Gibbs sampling.

Hence, we propose a second strategy, UNIMH, whereby we again use MH to sample from ν but also use MH to sample from s^2 . As with ν in DFMH, we sample s^2 by first a change of variable to its logarithm (i.e., $\psi = \log(s^2)$) and use a random walk MH algorithm based on a Gaussian proposal density for ψ . Similar to what was done for ν , the variance of this proposal density was only tuned for intermediate acceptance rates during the first half of burn-in to ensure a properly convergent MCMC chain. In UNIMH, 10 MH samples per MCMC cycle were specified for sampling ν and s^2 . Details on this strategy are further provided in Appendix B1.2.

3.2.4 Bivariate Metropolis Hastings sampling on ν and s^2 (BIVMH)

As previously noted, the posterior correlation between ν and s^2 can be high; hence, it might be advantageous to jointly sample both parameters together with a bivariate random walk MH sampler as demonstrated with another application by Ntzoufras (2011). Hence, we propose a third sampling algorithm that we label BIVMH. Here, we divided the burn-in period for this strategy into four stages of equal lengths with respect to the number of MCMC cycles; arguably, a more efficient implementation might be possible given that these stages may not necessarily need to be of the same length. In Stage 1, we sampled $\log(\nu)$ and $\log(s^2)$ from their respective FCD using the UNIMH strategy previously described, fine-tuning the variances of the two separate Gaussian proposal densities to ensure MH acceptance rates falling between 25% and 75%. In Stage 2, we sampled $\log(\nu)$ and $\log(s^2)$ using UNIMH, fixing the variances of their respective proposal densities to those values tuned at the end of Stage 1 while computing

the empirical correlation between the samples of $\log(\nu)$ and $\log(s^2)$ drawn within the same cycle. In Stage 3, $\log(\nu)$ and $\log(s^2)$ were jointly sampled together using a bivariate Gaussian proposal density with variances based on those tuned at the end of Stage 1 and a covariance based on the correlation computed from Stage 2. During Stage 3, we further fine-tuned the proposal variances to ensure intermediate acceptance rates for joint samples of $\log(\nu)$ and $\log(s^2)$ with the proposal covariance based on the same correlation derived in Stage 3. In Stage 4, we drew samples using the same joint MH random walk from the newly tuned bivariate Gaussian proposal density in Stage 3 but without further tuning in order to ensure a proper convergent MCMC chain. Upon the end of Stage 4, and hence burn-in, we saved samples for the hyperparameters of ν and s^2 (i.e., back-transformed) for MCMC-based fully Bayesian inference. Ten MH samples per MCMC cycle for ν and s^2 were drawn at each Stage. Details on this strategy are further provided in Appendix B1.3.

3.2.5 Simulation Study

In order to compare the efficiency of the three sampling strategies, DFMH, UNIMH and BIVMH under BayesA and BayesB modeling specifications, we simulated 15 replicated datasets using the HaploSim package in R (Coster, Bastiaansen et al. 2010). The simulated genome was composed of one chromosome of length 1 Morgan consisting of 100,000 equally spaced loci. For each of the 100 animals in the base population, every 5th locus on this chromosome was heterozygous (i.e., for a total of 20,000 such loci) whereas the remaining 80,000 loci were completely monomorphic, similar to that in Coster et al. (2010). Individuals were randomly mated to generate 100 animals within

each of 6000 subsequent generations in order to generate LD between loci. The number of recombinations per each meiosis event was drawn from a Poisson(1) distribution with the position of each recombination being randomly drawn from a uniform distribution on the chromosome (i.e., no interference). Furthermore, we specified the recurrent mutation rate to be 10^{-5} per locus per generation.

After Generation 6000, random matings were used to augment the population size to 1000 individuals in Generation 6001. In Generation 6001, we deleted loci with a minor allele frequency (MAF) less than 0.05 and randomly selected 30 from the remaining loci to be quantitative trait loci (QTL). Following Meuwissen et al. (2001), we simulated substitution effects α for these 30 QTL from a reflected gamma distribution with shape parameter 0.4 and scale parameter 1.66 such that the true breeding values (TBV) were genotype-based linear combinations of α . Phenotypes for animals in generation 6001 were generated based on heritability of 50%; i.e., such that $\sigma_e^2 = \text{var}(\text{TBV})$. Additionally, genotypes for 1000 offspring in generation 6002 were based on random matings of individuals in Generation 6001. Again, TBV were based on linear combinations of α based on QTL genotypes inherited from Generation 6001.

After discarding SNP with $\text{MAF} < 0.05$, we then selected every 1st, 4th and 10th SNP markers for inclusion in analyses in order to consider the effect of different marker densities; i.e., high (around 2394 SNPs), medium (around 598 SNPs), and low (around 239 SNPs). We compared the computational efficiency of inferring on key hyperparameters (e.g., v and s^2) for genetic architecture under these three different marker densities. We ignored fitting polygenic effects for all comparisons in the

simulation study to facilitate further computational feasibility, recognizing that the relative efficiency of each strategy should not differ otherwise.

We compared the computational efficiencies of the three MCMC strategies on each replicated dataset, considering each of three different marker densities. Now computational efficiency, as it pertains to a particular hyperparameter, was defined as the effective sample size (ESS) for the post-burn-in MCMC cycles divided by total CPU time; i.e. ESS/CPU recorded in #/seconds. That is, the greater ESS/CPU, the greater the computational efficiency for inferring the posterior density of that particular hyperparameter.

Given that many researchers do not infer upon some or even all hyperparameters in WGP models because of perceived inferential challenges, we thought it important to assess the impact of their misspecification on the accuracy of genomic prediction. Using the same simulated data as described previously, we focused on five different scenarios, all at the medium marker density (selecting every 4th marker). Each scenario was based on setting s^2 to be an arbitrary multiplicative constant of the average posterior mean at the medium marker density (\bar{s}_{med}^2) based on the complete (BayesA or BayesB) model that was used to infer upon the other hyperparameters (ν and π where applicable) as well. These five scenarios were to set 1) $s^2 = \bar{s}_{med}^2$, 2) $s^2 = 0.1 \bar{s}_{med}^2$, 3) $s^2 = 0.01 \bar{s}_{med}^2$, 4) $s^2 = 10 \bar{s}_{med}^2$, and 5) $s^2 = 100 \bar{s}_{med}^2$. Note that the specification of \bar{s}_{med}^2 depended upon which model (BayesA or BayesB) was employed, as described later.

We also wondered if one could roughly specify a good working value for s^2 by merely basing it on an estimate derived from, say, analysis of the same phenotype but

based on a SNP panel with a different marker density. As s^2 represents a typical value for the SNP-specific variances $\sigma_{g_j}^2$, then its value should be inversely related to the number of SNP markers as we have observed previously (YANG and TEMPELMAN 2012). For example, given that there were four times as many markers at the high marker density as there were at the medium marker density in the simulation study, an initial specification for s^2 at the high marker density is to use $s^2 = 0.25 \bar{s}_{med}^2$. Similarly, since there were half as many SNP markers for the low marker density specification, an initial specification in that case might be $s^2 = 2 \bar{s}_{med}^2$. These specifications for s^2 were compared for their effect on accuracy of breeding value prediction relative to the situation where s^2 is inferred upon along with all other hyperparameters under both BayesA and BayesB for all 15 replicated datasets.

In all cases, accuracy was defined as correlation between estimated breeding values (EBV) and TBV where $EBV = \mathbf{Z} \bar{\mathbf{g}}$ for $\bar{\mathbf{g}}$ is the posterior mean of \mathbf{g} and TBV is defined as before.

3.2.6 Data Application: Assessment of computational efficiency comparisons

In this dataset, 2,296 mice were genotyped for 12,147 SNP markers with a high pairwise LD of $r^2=0.6$ (Legarra, Robert-Granie et al. 2008). After data cleaning on genotypes (YANG and TEMPELMAN 2012), there were 1940 animals with 10,467 SNP markers. We selected 50, 100 and 200 SNP markers from each of the 19 autosomes to create three different marker densities using pre-corrected body weight at 6 weeks as our phenotypes. As in Yang and Tempelman (2012), we also modeled the random effects of cage in addition to SNP effects and polygenic effects in the WGP model using the

Gelman prior (GELMAN 2006) specified on the cage σ_c^2 and the polygenic variance σ_u^2 .

After merging phenotypes with the genotypes, we were left with 1917 animals with complete phenotypes and genotypes on 950, 1900 and 3800 SNP markers across the 19 autosomes.

3.3 Results

3.3.1 Simulation Study

By selecting every single, 4th and 10th SNP markers for inclusion, the average r^2 between adjacent SNPs, was 0.17, 0.24 and 0.32 for the three marker densities over the 15 replicated datasets. Inferences on s^2 in the three sampling strategies DFMH, UNIMH and BIVMH were compared under both BayesA (Appendix Figure B2.1A) and BayesB (Appendix Figure B2.1B) specifications. We observed estimates (posterior means) of s^2 decreased as the marker density increased and that estimates derived from BayesB were generally one order of magnitude greater than those in BayesA. Furthermore, estimates of π generally increased (Appendix Figure B2.2) whereas estimates of v generally decreased as marker density increased (Appendix Figure B2.3). All of these results are consistent with our previous work (YANG and TEMPELMAN 2012).

For quality control, we checked to see that the estimates for the key hyperparameters should be the same between the three computational strategies within each replicated dataset under the same model, allowing for Monte Carlo error. Pairwise scatterplots of the estimates of s^2 under the three different strategies for each of the three different marker densities in BayesA (Appendix Figure B2.4) and in BayesB (Appendix Figure B2.5) indicated generally good agreement as did for π using BayesB (Appendix

Figure B2.6) and for ν under BayesA (Appendix Figure B2.7) and BayesB (Appendix Figure B2.8). The greatest differences between DFMH and UNIMH or between DFMH and BIVMH were found in estimating ν under BayesB (Appendix Figure B2.8).

Figure 3.1 and 3.2 illustrate side-by-side boxplots of ESS/CPU for each of the three strategies under each of the three marker densities for ν and s^2 , respectively, under the BayesA model. In all cases, ESS/CPU were higher for BIVMH and UNIMH compared to DFMH ($P < 0.05$). For ν at high marker density ($r^2 = 0.32$), BIVMH had higher ESS/CPU than UNIMH ($P < 0.0001$). For s^2 at high and median marker densities ($r^2 = 0.32$ and 0.24), ESS/CPU were higher for BIVMH than UNIMH ($P < 0.0001$). Interestingly, the differences in efficiencies between the three strategies widened as marker density increased. Efficiencies for the three alternative sampling strategies were also compared under BayesB model for ν (Figure 3.3), s^2 (Figure 3.4), and π (Figure 3.5). We found that UNIMH and BIVMH had significantly greater computational efficiencies compared to DFMH for all three hyperparameters ($P < 0.05$). For ν , UNIMH had significantly higher ESS/CPU compared to BIVMH at low marker density ($P < 0.05$). For s^2 , UNIMH had significantly greater computational efficiencies compared to BIVMH at median marker density whereas BIVMH had higher ESS/CPU than UNIMH at low marker density ($P < 0.05$). For π , BIVMH had higher ESS/CPU compared to UNIMH at high and median marker densities ($P < 0.05$).

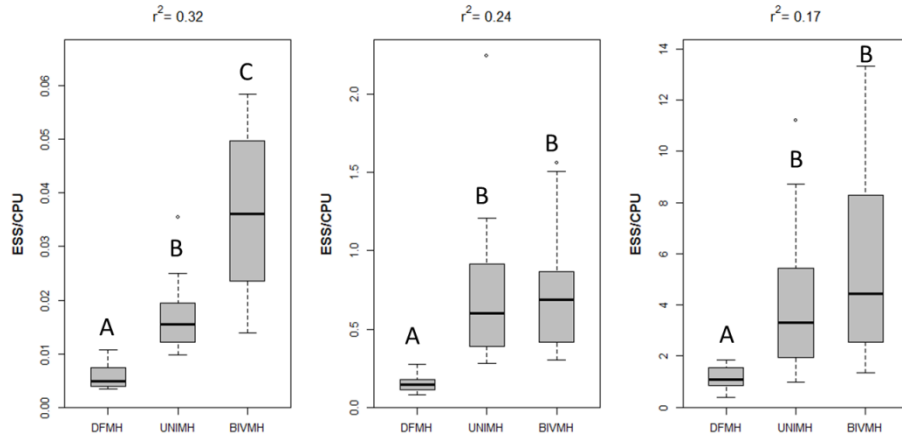


Figure 3.1: Boxplots of effective sample size for ν divided by total CPU time in seconds across 15 replicates for three different levels of LD comparing DFMH, UNIMH and BIVMH under BayesA model. Different letters indicate significant difference with $P < 0.05$.

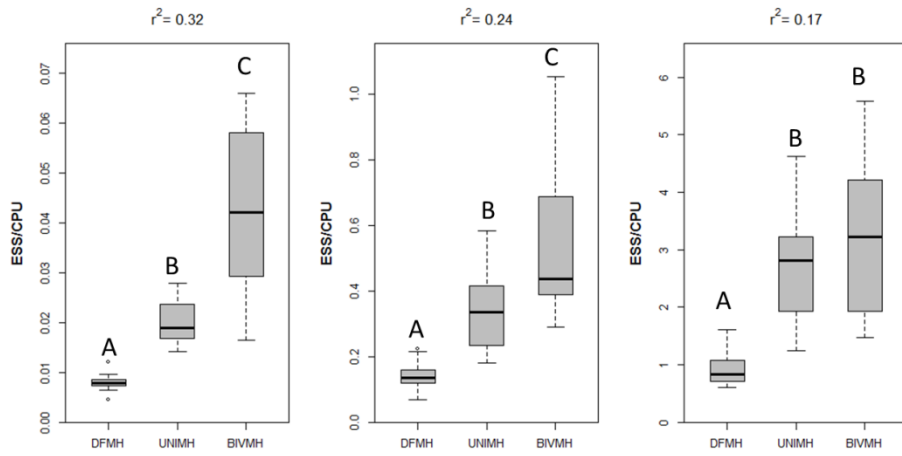


Figure 3.2: Boxplots of effective sample size for s^2 divided by total computational time in seconds across 15 replicates for three different levels of LD comparing DFMH, UNIMH and BIVMH under BayesA model. Different letters indicate significant difference with $P < 0.05$.

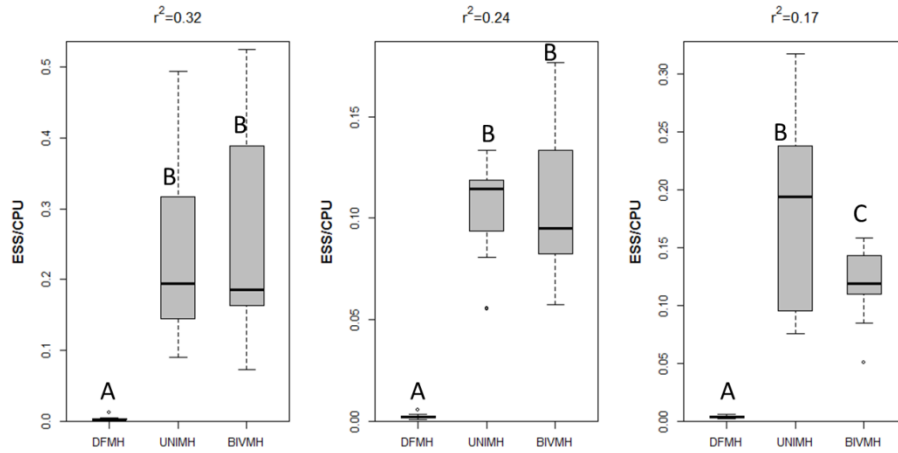


Figure 3.3: Boxplots of effective sample size for ν divided by total computational time in seconds across 15 replicates for three different levels of LD comparing DFMH, UNIMH and BIVMH under BayesB model. Different letters indicate significant difference with $P < 0.05$.

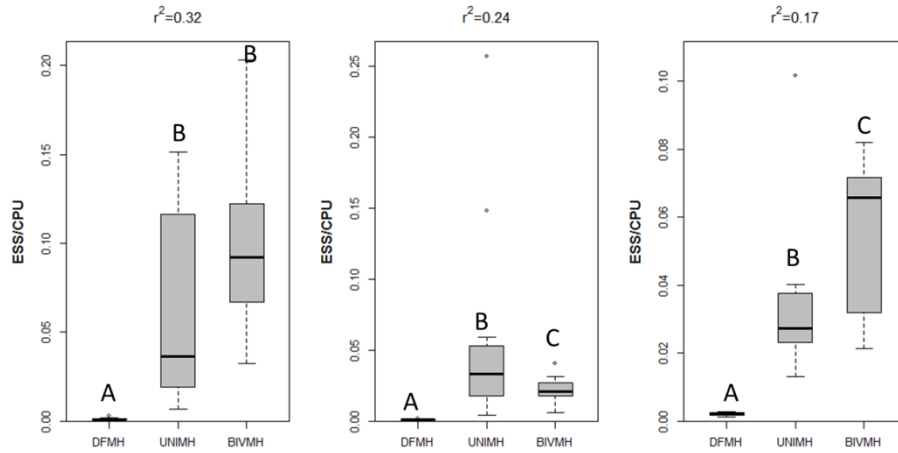


Figure 3.4: Boxplots of effective sample size for s^2 divided by total computational time in seconds across 15 replicates for three different levels of LD comparing DFMH, UNIMH and BIVMH under BayesB model. Different letters indicate significant difference with $P < 0.05$.

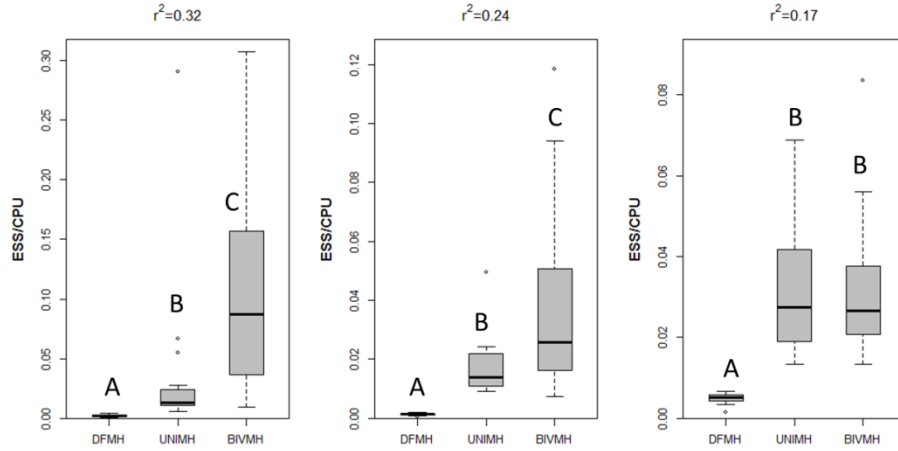


Figure 3.5: Boxplots of effective sample size for π divided by total computational time in seconds across 15 replicates for three different levels of LD comparing DFMH, UNIMH and BIVMH under BayesB model. Different letters indicate significant difference with $P < 0.05$.

We also separately looked at the components of computational efficiency; i.e., ESS and CPU/cycle in seconds for each parameter in both models and under all three strategies. As anticipated, DFMH was computationally less expensive in terms of CPU/cycle compared to the proposed strategies UNIMH and BIVMH; however, the ESS for the 40,000 MCMC cycles that were drawn in each analyses were such that UNIMH and BIVMH generally far exceeded that of DFMH. What was particularly ominous was how quickly the ESS measures degraded with increasing marker densities thereby suggesting that high density marker panels lead to analyses that require not only greater CPU/cycle but also a greater number of MCMC cycles to ensure that ESS values are sufficiently great enough to ensure reliable inference.

We were interested as to whether accuracy of breeding value prediction might depend on misspecification of hyperparameters, say, s^2 . We assessed the impact on

accuracy of breeding value predictions based on setting s^2 to a wide range of values based on various multiples (0.01x to 100x) of the average posterior mean ($\bar{s}_{med}^2 = 7 \times 10^{-4}$ for BayesA, $\bar{s}_{med}^2 = 4 \times 10^{-2}$ for BayesB) across the 15 replicates under the medium marker density. For BayesA (Figure 3.6) we determined no significant difference in accuracies when s^2 was understated (i.e., $s^2 = 0.1 \bar{s}_{med}^2$ and $s^2 = 0.01 \bar{s}_{med}^2$); however, breeding value accuracies were significantly compromised when s^2 was overstated ($P < 0.01$), particularly at $s^2 = 100 \bar{s}_{med}^2$ ($P < 0.0001$). For BayesB (Figure 3.7), we did not see any significant differences in accuracy of prediction between the various scenarios.

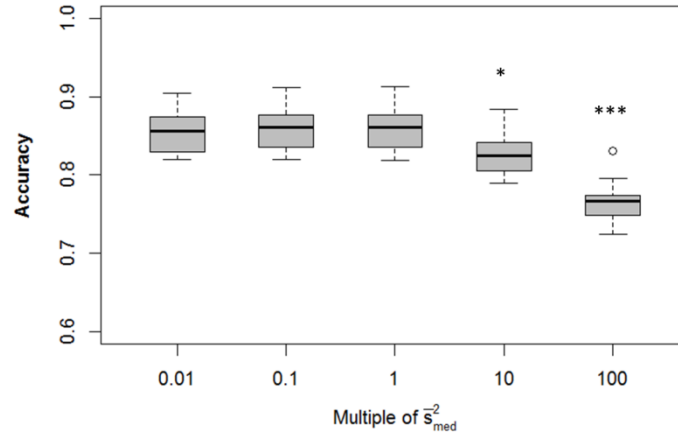


Figure 3.6: Boxplots of accuracies of breeding value predictions under BayesA model across 15 replicates at medium marker density (pairwise $r^2=0.24$) for s^2 set equal to different multiples of average posterior mean of s^2 ($\bar{s}_{med}^2 = 7 \times 10^{-4}$) in a fully Bayes analysis. Significant differences with 1 \bar{s}_{med}^2 are indicated by *($P < 0.01$), ***($P < 0.0001$).

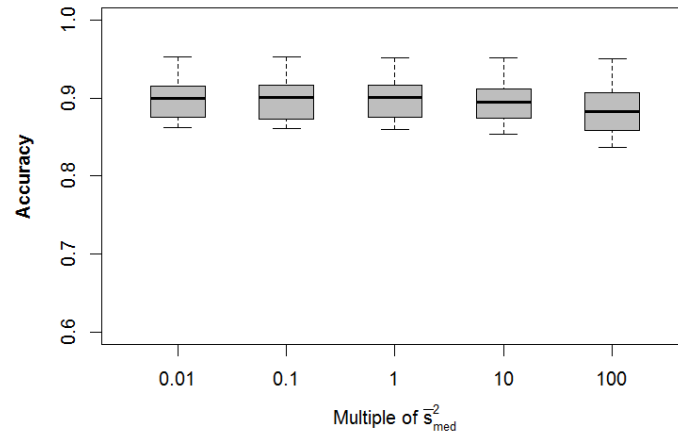


Figure 3.7: Boxplots of accuracies of breeding value predictions under BayesB model across 15 replicates at medium marker density (pairwise $r^2=0.24$) for s^2 set equal to different multiples of average posterior mean of s^2 ($\bar{s}_{med}^2 = 4 \times 10^{-2}$) in a fully Bayes analysis. Significant differences with 1 \bar{s}_{med}^2 are indicated by *($P < 0.01$), ***($P < 0.0001$).

We wondered if some of the non-significant differences in these comparisons could be partly attributed to a compensation in the inferences on other hyperparameters, specifically ν and π (in BayesB). Indeed we noted that as the specification on s^2 increased from $0.01 \bar{s}_{med}^2$ to $100 \bar{s}_{med}^2$, the posterior means of ν also increased under both BayesA (Figure 3.8) and BayesB (Figure 3.9). This was somewhat anticipated given the high posterior correlation attributed between these two hyperparameters. Note from the prior specification on $\{\sigma_{g_j}^2\}_{j=1}^m$ that $E(\sigma_{g_j}^2 | \sigma_{g_j}^2 > 0) = \frac{\nu s^2}{\nu - 2}$; that is, the average values of the MCMC draws of $\{\sigma_{g_j}^2\}_{j=1}^m$ will be somewhat constrained by $\frac{\nu s^2}{\nu - 2}$. So if s^2 is understated, the estimate of ν ($\nu > 2$) should decrease accordingly to compensate such that there is a good deal of flexibility in maintaining the value of $\frac{\nu s^2}{\nu - 2}$. However, if s^2 is overstated, then there is very little flexibility to accordingly bring down $\frac{\nu s^2}{\nu - 2}$ with an increased value of ν since $\frac{\nu s^2}{\nu - 2}$ can never be less than s^2 . We believe this is the reason why understating the value of s^2 is less serious than overstating it, at least for BayesA, as further indicated by our results. The misspecification also impacts estimates of π in a BayesB as further illustrated in Figure 3.10. This provides BayesB even more flexibility than BayesA for misspecification of s^2 ; that is, overstated values of s^2 merely distribute the number of non-zero $\{g_j\}_{j=1}^m$ over a few number of markers as indicated by higher values of π . This may be a key reason why we noticed non-significant differences for accuracy of breeding value prediction in BayesB between various specifications of s^2 in

Figure 3.7. Misspecification of hyperparameters could then be another reason why BayesB often outperforms BayesA in many other comparisons.

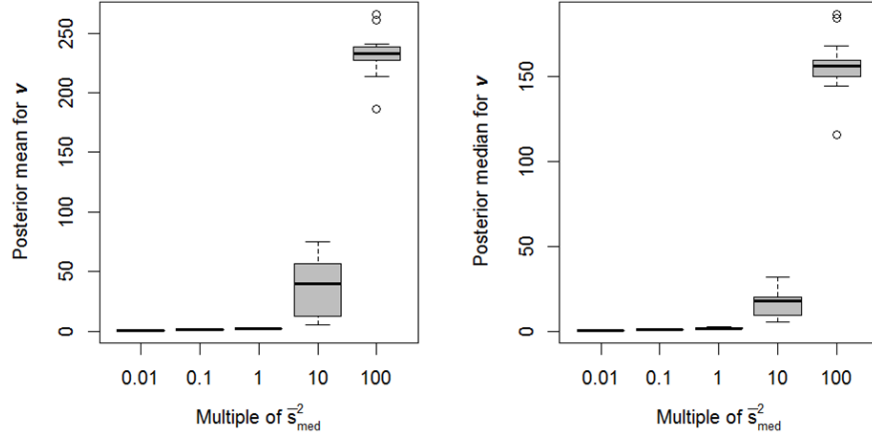


Figure 3.8: Boxplots of posterior mean and median for ν under BayesA model across 15 replicates at medium marker density (pairwise $r^2=0.24$) for s^2 set equal to different multiples of average posterior mean of s^2 ($\bar{s}_{med}^2=7 \times 10^{-4}$) in a fully Bayes analysis.

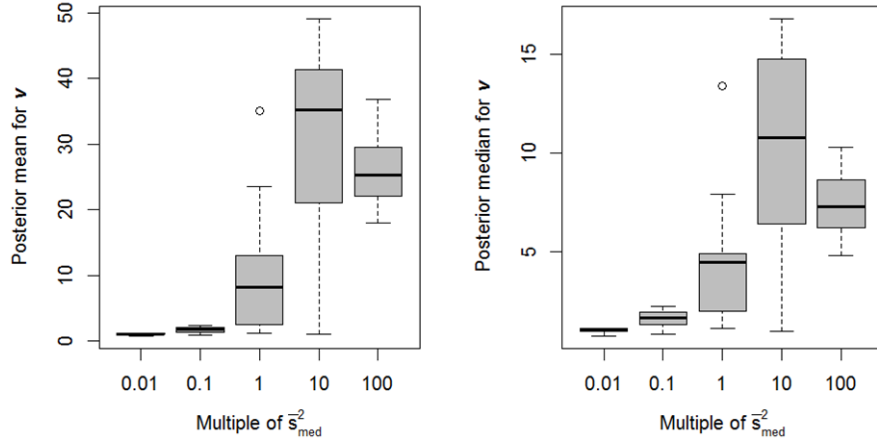


Figure 3.9: Boxplots of posterior mean and median for ν under BayesB model across 15 replicates at medium marker density (pairwise $r^2=0.24$) for s^2 set equal to different multiples of average posterior mean of s^2 ($\bar{s}_{med}^2=4 \times 10^{-2}$) in a fully Bayes analysis.

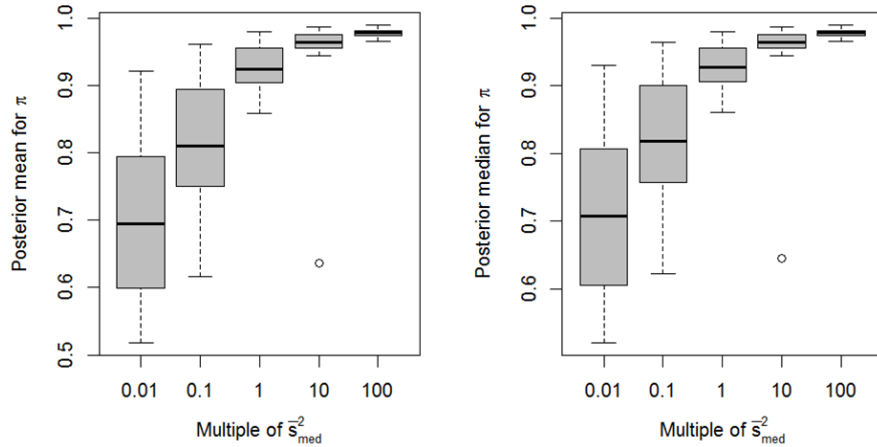


Figure 3.10: Boxplots of posterior mean and median for π under BayesB model across 15 replicates at medium marker density (pairwise $r^2=0.24$) for s^2 set equal to different multiples of average posterior mean of s^2 ($\bar{s}_{med}^2=4 \times 10^{-2}$) in a fully Bayes analysis.

We also wondered if estimates of s^2 based on analysis derived from certain marker densities could be extrapolated to other marker densities for analysis of the same phenotypes. Recall that $\bar{s}_{med}^2 = 7 \times 10^{-4}$ for BayesA, $\bar{s}_{med}^2 = 4 \times 10^{-2}$ for BayesB with the medium marker density panel. For the high marker density involving 4 times as many markers, we specified $s^2 = 4 \bar{s}_{med}^2$ whereas for the low marker density panel which had half as many markers, we specified $s^2 = 0.5 \bar{s}_{med}^2$. We found no significant differences in accuracies in any case (see Figures 3.11 and 3.12) except for a significant lower accuracy for extrapolation on s^2 at the higher marker density using BayesA (P=0.04).

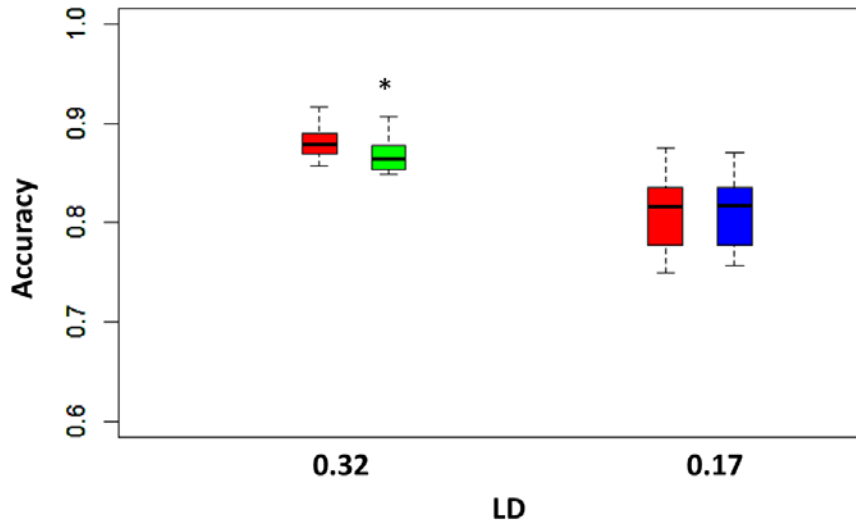


Figure 3.11: Accuracy of breeding value predictions under BayesA model across 15 replicates at high and low marker densities (pairwise LD $r^2=0.32$ and 0.17) using DFMH (red), DFMH with fixed scale $s^2=7 \times 10^{-4}/4=1.75 \times 10^{-4}$ (green) at high marker density and DFMH with fixed scale $s^2=7 \times 10^{-4} \times 2.5=1.75 \times 10^{-3}$ (blue) at low marker density. Significant difference in accuracy between DFMH (red) and DFMH with $s^2=1.75 \times 10^{-4}$ (green) was found at $P < 0.05$.

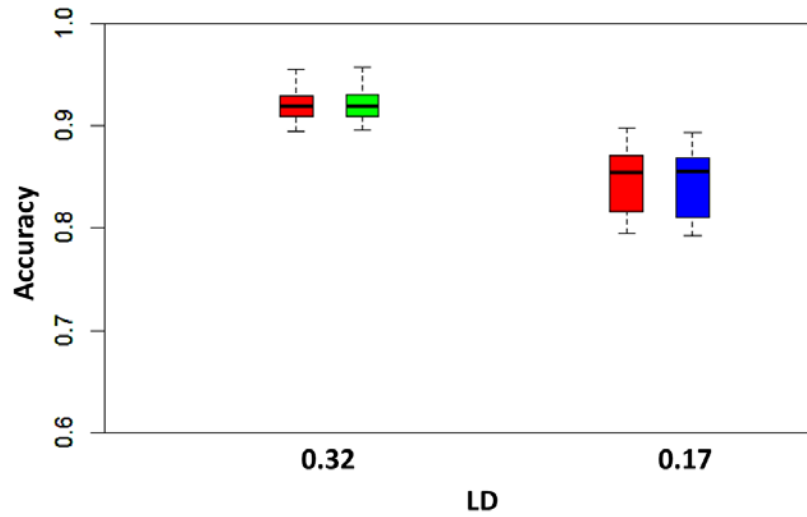


Figure 3.12: Accuracy of breeding value predictions under BayesB model across 15 replicates at high and low marker densities (pairwise LD $r^2=0.32$ and 0.17) using DFMH (red), DFMH with fixed scale $s^2=0.04/4=0.01$ (green) at high marker density and DFMH with fixed scale $s^2=0.04 \times 2.5=0.1$ (blue) at low marker density. No significant differences in accuracy among each pair of methods were found.

3.3.2 Application to Heterogeneous Stock Mice data

We summarize posterior inferences for the key hyperparameters under BayesA and BayesB analyses of the heterogeneous stock mice data in Appendix Table B2.1, Table B2.2 and Table B2.3 for the three marker densities: 950, 1900 and 3800 SNP. For the 950 SNP marker analysis (Appendix Table B2.1), posterior means of parameters were close to estimates previously provided for this same data by Yang and Tempelman (2012). For all analyses, the number of MCMC cycles post burn-in was the same. Under the BayesA model, the ESS for s^2 was twice as large using UNIMH and BIVMH compared to DFMH, while the ESS for ν was four times larger in UNIMH and BIVMH compared to DFMH; the ESS for s^2 and for ν were similar between UNIMH and BIVMH. Under the BayesB model, the ESS for ν using UNIMH and BIVMH was 7-8 times greater for ν and 2 times greater for π than that for DFMH. For the 1900 marker panel (Appendix Table B2.2), the ESS for UNIMH and BIVMH compared to DFMH were between 10-14 times greater for ν and were both around 3 times greater for s^2 under BayesA than for BayesB model, these ratios were respectively between 10 and 15 for ν , and close to 2 for s^2 and π . For the 3800 marker panel (Appendix Table B2.3), these respective ratios were between 12 and 13 for ν and around 4 for s^2 using BayesA, whereas they were between 10 and 15 for ν , about 3 for s^2 and about 4 for π using BayesB.

3.4 Discussion

Most researchers don't typically infer upon key hyperparameters (i.e., ν , s^2 and π) that partly determine the genetic architecture in BayesA/B WGP models. This is in part

due to the high posterior correlation that exists between some of these hyperparameters, in particular, ν and s^2 (Habier, Fernando et al. 2011; de los Campos, Hickey et al. 2013). Nevertheless, some (Riedelsheimer, Technow et al. 2012; Technow, Riedelsheimer et al. 2012; Technow and Melchinger 2013) have been successful in using techniques previously presented by Yi and Xu (2008) and Yang and Tempelman (2012) to infer upon these hyperparameters, and which closely mirrors that labeled DFMH as described in this paper.

We considered two alternative sampling strategies to DFMH, each involving the use of MH, in an attempt to improve the computational efficiency of WGP models as measured by the ratio ESS/CPU. Using simulation studies and empirical data analyses, we demonstrated that strategies borrowing more heavily on MH sampling had better computational efficiencies compared to DFMH. Simple modifications such as sampling s^2 with a MH rather than a Gibbs step (UNIMH) or joint sampling of s^2 and ν with a bivariate MH step (BIVMH) lead to substantial improvements in ESS/CPU. We concede that our investigation is not exhaustive with respect to assessing all possible strategies to improve computational efficiency in these models; in fact, there may be a hybrid involving some or all of the three presented sampling strategies that might be computationally more efficient. Deviations of MH sampling such as Langevin-Hastings could also have been explored and assessed here as well although its advantage relative to MH sampling has not been too convincing in other animal breeding models (Shariati and Sorensen 2008; Waagepetersen, Ibanez-Escriche et al. 2008). In other work that we do not report here, we attempted to base the covariance matrix for the proposal density in BIVMH on the negative Hessian of the joint FCD of $\log(\nu)$ and $\log(s^2)$. However, we

determined this matrix to be positive definite generally only when $\nu < 50$, thereby negating its use in this way. Recently, non-MCMC (i.e. expectation-maximization) schemes have been increasingly popular; however, it is often not straightforward how to estimate key hyperparameters in these implementations (KARKKAINEN and SILLANPAA 2012). In any case, we encourage further development and work in this area including the Bayesian LASSO model (de los Campos, Naya et al. 2009).

We have previously demonstrated that it may be advantageous to specify non-stationary correlation structures between adjacent SNP using a first-order antedependence specification (YANG and TEMPELMAN 2012). In work not reported here, we also evaluated the three alternative sampling strategies in the context of antedependence versions of BayesA and BayesB and drew conclusions virtually identical to what we draw here.

Overspecifying s^2 appeared to have deleterious effects on accuracy of genomic selection using BayesA models although no such effect was observed in BayesB models likely due to the counteracting influence of π . It appeared that underspecification of s^2 lead to more robust genomic predictions as there is greater flexibility for inference on π and s^2 to compensate for this. We also determined that it may be reasonable to consider specifying values for s^2 for one marker density based on a previous estimate from another marker density by taking into account the direct inverse relationship between s^2 and marker density.

At any rate, it should be fully appreciated that these hyperparameters should not be arbitrarily specified in BayesA models. We anticipate that these issues are also

pertinent to determining tuning parameters for various nonparametric approaches as well. We do recognize however computational challenges may be formidable for marker density panels that far exceed those that we considered in this paper. At the very least then, some hyperparameters should be estimated based on simple model-based approximations; for example, s^2 in BayesA should not be much different in magnitude from the variance component for SNP effects in a GBLUP (Meuwissen, Hayes et al. 2001) analysis; hence, a REML-like estimator could be used to provide a reasonable specification. If this is deemed to be computationally intractable relative to the marker density, then extrapolations based on analyses based on lower marker densities might be pursued similar to those presented in this paper.

3.5 Conclusions

In WGP Bayesian hierarchical models, log transformation and jointly drawing v and s^2 can improve MCMC efficiency for inference on all hyperparameters. Even separate univariate MH draws on v and s^2 is substantially more efficient than Gibbs sampling of s^2 . Overspecification of key hyperparameters s^2 can reduce accuracy of breeding value prediction under BayesA model. BayesB model is more robust to misspecification of s^2 due to inference on association probability π . However, it's important to estimate all hyperparameters since misspecification of s^2 can lead to poor inference on v and π .

Chapter 4 Random regression and reaction norm extensions of whole genome prediction models to account for genotype by environment interaction

4.1 Introduction

Whole genome prediction (WGP) has become a revolutionary process for selecting animals and plants for genetic merit on economically important traits using high density single nucleotide polymorphism (SNP) markers (Meuwissen, Hayes et al. 2001). Many WGP methods have been investigated to improve accuracy of breeding value (BV) prediction (de los Campos, Hickey et al. 2013). Meuwissen et al. (2001) proposed two hierarchical Bayesian methods, i.e. scaled- t density prior with and without point mass at zero, namely BayesB and BayesA, respectively. To infer upon key hyperparameters, fully hierarchical Bayesian WGP approaches based on BayesA have been developed and applied in many studies (Yi and Xu 2008; Jia and Jannink 2012; Yang and Tempelman 2012).

Genotype by environment ($G \times E$) interaction refers to how genotypes influence phenotypes differentially in different environments (FALCONER 1952). That is to say, the genetic merit, even ranking, of animals for certain quantitative traits could be substantially different across different environments. The existence of $G \times E$ has been found for various traits in various livestock and plant species (Deeb and Cahaner 2001; Berry, Buckley et al. 2003; Beerda, Ouweltjes et al. 2007; Bohmanova, Misztal et al. 2008; Knap and Su 2008; Hadjipavlou and Bishop 2009; Lillehammer, Hayes et al. 2009). Recently, it has been determined that some SNP and hence quantitative trait loci (QTL) effects are different across environments (Lillehammer, Arnyasi et al. 2007;

Lillehammer, Odegard et al. 2007; Lillehammer, Goddard et al. 2008; Lillehammer, Hayes et al. 2009). In fact, Lillehammer et al. (2007) determined in their analyses that some QTL may not have been otherwise inferred without allowing for $G \times E$. However, little work has been considered to jointly model SNP effects across different environments under a WGP framework (Burgueno, de los Campos et al. 2012). Burgueno et al. (2012) adopted factor analytic models to account for $G \times E$ based on SNP and/or pedigree derived relationships. Their model did not consider information due to environmental covariates that might potentially drive $G \times E$. If $G \times E$ is present, but is not considered in WGP models, then selection of animals for certain environments could be suboptimal. The existence of $G \times E$ further complicates the process of WGP validation; that is, sometimes the effects of markers estimated under one population (i.e., training set) are retested in another population or environment (i.e. validation set) (Daetwyler, Calus et al. 2013); a clear example is the use of parental genotypes and data as training data with progeny genotypes and data used as validation within the context of future environments as validation. If $G \times E$ effects are important, then this validation strategy may not work as intended.

Random regression (RR) and reaction norm (RN) models have played an important role in detecting $G \times E$ of a linear or even higher order nature (Calus, Groen et al. 2002; Berry, Buckley et al. 2003; Calus and Veerkamp 2003; Mattar, Silva et al. 2011; Cardoso and Tempelman 2012; Streit, Reinhardt et al. 2012). RR models have been typically used for modeling genetic merit of traits with repeated measurements over time (Berry, Buckley et al. 2003) whereas RN models have been applied to quantitative traits where genetic merit is typically modeled as a function of key environmental covariate(s)

(Streit, Reinhardt et al. 2012). For both RR and RN models, BV can be modeled as function of an intercept, reflecting an average environment, and a linear function of a key environmental covariate (Calus, Groen et al. 2002). In QTL mapping studies, the QTL specific intercept and slope effects of environmental covariates have been modeled to account for G×E (Lillehammer, Arnyasi et al. 2007; Lillehammer, Odegard et al. 2007; Lillehammer, Goddard et al. 2008). In a genome wide association study (GWAS) focusing on the detection of G×E, SNP specific intercept and slope effects of environmental covariates have been modeled (Lillehammer, Hayes et al. 2009). With increasing availability of high marker densities in WGP, we develop genomic RR/RN models by specifying SNP substitution effects as random intercept and linear functions of age or environmental covariates in a manner similar to Streit et al. (2013).

In “Bayesian alphabet” WGP models like BayesA or BayesB (MEUWISSEN and GODDARD 2010), SNP specific genetic variances are modeled. These variances are of no inherent interest but are used to specify a distribution for the SNP effects that are heavier tailed (e.g. Student t) than Gaussian. For RR/RN models used to specify G×E in WGP, 2x2 genetic variance-covariance matrices (VCV) of the SNP-specific intercepts and slopes are modeled. Conjugate priors on these trait-specific VCV, such as independent inverted Wishart (IW) densities, have been used for bivariate genomic analyses (CALUS and VEERKAMP 2011), thereby rendering marginal distributions on SNP intercept and slope effects as bivariate Student t . An alternative specification, the square root free Cholesky decomposition (CD) of the VCV, has been applied in bivariate trait analyses to model random and residual variance-covariance matrices (Bello, Steibel et al. 2010). The CD specification re-parameterizes VCV into generalized autoregressive parameters

(GARP) and innovation variances to provide potentially greater flexibility relative to IW based specifications (POURAHMADI 1999).

In this paper, we develop and test five possible RR/RN models based on IW or CD based specifications. The objectives of this study were to compare these five models to two conventional BayesA and BayesB models for assessing the accuracy of BV prediction in WGP, and to compare the five RR/RN models in the ability to detect G×E of a linear nature.

4.2 Materials and Methods

4.2.1 Random regression and reaction norm models

The random regression (RR) model for WGP can be denoted as follows:

$$y_{ik} = \mathbf{x}'_i \boldsymbol{\beta} + \mathbf{z}'_i (\mathbf{g}_1 + d_{ik} \mathbf{g}_2) + a_i + e_{ik}, \quad [1]$$

where y_{ik} is the k^{th} phenotype record for i^{th} animal ($k=1,2,\dots,t$; $i=1,2,\dots,n_i$); $\boldsymbol{\beta}$ is the vector of fixed effects, \mathbf{x}'_i is the incidence row vector connecting elements of $\boldsymbol{\beta}$ to animal i ; $\mathbf{z}'_i = [z_{i1} \ z_{i2} \ z_{i3} \ \dots \ z_{im}]$ is the vector of genotypes coded as 0, 1, or 2 copies of the minor allele on SNPs for animal i ; $\mathbf{g}_1 = \{g_{1j}\}_{j=1}^m$ is the vector of SNP-specific intercept effects; $\mathbf{g}_2 = \{g_{2j}\}_{j=1}^m$ is the vector of SNP-specific temporal slope effects; d_{ik} is the environmental covariate for record k on animal i and e_{ik} is the random residual. Finally, a_k is the random effect of animal characterized by a variance component σ_a^2 . This particular random effect may be either a specification on residual polygenic or permanent environmental effects or both; nevertheless, it is particularly required in addition to the

other terms when there are repeated records per animal. The environmental covariates d_{ik} are assumed to be known without error. In Equation [1], one could write the total breeding value (BV) for animal i in an environment characterized by covariate d_{ik} as $\mathbf{z}'_i(\mathbf{g}_1 + d_{ik}\mathbf{g}_2)$ which, in turn, is a function of the intercept BV $\mathbf{z}'_i\mathbf{g}_1$ and the slope BV $\mathbf{z}'_i\mathbf{g}_2$. One could then think of $\mathbf{z}'_i\mathbf{g}_1$ as being a measure of overall genetic merit (if d_{ik} is recentered to an average of 0) for animal i whereas $\mathbf{z}'_i\mathbf{g}_2$ is a the genetic merit of that same animal's environmental sensitivity; i.e., the greater the value of $|\mathbf{z}'_i\mathbf{g}_2|$, the greater the sensitivity of that animal's genetic merit to different environments as represented by different values of d_{ik} .

We further write the RR-WGP model in matrix notation as follows:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{g}_1 + \mathbf{DZ}\mathbf{g}_2 + \mathbf{W}\mathbf{a} + \mathbf{e}, \quad [2]$$

where $\mathbf{e} = \{e_{ik}\}$, \mathbf{D} is a nt by nt diagonal matrix with the environmental covariates $\{d_{ik}\}$ along the diagonal, $\mathbf{X} = \{\mathbf{x}'_i \otimes \mathbf{1}_{tx1}\}_{i=1}^n$, and $\mathbf{Z} = \{\mathbf{z}'_i \otimes \mathbf{1}_{tx1}\}_{i=1}^n$ with \otimes denoting the Kronecker product. We assume $\mathbf{e} \sim N(\mathbf{0}, \mathbf{R} = \mathbf{I}\sigma_e^2)$ in this paper although generalizations to heterogeneous residual variances over time would be possible too. Also, we might specify arbitrarily informative or diffuse priors $p(\boldsymbol{\beta})$ on “fixed effects” $\boldsymbol{\beta}$, $p(\sigma_e^2)$ on σ_e^2 and $p(\sigma_a^2)$ on σ_a^2 .

Reaction norm (RN) models could somewhat be considered simplifications of RR models whereby typically only one phenotypic record (y_i) is observed per animal, although the environmental covariate (d_i) unique to animal i may vary across animals

(i.e., $t = 1$). Sometimes these covariates need to also be inferred (Su, Madsen et al. 2006) but for the purposes of this study we'll also consider d_i as known. The RN model can then be written as:

$$y_i = \mathbf{x}_i' \boldsymbol{\beta} + \mathbf{z}_i' (\mathbf{g}_1 + d_i \mathbf{g}_2) + e_i, \quad [3]$$

As with the RR model in Equations [1] and [2], we could add an effect for animal, or equivalently, residual polygenic effects based on a known correlation (i.e., numerator relationship) matrix, \mathbf{A} , between animals if the number of SNP is not considered to be large enough to model genetic variability.

4.2.2 Conventional BayesA and BayesB (BayesA\BayesB)

In conventional WGP models, all elements of \mathbf{g}_2 are zero. BayesB specifies a mixture prior of a point mass at zero with non-association probability $(1 - \pi)$ and a Student- t density with degrees of freedom ν and scale parameter s^2 with association probability π . BayesA is a special case of BayesB when $\pi = 1$ (Meuwissen, Hayes et al. 2001). Priors such as $p(\nu)$, $p(s^2)$, and/or $p(\pi)$ can be specified on these hyperparameters for BayesA or BayesB in order to properly “tune” them or account for their uncertainty as we and others have done previously (Yi and Xu 2008; Technow, Riedelsheimer et al. 2012; Yang and Tempelman 2012) or strongly advocated (JIA and JANNINK 2012).

4.2.3 Bivariate Normality (IW-BayesC)

The simplest specification for SNP-effects and slopes might be based on multivariate normality. Suppose we reorder $\mathbf{g} = [\mathbf{g}_1' \ \mathbf{g}_2']'$ instead as

$\mathbf{g}^* = [g_{11} \ g_{21} \ g_{12} \ g_{22} \ \dots \ g_{1m} \ g_{2m}]' = [\mathbf{g}_{.1}' \ \mathbf{g}_{.2}' \ \dots \ \mathbf{g}_{.m}']$ where $\mathbf{g}_{.j}$ represents the random intercept and slope effects of SNP j . Here $\mathbf{g}_{.j}$ are all specified to be independently multivariate normal with null mean vector and common variance covariate matrix $\Sigma_{\mathbf{g}}$ where

$$\Sigma_{\mathbf{g}} = \begin{bmatrix} \sigma_{\mathbf{g}_1}^2 & \sigma_{\mathbf{g}_1\mathbf{g}_2} \\ \sigma_{\mathbf{g}_1\mathbf{g}_2} & \sigma_{\mathbf{g}_2}^2 \end{bmatrix} \quad [4]$$

This specification, more or less, represents a bivariate extension of what Habier et al. (2011) describes as BayesC π when $\pi = 0$ where π defines the probability of non-association. However, as we illustrate later, this is effectively equivalent to a classical mixed model analysis. One can specify a conjugate Inverted Wishart prior on $\Sigma_{\mathbf{g}}$ with $\Sigma_{\mathbf{g}} \sim IW(v_0, \Sigma_0)$, and we denote this extension as IW-BayesC.

4.2.4 Bivariate Student t and Variable Selection (IW-BayesA\IW-BayesB)

We consider an extension to IW-BayesC, whereby intercept and slope effects are specified to have heterogeneous variance-covariance matrices across SNP. For SNP j , we specify $\mathbf{g}_{.j}$ to be conditionally bivariate normal; i.e.,

$$\mathbf{g}_{.j} \sim \mathbf{N} \left(\mathbf{0}_{2 \times 1}, \mathbf{G}_j = \begin{bmatrix} \sigma_{g_{1j}}^2 & \sigma_{g_{1j}g_{2j}} \\ \sigma_{g_{1j}g_{2j}} & \sigma_{g_{2j}}^2 \end{bmatrix} \right). \quad [5]$$

We then specify all \mathbf{G}_j to have independent conjugate inverted Wishart prior densities; i.e. $\mathbf{G}_j \sim IW(v_{\mathbf{g}}, \Sigma_{\mathbf{g}})$ characterized by a degrees of freedom parameter $v_{\mathbf{g}}$ and scale matrix

$\Sigma_{\mathbf{g}} = \begin{bmatrix} \sigma_{\mathbf{g}_1}^2 & \sigma_{\mathbf{g}_1\mathbf{g}_2} \\ \sigma_{\mathbf{g}_1\mathbf{g}_2} & \sigma_{\mathbf{g}_2}^2 \end{bmatrix}$. We denote this specification as IW-BayesA, noting the obvious

bivariate extension to BayesA as first proposed by Meuwissen et al. (Meuwissen, Hayes et al. 2001). Note that as $v_g \rightarrow \infty$, $\mathbf{G}_j = \Sigma_g \forall j$ such that IW-BayesA reverts back to IW-BayesC. We specify a prior $p(v_g)$ on v_g and a conjugate Wishart prior $\Sigma_g \sim W(v_0, \mathbf{S}_0)$ on Σ_g . As alluded to by Munilla and Cantet (2012) and Bello et al. (2010), the variability of the three components (i.e., 2 SNP-specific variances and a SNP-specific covariance) of \mathbf{G}_j using IW-BayesA/B is primarily controlled by one hyperparameter: v_g .

Mirroring the extension of BayesA to BayesB by Meuwissen et al. (2001), we also further modified IW-BayesA by specifying a mixture prior on \mathbf{G}_j such that $\mathbf{G}_j = \mathbf{0}_{2 \times 2}$ with probability $(1 - \pi)$ and $\mathbf{G}_j \sim \text{IW}(v_g, \Sigma_g)$ with probability π . We name this procedure IW-BayesB specifying a prior $p(\pi)$ on π . This specification is perhaps more dubious than IW-BayesA, given its all or none assumption with respect to SNP effects on both intercept and slope, whereas IW-BayesA likely has more flexibility to specify large SNP effects for, say, the intercept, but near-zero for slope.

4.2.5 Cholesky decomposition specifications (CD-BayesA\CD-BayesB)

Based on our previous experiences, e.g. (Bello, Steibel et al. 2010), we conjectured that specification of inverted Wishart prior densities on \mathbf{G}_j might be rather inflexible as such specifications either imply that all SNP have either non-zero effects for both intercept and slope (IW-BayesA, IW-BayesC) or, if they don't, both effects are 0 (i.e., IW-BayesB), thereby not allowing for the possibility that some SNP effects are overall important (i.e. non-zero intercept) but environmentally robust (i.e., zero slope).

Furthermore, these IW specifications are additionally inflexible in that the heterogeneity of each single component of \mathbf{G}_j is controlled by a single parameter ν_g .

We subsequently developed an alternative parameterization based on the square root free Cholesky decomposition (CD) of \mathbf{G}_j as based on our previous work (Bello, Steibel et al. 2010). The CD parameterization provides potentially greater flexibility by modeling the following relationship between \mathbf{g}_2 and \mathbf{g}_1 :

$$\mathbf{g}_2 = \Psi \mathbf{g}_1 + \mathbf{g}_{2|1} \quad [6]$$

Here $\mathbf{g}_{2|1} = \{g_{2|1,j}\}$ is the vector of SNP-specific slope effects conditional on intercept effects whereas $\Psi = \text{diag}\{\phi_j\}_{j=1}^m$ represents a diagonal matrix of SNP-specific associations between intercept and slope effects. Hence, we can re-write the RN/RR model [2] as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{g}_1 + \mathbf{DZ}(\Psi \mathbf{g}_1 + \mathbf{g}_{2|1}) + \mathbf{e}, \quad [7a]$$

or

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + (\mathbf{Z} + \mathbf{DZ}\Psi)\mathbf{g}_1 + \mathbf{DZ}\mathbf{g}_{2|1} + \mathbf{e}. \quad [7b]$$

Note for SNP j that if $\phi_j \approx 0$, intercept effects are independent of slope effects. If $g_{2|1,j} \approx 0$ and $\phi_j \neq 0$ then intercept and slope effects are perfectly correlated. If $g_{2|1,j} \approx 0$ and $\phi_j \approx 0$, then the SNP is said to be *environmentally robust* (i.e., $g_{2j} \approx 0$).

For SNP j , we specify $g_{1j} \sim N(0, \sigma_{g_{1j}}^2)$ with $\sigma_{g_{1j}}^2 \sim \chi^{-2}(\nu_1, \nu_1 s_1^2)$ whereas $g_{2|1,j} \sim N(0, \sigma_{g_{2|1,j}}^2)$ with $\sigma_{g_{2|1,j}}^2 \sim \chi^{-2}(\nu_2, \nu_2 s_2^2)$. In essence, these two mixtures specify

two separate univariate Student t densities for elements of \mathbf{g}_1 and $\mathbf{g}_{2|1}$. Furthermore, we specify independent normal priors on SNP-specific association parameters between intercept and slope effects; i.e., $\phi_j \sim N(\mu_\phi, \sigma_\phi^2) \forall j$. We label this model as CD-BayesA. Alternatively, let's consider a variable selection extension of CD-BayesA such that $\sigma_{g_{1j}}^2$ and $\sigma_{g_{2|1j}}^2$ have these same respective inverted chi-square priors but with corresponding probabilities π_1 and π_2 such that $\sigma_{g_{1j}}^2 = 0$ and $\sigma_{g_{2|1j}}^2 = 0$ with probabilities $(1 - \pi_1)$ and $(1 - \pi_2)$. For obvious reasons, we then label this model then as CD-BayesB. For both models, we specify diffuse or informative priors on μ_ϕ and on σ_ϕ^2 .

4.2.6 Bayesian inference

In order to conduct fully Bayesian inference using Markov Chain Monte Carlo methods, it is necessary to derive the full conditional densities (FCD) for each unknown parameter to be inferred. For each of the aforementioned RN/RR models, we present these FCD in Appendix C1.

4.2.7 Simulation Study

In order to discern the ability of the various models to differentially fit various naively defined genetic architectures, we conducted a simple simulation study. We targeted six specific scenarios as outlined in Table 4.1. Key specifications were based on an overall or average genetic correlation ($\bar{\rho}_{\mathbf{g}_1\mathbf{g}_2}$) between intercept and slope, as further described and defined in Appendix C1.3, targeting values of $\bar{\rho}_{\mathbf{g}_1\mathbf{g}_2} = 0$, $\bar{\rho}_{\mathbf{g}_1\mathbf{g}_2} = 0.5$, and $\bar{\rho}_{\mathbf{g}_1\mathbf{g}_2} = 0.8$. We also investigated the effects of the number of QTL influencing both

intercept and slope ($M_{both} = 20, 50, \text{ or } 100$), the number of QTL influencing the intercept only ($M_{int} = 20, 50$) and the heritability ($h^2 = 0.20 \text{ or } 0.50$). One may think of M_{int} as the number of *environmentally robust* QTL (i.e., consistent genetic effects across environments) whereas M_{both} denotes the number of *environmentally sensitive* QTL; i.e. QTL whose effects are influenced by environmental effects.

Table 4.1: Summary of six scenarios in LD simulation

| Scenario | Number of QTLs for both intercept and slope (M_{both}) | Number of QTLs for intercept only (M_{int}) | Overall genetic correlation ($\bar{\rho}_{g_1g_2}$) [and range across replicates] | Average heritability (h^2) |
|----------|---|---|--|--------------------------------------|
| 1 | 100 | 0 | 0 [-0.07, 0.07] | 0.5 |
| 2 | 100 | 0 | 0.5 [0.39, 0.61] | 0.5 |
| 3 | 100 | 0 | 0.8 [0.68, 0.85] | 0.5 |
| 4 | 50 | 50 | 0.5 [0.40, 0.66] | 0.5 |
| 5 | 50 | 50 | 0.5 [0.40, 0.66] | 0.2 |
| 6 | 20 | 20 | 0.5 [0.37, 0.71] | 0.5 |

The first three scenarios (Scenarios 1-3) all entailed $M_{both} = 100$ and $h^2 = 0.50$ and characterized situations that seemingly best agreed with an IW-BayesB or a CD-BayesB situation i.e., $M_{int} = 0$. The only differences between each of these three scenarios were differences in $\bar{\rho}_{g_1g_2}$. Scenarios 4-6 seemingly best agreed with the CD-BayesB specifications (i.e., $M_{int} \neq 0$) and were studied in order to assess the effect of h^2 (0.5 vs 0.2 for Scenarios 4 vs. 5) and M_{int} (50 versus 20 for Scenarios 4 vs. 6).

Twenty replicated datasets were generated under each of the six different scenarios. For each replicate, we used the R package HaploSim (Coster, Bastiaansen et al. 2010) to generate 6000 historical generations based on a constant population size of 100 animals as based on 200 unique haplotypes in the base generation. For all cases, the genome was originally composed of one chromosome with 1 Morgan in length and having 100,000 loci. For 20,000 of these loci, the biallelic minor allele frequency was 0.5 whereas the remaining loci (i.e., 80,000) were specified to be monomorphic in the base population. The number of recombinations for each meiosis event was drawn from a Poisson(1) distribution with genomic positions for recombination randomly chosen from a uniform distribution. For the 6000 historical generations, we specified the recurrent mutation rate for all loci as 10^{-5} per locus per generation. After 6000 generations, two additional Generations 6001 and 6002 were generated to expanded to randomly mated population sizes of $n = 2000$ animals each.

For each replicate, we deleted SNP with minor allele frequency (MAF) < 0.05 in Generation 6001. Around 2200 SNPs remained after data editing. The genotype matrix \mathbf{Z} was then based on the number of minor alleles (0, 1 or 2) at each locus for each animal.

We randomly chose $M_{both} + M_{int}$ of these SNP to be QTL. Variances for QTL-specific intercept effects $\sigma_{g_{1j}}^2$ were drawn from scaled inverted chi-square distributions with scale $s_1^2 = 2$ and degrees of freedom $\nu_1 = 5$. Variances for any QTL-specific slope effects conditional on intercept effects, (i.e., $\sigma_{g_{2|j}}^2$) for each of M_{both} QTL, were also drawn from scaled inverted chi-square densities always with $\nu_2 = 5$ and with $s_2^2 = 2$ ($\bar{\rho}_{g_1g_2} = 0$), $s_2^2 = 1.5$ ($\bar{\rho}_{g_1g_2} = 0.5$), or $s_2^2 = 0.72$ ($\bar{\rho}_{g_1g_2} = 0.8$). QTL effects for intercepts $\{g_{QTL,1j}\}_{j=1}^{j=M_{both}+M_{int}}$ and conditional slopes $\{g_{QTL,2|j}\}_{j=1}^{j=M_{both}}$ were then independently generated from normal distributions with null means and their corresponding variances $\sigma_{g_{1j}}^2$ and $\sigma_{g_{2|j}}^2$. Hence, QTL effects for $\{g_{QTL,1j}\}_{j=1}^{j=M_{both}+M_{int}}$ and $\{g_{QTL,2|j}\}_{j=1}^{j=M_{both}}$ were each Student t -distributed.

The association parameters $\{\phi_{QTL,j}\}_{j=1}^{j=M_{both}}$ between intercept and slope for each of M_{both} QTL were generated from independent normal distributions, always with variance $\sigma_{QTL,\phi}^2 = 0.05$ and with mean $\mu_{QTL,\phi} = 0$, for $\bar{\rho}_{g_1g_2} = 0$, $\mu_{QTL,\phi} = 0.5$, for $\bar{\rho}_{g_1g_2} = 0.5$, and $\mu_{QTL,\phi} = 0.8$, for $\bar{\rho}_{g_1g_2} = 0.8$. Subsequently, the effect for slope for QTL j was determined to be $g_{QTL,2j} = g_{QTL,1j} + \phi_{QTL,j}g_{QTL,2|j}$; $j=1,2,\dots,M_{both}$. An environmental covariate, d_i , unique to each animal, was randomly drawn from $N(0,0.36)$. Writing $\mathbf{Z}_{QTL,int}$ as the subset of the $n \times M_{int}$ SNP genotypes in \mathbf{Z} that are designated to be QTL for intercepts only and $\mathbf{Z}_{QTL,both}$ as the subset of the $n \times M_{both}$ SNP genotypes in \mathbf{Z} that are designated to be QTL for both intercepts and slopes, the true breeding values $\{TBV\}_{i=1}^{i=n}$ in each of

Generations 6001 and 6002 were generated using

$$\{TBV_i\}_{i=1}^{i=n} = \mathbf{Z}_{QTL,both} \left\{ g_{QTL,1j} \right\}_{j=1}^{j=M_{both}} + \mathbf{Z}_{QTL,int} \left\{ g_{QTL,1j} \right\}_{j=M_{both}+1}^{j=M_{both}+M_{int}} + \text{diag}(d_i) \mathbf{Z}_{QTL,both} \left\{ g_{QTL,2j} \right\}_{j=1}^{j=M_{both}} \quad [8]$$

Residuals (e_i) were generated for the record on each animal were drawn from a normal distribution with null mean and variance (σ_e^2) as dependent upon h^2 ; i.e

$$\sigma_e^2 = \frac{\text{var}\left(\{TBV_i\}_{i=1}^{i=n}\right)(1-h^2)}{h^2} \quad [9]$$

Hence, phenotypic records were generated as $y_i = TBV_i + e_i$.

Since the existence of GxE would imply that we select for genetic merit tailored for specific environments, we compared the accuracy of predicting TBV among the five aforementioned RN methods as well the two conventional methods (BayesA, BayesB) at three environmental covariate values: $d = -1.2$, $d = 0.0$ and $d = 1.2$ representing -2, 0 and + 2 standard deviations, respectively, for d . Accuracy for a particular value of d was

defined as the correlation between $\{TBV_{i,d}\}_{i=1}^{i=n}$ and $\{EBV_{i,d}\}_{i=1}^{i=n}$ where $\{TBV_{i,d}\}$ for

Generation 6002 animals is based on Equation [8] but with all $d_i = d$ whereas

$$EBV_{i,d*} = \mathbf{z}_i' \hat{\mathbf{g}}_1 + d \mathbf{z}_i' \hat{\mathbf{g}}_2 \text{ for } \mathbf{z}_i' \text{ being the SNP genotypes for animal } i \text{ in Generation 6002}$$

were based on estimated effects using only Generation 6001 data; i.e., $\hat{\mathbf{g}}_1$ and $\hat{\mathbf{g}}_2$ are the respective posterior means for SNP-specific intercepts and slopes. For each of the 5 RR/RN models, we also assessed the relative accuracy of predicting the components of

the BV due to intercept and slope separately in all six scenarios among the five methods. Intercept BV accuracy was determined as the correlation between true intercept BV (first two expressions on right side of Equation 8) and estimated intercept BV ($\mathbf{Z}\hat{\mathbf{g}}_1$) whereas slope BV accuracy was determined as the correlation between true slope BV ($\mathbf{Z}_{QTL,both} \left\{ g_{QTL,2j} \right\}_{j=1}^{j=M_{both}}$) and estimated slope BV ($\mathbf{Z}\hat{\mathbf{g}}_2$). We compared prediction accuracy of BV using a Wilcoxon signed rank test between each pair of the RR/RN and conventional methods.

4.2.8 MSU Pig Resource Population data

The genotypes used for this analysis were based on a commercial platform for low density genotyping (8434 SNP) in swine marketed as the Neogen Porcine GeneSeek Genomic Profiler LD (version 1) (GeneSeek, a Neogen Company, Lincoln, NE) (Badke, Bates et al. 2013). We received complete phenotypes and genotypes information on 928 F2 animals derived from a Duroc \times Pietrain resource population at Michigan State University (Edwards, Ernst et al. 2008; Choi, Steibel et al. 2010). Any SNP with MAF < 0.01 were deleted. For adjacent SNP in complete LD with pairwise $r^2=1$, we deleted one SNP at random from each pair. Then we excluded SNPs having P -value < 10^{-4} for the Hardy Weinberg equilibrium test. Genotypes for the remaining 5271 SNPs were standardized using $(z_{ij} - 2p_j) / \sqrt{2p_j(1-p_j)}$ where z_{ij} is genotype of j^{th} SNP on i^{th} animal and p_j is allele frequency of the reference (“0”) allele for j^{th} SNP (de los Campos, Hickey et al. 2013). Pedigree information was available for the 928 F2 animals including their parents and grandparents. There were 140 unique full-sib families in total among the

928 F2 animals. The average full-sib family size was around 6. Map information was available on SNPs for each of the 18 autosomes from Duarte et al. (Duarte, Bates et al. 2013).

We focused on back fat thickness as our response variable. Back fat thickness was measured at the 10th rib by B-mode ultrasound at weeks 10, 13, 16, 19 and 22 (Edwards et al. 2008). We used the following RR model for data analysis:

$$y_{ijkl} = \mu + \sum_{\omega=1}^4 \beta_{\omega} d_i^{\omega} + sex_j + \sum_{\omega=1}^4 (sex_j \times \beta_{\omega} d_i^{\omega}) + litter_k + u_{1l} + d_i u_{2l} + \mathbf{z}'_l (\mathbf{g}_1 + d_i \mathbf{g}_2) + e_{ijkl} \quad [10]$$

Here, μ is the overall mean; d_i^{ω} is the ω^{th} order power or polynomial ($\omega=1,2,3,4$) on week i with β_{ω} being the corresponding partial regression coefficient, sex_j is the fixed effect of sex of animal j , $(sex_j \times \beta_{\omega} d_i^{\omega})_j$ is the fixed effect of the interaction between d_i^{ω} and sex_j , $litter_k$ is random effect for litter k , u_{1l} is permanent environmental intercept effect for animal l , u_{2l} is the permanent environmental slope effect for animal l , $\mathbf{z}'_l = [z_{l1} \ z_{l2} \ z_{l3} \ \dots \ z_{lm}]$ is the vector of genotypes on animal l , \mathbf{g}_1 is the vector of SNP-specific intercept effects; \mathbf{g}_2 is the vector of SNP-specific slope effects; d_i is the recoded covariate for week i ; e_{ijkl} is the residual effect on y_{ijkl} . We further rescaled d_i as -1, -0.5, 0, 0.5 and 1 for weeks 10, 13, 16, 19 and 22.

Random effect specifications were as follows. Litter effects were presumed to be normally and independently distributed with null mean and variance component σ_{litter}^2 . We decided not to fit polygenic effects since they are strongly confounded with both litter

effects, permanent environment effects and SNP effects and hence would severely impede MCMC performance. The permanent environmental effects $\mathbf{u} = [\mathbf{u}_1' \ \mathbf{u}_2']'$ for $\mathbf{u}_1 = \{u_{1l}\}$ and $\mathbf{u}_2 = \{u_{2l}\}$ were assumed to be multivariate normal with null mean vector and variance-covariance matrix $\mathbf{I} \otimes \mathbf{\Sigma}$ where \mathbf{I} is the identity matrix and $\mathbf{\Sigma}$ is the 2 x 2 unstructured covariance matrix between permanent environmental intercept and slope.

With availability of pedigree information, we also applied a conventional polygenic model (RR-BLUP) as a control-based comparison for the genomic (i.e., SNP) based RR models. Fixed effects, random litter effects and permanent environmental effects were defined the same as in Equation [10], except that we replaced SNP effects by polygenic effects for both intercept and slope as following:

$$y_{ijkl} = \mu + \sum_{\omega=1}^4 \beta_{\omega} d_i^{\omega} + sex_j + \sum_{\omega=1}^4 (sex_j \times \beta_{\omega} d_i^{\omega}) + litter_k + u_{1l} + d_i u_{2l} + a_{1l} + d_i a_{2l} + e_{ijkl} \quad [11]$$

We assumed that the polygenic effects $\mathbf{a} = [\mathbf{a}_1' \ \mathbf{a}_2']'$ for intercept $\mathbf{a}_1 = \{a_{1l}\}$ and slope $\mathbf{a}_2 = \{a_{2l}\}$ have a joint multivariate normal distribution with null mean vector and variance-covariance matrix $\mathbf{A} \otimes \mathbf{\Sigma}_a$, where \mathbf{A} is the numerator relationship matrix and $\mathbf{\Sigma}_a$ is the 2 x 2 unstructured covariance matrix between polygenic intercept and slope effects.

All of the five aforementioned RR models along with RR-BLUP were compared to conventional WGP specifications (i.e. BayesA and BayesB) based on 20 random across litter cross validation splits of the data using Wilcoxon signed rank test. That is, for each split, data on all individuals from 126 (90%) litters were randomly chosen as the

training data whereas the data on all subjects in the remaining 14 (10%) litters were chosen to be validation data. In this cross validation, hyperparameters and variance components were fixed to estimates derived from analyses of the entire data. In all cases, except for RR-BLUP, inferences were based on posterior means of the MCMC samples from the posterior distributions. To expedite analyses under RR-BLUP model, we used ASReml 3.0 to provide analytical BLUP solutions rather than using MCMC, noting that conditionally on variance components, these MCMC and BLUP inferences are identical. For all eight models, the Pearson correlation between these predictions and the actual training data was used as a measure of performance of the competing methods.

4.2.9 Priors used for data analyses

For all analyses in this paper, we specified a vaguely informative prior $\nu \sim p(\nu) \propto (\nu + 1)^{-2}$ and Gelman's prior $s^2 \sim \chi^{-2}(-1, 0)$, which is also informative when bounded, for conventional BayesA as we've done previously (YANG and TEMPELMAN 2012). For conventional BayesB, proper conjugate priors $s^2 \sim \text{Gamma}(0.1, 0.1)$ and $\pi \sim p(\pi | \alpha_\pi, \beta_\pi) = \text{Beta}(\alpha_\pi = 1, \beta_\pi = 8)$ were used. For IW-BayesC, a conjugate inverted Wishart prior was specified on scale matrix Σ_g with $\Sigma_g \sim IW(\nu_0 = -3, \mathbf{S}_0 = \mathbf{0}_{2 \times 2})$. For IW-BayesA and IW-BayesB, we specified a prior on ν_g such that $p(\nu_g) \propto (\nu_g + 1)^{-2}$ and a conjugate Wishart prior; i.e., $\Sigma_g \sim W(\nu_0, \mathbf{S}_0)$ on the scale matrix Σ_g . As in conventional BayesB, the same proper prior was specified on π , i.e. $\pi \sim \text{Beta}(\alpha_\pi = 1, \beta_\pi = 8)$ in IW-BayesB.

In CD-BayesA and CD-BayesB, we used the same non-informative priors

$p(\nu_1) \propto (\nu_1 + 1)^{-2}$ and $p(\nu_2) \propto (\nu_2 + 1)^{-2}$ for intercept and conditional slope effects. In addition, Gelman's prior $s_1^2 \sim \chi^2(-1, 0)$ and $s_2^2 \sim \chi^2(-1, 0)$ were used in CD-BayesA, while $s_1^2 \sim \text{Gamma}(0.1, 0.1)$ and $s_2^2 \sim \text{Gamma}(0.1, 0.1)$ were specified in CD-BayesB.

For CD-BayesB, we specified the proper priors $\pi_1 \sim \text{Beta}(\alpha_{\pi_1} = 1, \beta_{\pi_1} = 8)$ and

$\pi_2 \sim \text{Beta}(\alpha_{\pi_2} = 1, \beta_{\pi_2} = 8)$. For both CD-BayesA and CD-BayesB, mean of SNP specific association parameters between intercept and slope μ_ϕ was specified using

$p(\mu_\phi) = N(\tau = 0, \zeta^2 = 1)$, and variance for association parameters σ_ϕ^2 was specified with

Gelman's prior $p(\sigma_\phi^2) = \chi^2(-1, 0)$.

4.3 Results

4.3.1 Simulation Study

In Figure 4.1, we illustrate comparisons of the accuracies for each of the seven competing models at each of three different environmental covariate values: $d = -1.2, 0$ and 1.2 for each of the six scenarios as previously described in Table 4.1. Again, these results were based on 20 replicated datasets for each scenario. We focus first on the comparisons in Scenarios 1-3 where the simulated genetic architecture seemed more congruent with an IW-BayesA or CD-BayesA like specification (i.e., $M_{int} = 0$). When $\bar{\rho}_{g_1, g_2} = 0$ in Scenario 1 (Figure 4.1(1)), there were significant ($P < 0.05$) differences in accuracy of nearly 30 percentage points (i.e., $>84\%$ vs. $<56\%$ at $d = -1.2$, $>82\%$ vs. $<52\%$ at $d = 1.2$) in favor of each the five RN models over the two conventional models

(BayesA/BayesB). Such differences were also significant when $\bar{\rho}_{g_1g_2} = 0.5$ (Scenario 2 in Figure 4.1(2)) and when $\bar{\rho}_{g_1g_2} = 0.8$ (Scenario 3 in Figure 4.1(3)) although these differences became increasingly asymmetric between $d = -1.2$ (i.e., progressively larger) and $d = 1.2$ (i.e., progressively smaller). For each of these three scenarios, IW-BayesA had the highest accuracy relative to all other models in $d = -1.2$ and 1.2 ($P < 0.05$) whereas it did not appear to be different from either of the CD models when $d = 0$. Furthermore, the differences between all of the competing models were trivial ($< 2\text{-}3\%$) between all RN and conventional models at $d = 0$ such that IW-BayesB and IW-BayesC were not even judged to be different from BayesA and BayesB.

We further compared accuracies at $d = -1.2, 0$ and 1.2 in three other scenarios (4, 5, and 6) where the simulated genetic architecture seemed more congruent with a CD-BayesB like specification (i.e., $M_{int} > 0$) based on $\bar{\rho}_{g_1g_2} = 0.5$. In Scenario 4 (Figure 4.1(4)) where $M_{both} = M_{int} = 50$ and $h^2 = 0.5$, both conventional BayesA and BayesB were inferior ($P < 0.05$) to all RN models at both $d = -1.2$ and 1.2 as expected. However, IW-BayesA surprisingly outperformed ($P < 0.05$) all other RN methods at both $d = -1.2$ and 1.2 , although such differences were small (i.e., $< 2\text{-}3\%$). Differences between all models were even smaller at $d = 0$ (i.e., $< 1\text{-}2\%$). The specifications for Scenario 5 (Figure 4.1(5)) differed only from Scenario 4 with a lower $h^2 = 0.2$. In that case, IW-BayesA also had significantly ($P < 0.05$) higher accuracy compared to other RN models at $d = -1.2$ and $d = +1.2$ except for CD-BayesB at $d = +1.2$. At $d = 0$, CD-BayesB ($P < 0.05$) outperformed other models although all differences were very small. The final Scenario 6 (Figure 4.1(6)) differed only from Scenario 4 by lower numbers of QTL (i.e. $M_{both} = M_{int} = 20$). Not only was IW-BayesC significantly much lower in accuracy compared to all other RN

models at all three values of d , it was even lower in accuracy compared to conventional (i.e., BayesA and BayesB) models at $d = 0$. CD-BayesB had the highest accuracy among the five RN methods at each value of d although that difference was not significant compared to IW-BayesA and IW-BayesB at $d = -1.2$.

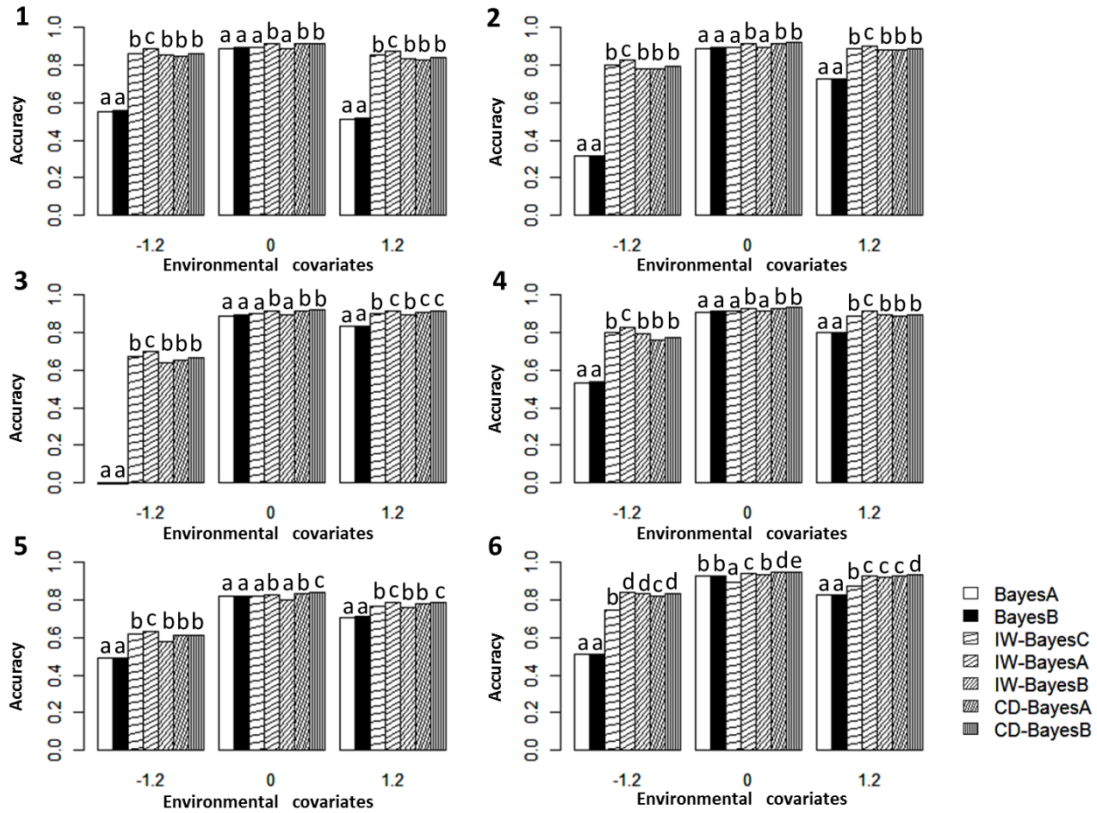


Figure 4.1: Average accuracy of breeding value prediction for seven methods (BayesA\BayesB\IW-BayesA\IW-BayesB\IW-BayesC\CD-BayesA\CD-BayesB) at three environmental covariates in six scenarios, i.e. (1) $M_{\text{both}} = 100$, $M_{\text{int}} = 0$, $\bar{\rho}_{g_1g_2} = 0$, $h^2 = 0.5$; (2) $M_{\text{both}} = 100$, $M_{\text{int}} = 0$, $\bar{\rho}_{g_1g_2} = 0.5$, $h^2 = 0.5$; (3) $M_{\text{both}} = 100$, $M_{\text{int}} = 0$, $\bar{\rho}_{g_1g_2} = 0.8$, $h^2 = 0.5$; (4) $M_{\text{both}} = 50$, $M_{\text{int}} = 50$, $\bar{\rho}_{g_1g_2} = 0.5$, $h^2 = 0.5$; (5) $M_{\text{both}} = 50$, $M_{\text{int}} = 50$, $\bar{\rho}_{g_1g_2} = 0.5$, $h^2 = 0.2$; (6) $M_{\text{both}} = 20$, $M_{\text{int}} = 20$, $\bar{\rho}_{g_1g_2} = 0.5$, $h^2 = 0.5$; Different letters indicate significant difference at $P < 0.05$.

We attempted to better understand these results by focusing on which of the RN methods performed best for inferring upon the intercept component of the BV and the slope component of the BV. For all six simulation scenarios, we found that CD-BayesB always had the significantly highest accuracy for intercept BV compared to all other RN models (Figure 4.2A) whereas IW-BayesC and IW-BayesB were generally amongst the worst. For slope BV accuracy, there was no evidence of differences between any of the models at the low heritability Scenario 5 (Figure 4.2B). However, IW-BayesA did outperform other RN methods in Scenarios 1-4 except for Scenario 3 where CD-BayesB was not found to be inferior to IW-BayesA. In Scenario 6, IW-BayesA only outperformed IW-BayesC. IW-BayesC and IW-BayesB had generally among the lowest slope BV accuracy although CD-BayesA was poorest in Scenario 4. Hence the general advantages for predicting environment-specific BV for CD-BayesB appeared to accrue from its greater accuracy on inferring intercept components of genetic merit whereas that for IW-BayesA appeared to accrue from inferring the slope components.

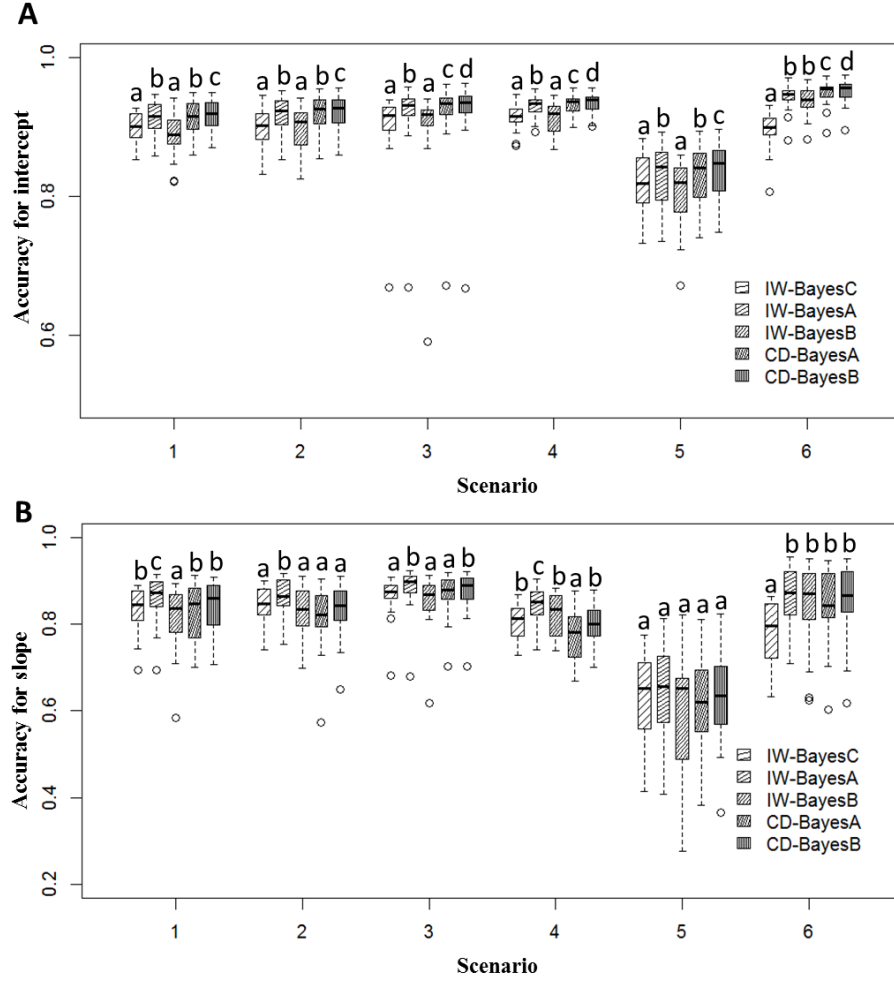


Figure 4.2: Accuracy of intercept (A) and slope (B) breeding value prediction for five RN methods (IW-BayesA\IW-BayesB\IW-BayesC\CD-BayesA\CD-BayesB) under six scenarios: i.e. (1) $M_{\text{both}} = 100$, $M_{\text{int}} = 0$, $\bar{\rho}_{g_1g_2} = 0$, $h^2 = 0.5$; (2) $M_{\text{both}} = 100$, $M_{\text{int}} = 0$, $\bar{\rho}_{g_1g_2} = 0.5$, $h^2 = 0.5$; (3) $M_{\text{both}} = 100$, $M_{\text{int}} = 0$, $\bar{\rho}_{g_1g_2} = 0.8$, $h^2 = 0.5$; (4) $M_{\text{both}} = 50$, $M_{\text{int}} = 50$, $\bar{\rho}_{g_1g_2} = 0.5$, $h^2 = 0.5$; (5) $M_{\text{both}} = 50$, $M_{\text{int}} = 50$, $\bar{\rho}_{g_1g_2} = 0.5$, $h^2 = 0.2$; (6) $M_{\text{both}} = 20$, $M_{\text{int}} = 20$, $\bar{\rho}_{g_1g_2} = 0.5$, $h^2 = 0.5$; Different letters indicate significant difference at $P < 0.05$.

4.3.2 MSU Pig Resource Population data

Predictive ability was calculated as the correlation between observed phenotypes y_{ijkl} in the validation dataset and their predicted values based on inferences from the training dataset. The predictive abilities for each of the eight models for each of the 20 different cross-validation sets are illustrated in Figure 4.3. The five RR Bayesian methods had ~2.5% higher ($P < 0.0001$) predictive ability than the two conventional methods and RR-BLUP model. In addition, no significant differences in predictive ability were found among the five SNP effects based RR methods.

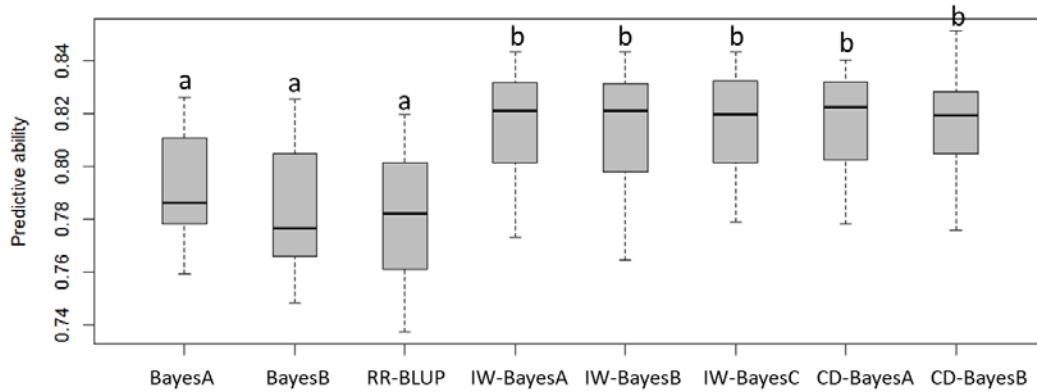


Figure 4.3: Predictive ability for eight methods (BayesA\BayesB\RR-BLUP\IW-BayesA\IW-BayesB\IW-BayesC\CD-BayesA\CD-BayesB) from cross-validation analysis using back fat thickness in MSU Pig Resource Population data. Different letters indicate significant difference at $P < 0.0001$.

In an attempt to better understand these results, we focused on the non-mixture methods (conventional BayesA, IW-BayesA and CD-BayesA) for estimated intercept BV and slope BV respectively. Based on estimated intercept BV using the complete final analyses data, Figure 4.4 showed a relatively high correlation (~ 0.9996) between the two RR methods IW-BayesA and CD-BayesA. In contrast, low correlations were determined between conventional BayesA and IW-BayesA (~ 0.8606) and between conventional BayesA and CD-BayesA (~ 0.8589). As conventional BayesA does not model genomic effects for slope on weeks of age, we can only compare IW-BayesA and CD-BayesA for the estimated slope BV. Figure 4.5 demonstrated that there was a high correlation (~ 0.9845) between IW-BayesA and CD-BayesA for the estimated slope component of BV.

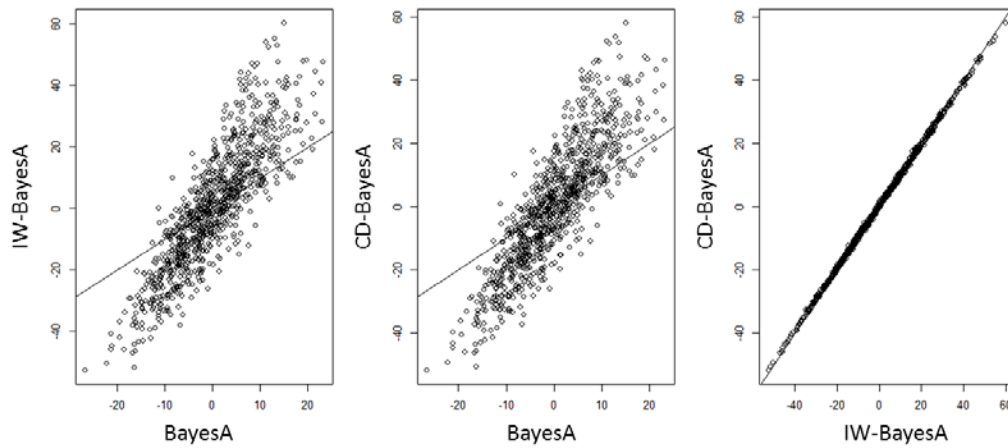


Figure 4.4: Estimated intercept breeding values from conventional BayesA, IW-BayesA and CD-BayesA using the complete final analyses data on back fat thickness in MSU Pig Resource Population. Reference line is $y=x$.

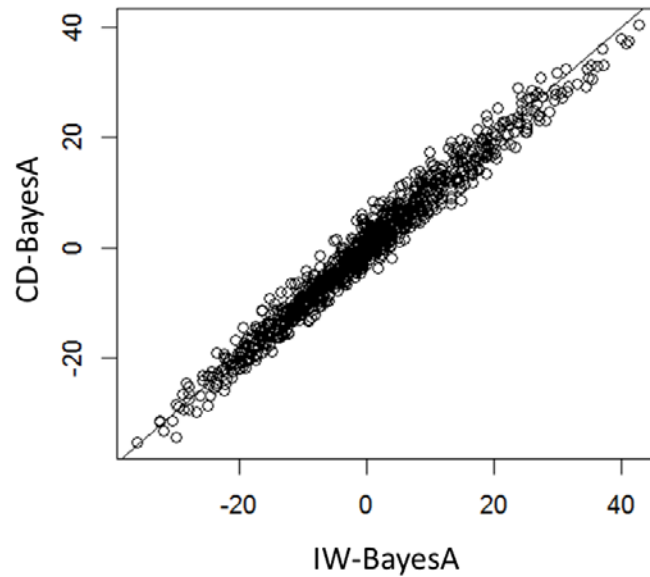


Figure 4.5: Estimated slope breeding values from IW-BayesA and CD-BayesA using the complete final analyses data on back fat thickness in MSU Pig Resource Population. Reference line is $y=x$.

To further investigate the difference in detecting QTLs, we computed the absolute values of posterior means of SNP effects at three different ages (10, 16 and 22 weeks) for these same three methods in Figures 4.6-4.8. In the conventional BayesA model, the SNP effects by necessity are estimated to be the same at any age and hence only one plot was provided; this plot demonstrated that three chromosomes (2, 6, and 11) had some relatively large SNP peaks. IW-BayesA (Figure 4.7) and CD-BayesA (Figure 4.8) also demonstrated peaks in these and other chromosomes at all three ages. However, as might be anticipated, SNP effects under these RR models tended to increase with increasing ages. On chromosome 6, we found two relatively large SNP peaks with IW-BayesA and CD-BayesA, respectively. With estimated SNP intercept and slope effects from IW-BayesA and CD-BayesA, we can further demonstrate regression lines of estimated SNP effects on rescaled weeks of age (-1, -0.5, 0, 0.5, 1) for the two SNP markers (Appendix Figure C2.1).

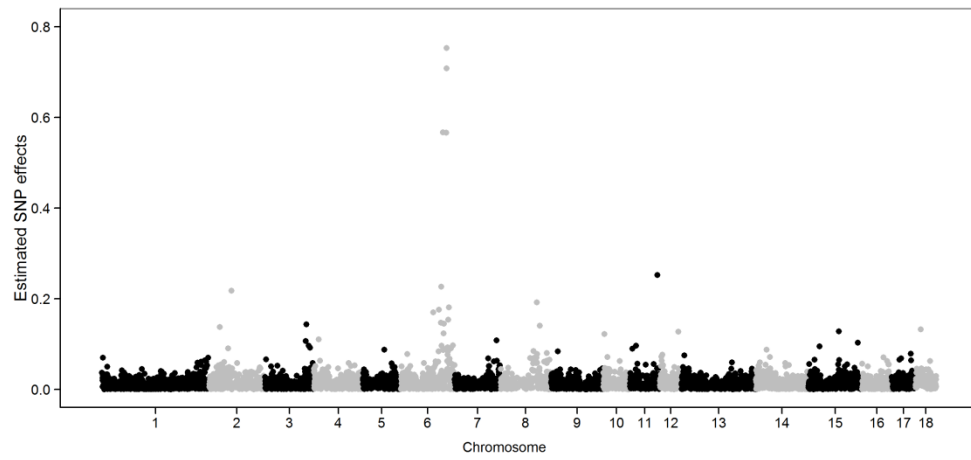


Figure 4.6: Estimated SNP effects from conventional BayesA against marker position using the complete final analyses data on back fat thickness in MSU Pig Resource Population.

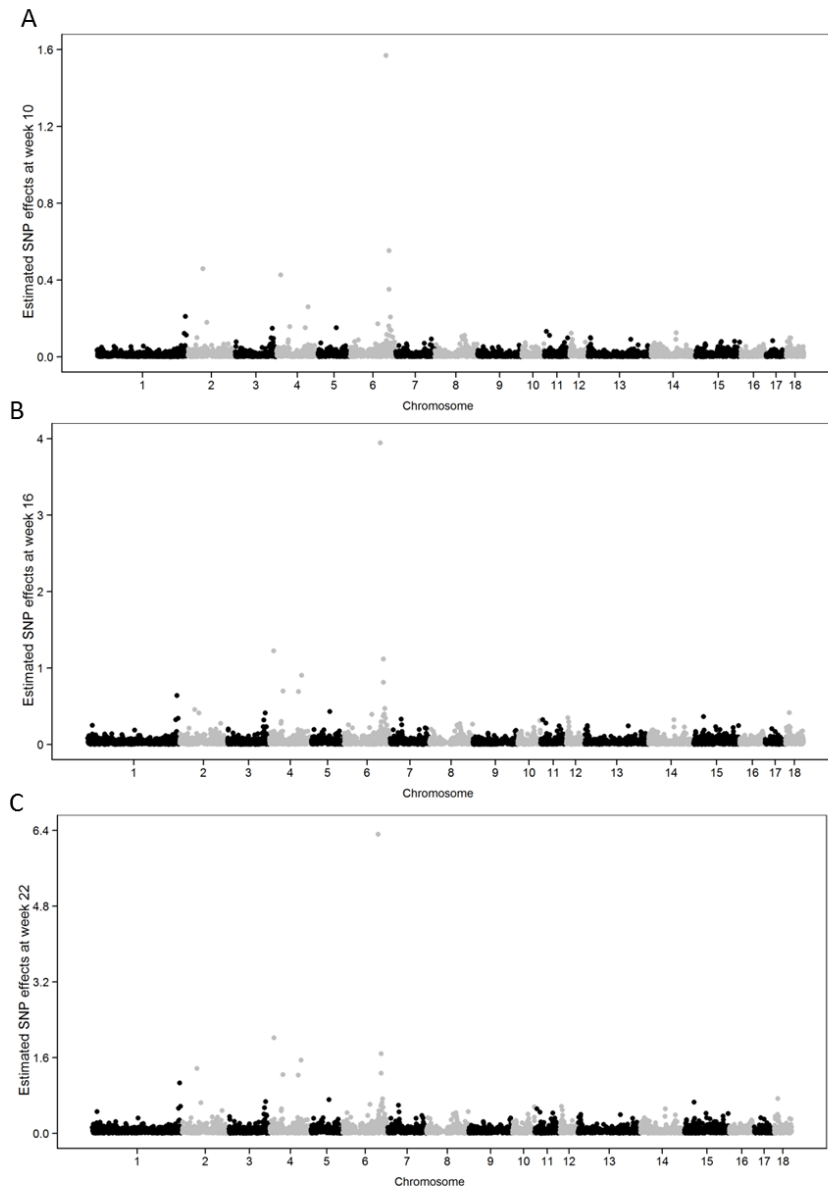


Figure 4.7: Estimated SNP effects from IW-BayesA against marker position using the complete final analyses data on back fat thickness in MSU Pig Resource Population when A) at week 10, B) at week 16, C) at week 22.

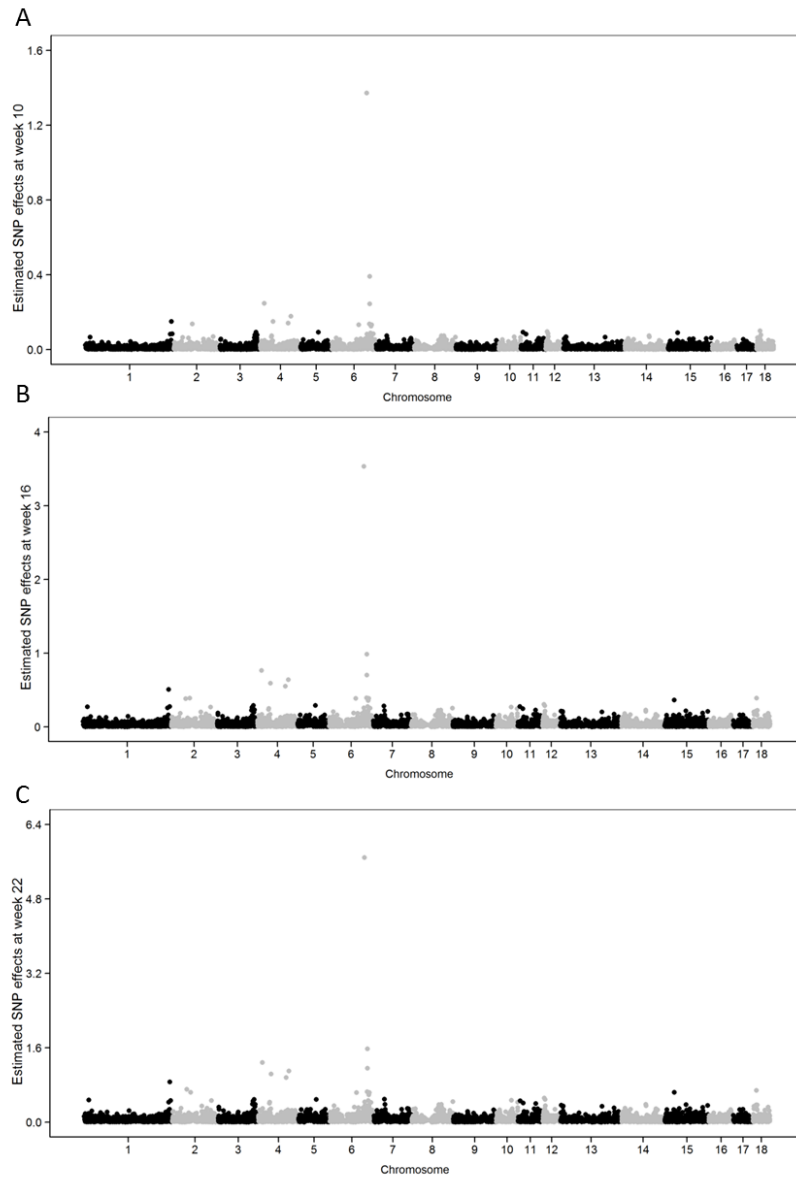


Figure 4.8: Estimated SNP effects from CD-BayesA against marker position using the complete final analyses data on back fat thickness in MSU Pig Resource Population when A) at week 10, B) at week 16, C) at week 22.

4.4 Discussion

Although RR and RN models have been extensively used for modeling G×E in classical polygenic models (CARDOSO and TEMPELMAN 2012), they have not been as extensively adapted for WGP models. Several efforts have been made to infer upon G×E using RN models in QTL mapping and GWAS studies (Lillehammer, Arnyasi et al. 2007; Lillehammer, Odegard et al. 2007; Lillehammer, Goddard et al. 2008; Lillehammer, Hayes et al. 2009). To improve power of QTL detection, Lillehammer et al. (2007) proposed RN models to estimate the QTL intercept and slope effects based on haplotypes with identity by descent (IBD) information. They applied their models to a Norwegian Red cattle population using herd-year mean estimates as environmental covariates (Lillehammer, Arnyasi et al. 2007; Lillehammer, Goddard et al. 2008) which is typical of RN models. Lillehammer et al. (2009) compared their models with and without pedigree information in an Australian dairy bull population to scan one SNP at a time based on genotypes. WGP models (i.e., joint analysis of all SNP) have been advocated to be very important for conducting GWAS (Wang, Misztal et al. 2012); hence, we explored and compared the alternative WGP RR alternatives for that purpose as well.

We consider both RR and RN WGP models in this paper since the modeling issues are almost identical, albeit the circumstances are rather different. RN models are intended for those situations where there is typically one measure per animal and environmental effects and animal BV might be characterized by linear functions of these covariates, (DE JONG 1995). RR models are intended for longitudinal data collection in the sense that there are repeated measurements for each animal over time, such as back fat thickness in pigs as analyzed in this paper. Because of this repeated measures dynamic,

it becomes even more imperative to model animal effects (i.e. additive genetics and/or permanent environmental effects) in RR genomic models relative to RN models.

Alternative strategies for modeling GxE might be pursued based on, for example, factor analysis, when environments cannot be classified by linear functions of covariates (Burgueno, de los Campos et al. 2012).

The simplest RN/RR specification that we considered was IW-BayesC, being essentially identical to a classical mixed model specification. Unlike any of the other four RR/RN specifications, IW-BayesC assumes all pairs of SNP-specific intercept and slope effects to be normally distributed whereas each of the other methods specify either t -distributed or null effects. Furthermore, being simpler, the only hyperparameters requiring inference in IW-BayesC are variance components to estimate; this is not trivial given the difficulty for inferring upon degrees of freedom, for example, in heavier tailed specifications (Habier, Fernando et al. 2011). In fact, IW-BayesC is identical in principle to a classical RR or RN approach for mixed effects modeling. That is, we can set up the mixed model equations from the RR/RN-WGP model [2] as follows:

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{DZ} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{I}g^{11} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{DZ} + \mathbf{I}g^{12} \\ (\mathbf{DZ})'\mathbf{R}^{-1}\mathbf{X} & (\mathbf{DZ})'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{I}g^{12} & (\mathbf{DZ})'\mathbf{R}^{-1}\mathbf{DZ} + \mathbf{I}g^{22} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{g}}_1 \\ \hat{\mathbf{g}}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \\ (\mathbf{DZ})'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix} \quad [11]$$

where $\begin{bmatrix} g^{11} & g^{12} \\ g^{12} & g^{22} \end{bmatrix} = \boldsymbol{\Sigma}_g^{-1}$ whereby one could readily base σ_e^2 and $\boldsymbol{\Sigma}_g$ on their REML

estimates. Of course, it might also be necessary to further specify polygenic effects and/or permanent environmental effects in [11], particularly for the RR case. In the strictest sense, this computational approach is not exactly equivalent to IW-BayesC,

which is a fully Bayes procedure that would take into account uncertainty in σ_e^2 and Σ_g . Nevertheless, based on experiences with our simulation study (not reported), inferences on variance components, SNP effects, and, hence, BV are effectively the same, provided that relatively diffuse priors on σ_e^2 and Σ_g are specified in IW-BayesC. Furthermore, this mixed model/REML approach is computationally far more efficient compared to having to conduct MCMC on IW-BayesC.

This efficiency can be further enhanced when the number of markers well exceeds the number of animals for which genomic BV are estimated. That is, one can design a set of mixed model equations equivalent to [11] but with potentially much smaller dimensionality by directly solving for genomic intercept and slope BV in a genomic animal model rather explicitly modeling SNP-specific effects (Habier, Fernando et al. 2007; Strandén and Garrick 2009). Extending Strandén and Garrick (2009) further and assuming, say, one record per genotyped animal in a RN-like situation we can reparameterize Equation [11] as follows

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}' & \mathbf{X}'\mathbf{D} \\ \mathbf{X} & \mathbf{I} + (\mathbf{Z}\mathbf{Z}')^{-1} \sigma_e^2 g^{11} & \mathbf{D} + (\mathbf{Z}\mathbf{Z}')^{-1} \sigma_e^2 g^{12} \\ \mathbf{D}'\mathbf{X} & \mathbf{D} + (\mathbf{Z}\mathbf{Z}')^{-1} \sigma_e^2 g^{11} & \mathbf{D} + (\mathbf{Z}\mathbf{Z}')^{-1} \sigma_e^2 g^{22} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}}_1 \\ \hat{\mathbf{u}}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{y} \\ \mathbf{D}'\mathbf{y} \end{bmatrix} \quad [12]$$

where $\mathbf{u}_1 = \mathbf{Z}\mathbf{g}_1$ and $\mathbf{u}_2 = \mathbf{Z}\mathbf{g}_2$. Extensions by Misztal et al. (Misztal, Legarra et al. 2009) to RR and RN models where records from genotyped and ungenotyped animals are combined into the analysis would be relatively straightforward as well.

One objective of this paper was to develop and compare five alternative RR/RN models against each other and two conventional WGP models BayesA and BayesB. In the simulation study, we first investigated the effect of an “average” genetic correlation

($\bar{\rho}_{g_1g_2}$) between intercept and slope. In Scenarios 1-3 of the simulation study, each mimicking a CD BayesA-like or even IW-BayesA-like process (because $v_1 = v_2$), we considered three different levels of $\bar{\rho}_{g_1g_2}$ representing low, median and high positive genetic correlations between traits. We compared accuracy of genomic prediction in environments characterized by low, average and high values. We found significantly higher accuracies for the five RR/RN methods compared to the two conventional WGP methods at the extreme environments ($d=-1.2$ and $d=1.2$) at all specifications of $\bar{\rho}_{g_1g_2}$. We found the difference in accuracies between the RN versus the conventional models became greater with increasing $\bar{\rho}_{g_1g_2}$ at $d = -1.2$ whereas, curiously, we found the converse trends at $d = 1.2$. This result might be due to the fact that the intercept is defined at $d = 0$. Hence, positive genetic correlations between intercept and slope would build positive associations between genomic evaluations at $d = 0$ with those genomic evaluations based on $d \geq 0$, but negative associations between genomic evaluations at $d = 0$ and those genomic evaluations based on d appreciably less than 0; i.e., where SNP or animal-specific reaction norms start to “crisscross” each other. In other words, if we had specified $\bar{\rho}_{g_1g_2}$ to be negative, we would have found the opposite trends.

We found that IW-BayesA tended to have significantly greater accuracies in Scenarios 1-3 than all other RN models at $d=1.2$ and $d=-1.2$. This was initially surprising to us; however, we would note two considerations. Firstly, the degrees of freedom specification were the same for QTL effects for both intercept and in slope, i.e., $v_1 = v_2$. Hence, CD-BayesA might confer little or no advantage to IW-BayesA then because of the greater parsimony of a single degrees of freedom specification of the latter. Secondly,

with data simulated based on LD, one might anticipate that the assumption of independence between effects of adjacent SNP markers is somewhat distorted (YANG and TEMPELMAN 2012), although independence is specified for every model in this study. Subsequently it may be rather difficult to predict the relative performance of WGP models under LD, since the true model cannot be specified parametrically under LD, even when QTL effects are generated from known distributions. In fact, we conducted a separate simulation study (results not shown) whereby the data generation strategy exactly match the assumptions of either of the two models, CD-BayesA and IW-BayesA; but based on an assumption of linkage equilibrium between SNP markers. In those comparisons, CD-BayesA did outperform IW-BayesA when the data generation model was based on a CD-BayesA model and vice versa. Similar conclusions have been recently drawn by Wimmer et al. (2013) who found that the presence of high levels of LD, high levels of complexity ($(M_{both} + M_{int})/m$) of genetic architecture, and low levels of determinedness (n/m) will tend to mute differences in performance between various Bayesian alphabet models.

At an average level of performance ($d=0$), there appeared to be very little difference between any of the models, including the two conventional models not based on any RN specification whatsoever. This was not surprising since one would expect that conventional WGP models would, by default, predict to an average environment.

Scenarios 4-6 in the simulation study were intended to mimic a CD-BayesB like process whereby only a fraction of the QTL effects having general performance effects (i.e. intercept) also showed environmental sensitivity (i.e., non-zero slope effects). Scenario 4 (heritability = 0.5) and Scenario 5 (heritability = 0.2) were the same in that

half of the 100 QTL were environmentally sensitive. As expected, all RN models had lower accuracy in Scenario 5 compared to Scenario 4 because of the lower heritability, although somewhat surprisingly IW-BayesA generally maintained its advantage over all models.

Realizing that the total number of QTL have also been known to influence prediction accuracy comparisons between conventional BayesA and BayesB (MEUWISSEN and GODDARD 2010), we considered Scenario 6 which involved a total of 40 QTL, $M_{int} = 20$ environmentally robust QTL and $M_{both} = 20$ environmentally sensitive QTL, 2/5 of what was specified in Scenario 4 with everything else being the same. As anticipated, CD-BayesB finally started to emerge as the most accurate of the 4 RN methods particularly at $d = 0$ and $d = 1.2$. Again, these results are in agreement with Wimmer et al (2013) who determined that variable selection methods (like BayesB or CD-BayesB here), perform best under genetic architectures with low complexities. Conversely, it was this scenario where the performance of IW-BayesC started to plummet, even being inferior to conventional BayesA/BayesB analyses at $d = 0$. Of course, we should be quick to note that QTL effects were simulated from heavy-tailed t -distributions, perhaps thereby stacking the odds against IW-BayesC.

A particularly odd result was that the comparisons on accuracies did not necessarily match up with estimated accuracies of predicting their components, i.e. intercept and slope BV. That is, CD-BayesB was among the best for inferring upon intercept BV in Figure 4.2A, whereas IW-BayesA was typically among the best for inferring upon slope BV. Nevertheless, this does help explain why IW-BayesA was often

the best for inferring upon genomic BV at $d = -1.2$ and $d = 1.2$, whereas CD-BayesB generally dominated at $d = 0$.

Our simulation study was based on arbitrary specifications of genetic architecture based on $(M_{both} + M_{int})$ QTL randomly located on a 1M chromosome; based on arguments provided by Meuwissen and Goddard (2010), one might readily extrapolate these simulation results to the case of $n_{chr} * (M_{both} + M_{int})$ QTL for a n_{chr} M genome based on n_{chr} chromosomes. We realize that other determinants such as marker density can also influence the comparisons among the five RN WGP models. Furthermore, if we had specified even greater $M_{both} + M_{int}$ QTL, the more complex genetic architecture might then more likely reflect the IW-BayesC assumptions ($\pi_g = 1, \nu_g \rightarrow \infty$ such that $\Sigma_j = \Sigma_g \forall j$).

Given our simulation results then, it perhaps was not too surprising that we did not observe any meaningful differences between the various models with an application to data from a pig resource population. Firstly, the model applied was a RR, rather than a RN model, implying that there is greater phenotypic information provided by the repeated records in a RR context thereby potentially muting any real differences between the various candidate models. Furthermore, the genetic architecture of the trait analyzed was presumably far more complex and the level of determinedness far less than anything considered in the simulation study.

Based on an analysis using microsatellite markers in the MSU pig resource population, Choi et al. (2010) found highly significant QTL for back fat thickness at week of age 10, 13, 16, 19 and 22 on chromosome 6 using a QTL mapping approach without considering G×E. Our results based on a RR WGP analysis using a low density

SNP chip also indicated potential QTL on chromosome 6. Nevertheless, the RR WGP specifications allowed us to explicitly model these potential QTL effects as a function of age.

4.5 Conclusions

Five RR/RN methods have been developed in this paper under the frame work of WGP. Based on a RN simulation study and a RR data analysis in pigs, RR/RN WGP models provide greater accuracies in genomic evaluations compared to more conventional WGP models. We believe that it's important to account for SNP specific intercept and slope effects in RN or RR data situations where SNP genotypes are available. Nevertheless, differences in predictive performance between the various RR/RN WGP models were not overwhelming such that simpler specifications such as IW-BayesA may be suitable for analyses that involve high degrees of genetic complexity or low levels of determinedness as previously mentioned by Wimmer et al (2013). Conversely, based on our simulation results, we anticipate that CD-BayesB might show greater promise when marker density is large relative to the number of QTL; i.e., low degree of complexity. It is important that efficient software and/or algorithms be developed for these models in order to allow for meaningful comparisons in these situations.

Chapter 5 Exploring alternative specifications for bivariate trait whole genome prediction models

5.1 Introduction

With the advent of genotyping and sequencing technologies, whole genome prediction (WGP) has become commonly used for genetically selecting animals and plants for economically important traits (de los Campos, Hickey et al. 2013). Numerous approaches, including non-parametric methods (Gianola, Wu et al. 2010), Bayesian parametric “alphabet” methods (Meuwissen, Hayes et al. 2001; Gianola, de los Campos et al. 2009; Habier, Fernando et al. 2011), and generalized expectation-maximization methods (KARKKAINEN and SILLANPAA 2012) for single trait analyses have been developed. There may be, however, other untapped opportunities to improve prediction accuracy in WGP. It is well known, for example, that many economically important traits are genetically correlated. Multiple trait analyses have been recently used to account for correlations among traits due to specific genes in genome wide association studies (ZHU and ZHANG 2009), including for differential mapping of pleiotropic versus non-pleiotropic QTLs (Banerjee, Yandell et al. 2008).

A large number of genetic evaluation methods have been developed and applied to jointly analyze correlated traits in livestock (Gianola and Sorensen 2004; Banerjee, Yandell et al. 2008). Some of these methods involve independent analyses on sets of transformed variables using techniques based on, for example, factor analysis, principal component analysis, canonical analysis and cluster analysis (Weller, Wiggans et al. 1996; Musani, Zhang et al. 2006; de los Campos and Gianola 2007; Vichi and Saporita 2009).

For quantitative genetic analysis, these methods generally require a two-step approach of reducing either number of traits and/or number of genetic effects. Another approach to multiple trait modeling is to model the linear regression relationships among traits in a multilayer system, namely structural equation models (SEM) or path coefficient models (GIANOLA and SORENSEN 2004) although this might seem rather complex with large numbers of SNPs. Banerjee et al. (2008) used seeming unrelated regression (SUR) to identify pleiotropic QTL for multiple traits. This method allows each trait to have a separate set of QTL or trait specific QTL and facilitates a computational efficient sampling algorithm. However, their method models trait correlations due to residuals rather than due to QTL, thereby providing no information on genetic correlation between traits.

In order to further improve WGP accuracy, several efforts have been made to develop Bayesian approaches in multiple trait models (CALUS and VEERKAMP 2011; JIA and JANNINK 2012). Calus and Veerkamp (2011) demonstrated that, for traits having a high genetic correlation with each other, multiple trait WGP model analyses lead to higher WGP accuracies compared to single trait analyses, particularly for the lower heritability trait. Among multiple trait WGP models investigated by Calus and Veerkamp (2011), BayesSSVS, a variable selection method with a spike and slab prior on the SNP effects, outperformed other models that assumed a normal density on all SNP. Jia and Jannink (2012) further confirmed that the advantage of multiple trait WGP models over single trait counterparts was greatly influenced by several factors, i.e. heritabilities and genetic correlations between traits as well as the number of QTLs.

BayesA and BayesB are popular “Bayesian alphabet” models used for single trait WGP. BayesA specifies a scaled- t prior on SNP effects and is a special case of BayesB which, similar to BayesSSVS, specifies a mixture prior of point mass at zero and scaled- t density (Meuwissen, Hayes et al. 2001). These specifications often have higher WGP prediction accuracies compared to procedures based on Gaussian distribution assumptions (e.g. ridge regression or GBLUP) if SNP effects deviate substantially from normality. Calus and Veerkamp (2011) recently extended BayesA for use in multiple trait analyses and determined similar advantages over multiple trait GBLUP predictions.

By estimating SNP-specific pleiotropic effects for multiple traits, we may infer upon the most important pleiotropic regions in the genome given the relative locations of SNP markers (van Binsbergen, Veerkamp et al. 2012). However, the assumption of a multiple trait BayesA model, derived from conjugate inverted Wishart (IW) prior densities on the SNP specific variance-covariance matrices (VCV), might be potentially inflexible since the uncertainty in all elements of a VCV is based on a single degrees of freedom parameter (MUNILLA and CANTET 2012). An alternative parameterization on VCV for random and residual effects was proposed by Bello et al. (2010) who suggested that the square root free Cholesky decomposition (CD) of the VCV in bivariate mixed models might allow greater flexibility as uncertainty can be differentially expressed on each element of a VCV using such a parameterization.

In this study, our objectives were: 1) To reaffirm the greater accuracies of prediction provided by bivariate trait models relative to single trait conventional WGP approaches and 2) To assess whether there may be greater flexibility, and hence greater

WGP accuracy, using CD-based parameterizations compared to IW based specifications on SNP-specific variance-covariance matrices.

5.2 Methods and Materials

5.2.1 Whole genome prediction models

In a bivariate trait WGP model, SNP substitution effects are estimated for two traits simultaneously. The general bivariate trait WGP model can be denoted by the following:

$$y_{ij} = \mathbf{x}'_{ij} \boldsymbol{\beta}_j + \mathbf{z}'_i \mathbf{g}_j + e_{ij}, \quad [1]$$

where y_{ij} is the phenotype record for i^{th} animal on j^{th} trait ($i = 1, 2, \dots, n; j = 1, 2$); $\boldsymbol{\beta}_j$ is the vector of fixed effects on trait j , \mathbf{x}'_{ij} is the incidence row vector connecting elements of $\boldsymbol{\beta}_j$ to animal i ; $\mathbf{z}'_i = [z_{i1} \ z_{i2} \ z_{i3} \ \dots \ z_{im}]$ is the vector of genotypes coded as 0, 1, or 2 copies of the minor allele on SNPs for animal i ; $\mathbf{g}_j = \{g_{jk}\}_{k=1}^m$ is the vector of SNP substitution effects on trait j ; e_{ij} is the random residual for i^{th} animal on j^{th} trait. We can rewrite Equation [1] using matrix notation as:

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{Z} & \mathbf{0} \\ \mathbf{0} & \mathbf{Z} \end{bmatrix} \begin{bmatrix} \mathbf{g}_1 \\ \mathbf{g}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix}, \quad [2]$$

where $\mathbf{y}_i = \{y_{ij}\}_{i=1}^n$, $\mathbf{X}_i = \{\mathbf{x}'_{ij}\}_{i=1}^n$, and $\mathbf{Z} = \{\mathbf{z}'_i\}_{i=1}^n$. The animals' genomic merit for the two traits can be subsequently represented as $\mathbf{u}_1 = \mathbf{Z}\mathbf{g}_1$ and $\mathbf{u}_2 = \mathbf{Z}\mathbf{g}_2$, respectively. For the various bivariate trait WGP models investigated, we assumed that pairs of residuals on

animal i ; i.e. $\mathbf{e}_i = [e_{i1} \ e_{i2}]'$, $i=1,2,\dots,n$, follow independent bivariate normal densities

with a vector null mean and a common variance covariance matrix $\Sigma_e = \begin{bmatrix} \sigma_{e_1}^2 & \sigma_{e_1 e_2} \\ \sigma_{e_1 e_2} & \sigma_{e_2}^2 \end{bmatrix}$.

Similarly, effects of SNP k on the two traits follow independent bivariate normal densities with a vector null mean and a common variance covariance matrix

$\Sigma_g = \begin{bmatrix} \sigma_{g_1}^2 & \sigma_{g_1 g_2} \\ \sigma_{g_1 g_2} & \sigma_{g_2}^2 \end{bmatrix}$. Diffuse proper Gaussian or flat priors are typically specified on β_1

and β_2 (SORENSEN and GIANOLA 2002). For the residual variance covariance matrix Σ_e , we might typically specify a conjugate inverted Wishart prior with degrees of freedom ν_0 and scale matrix Σ_0 .

5.2.2 Univariate BayesA and BayesB (uBayesA\|uBayesB)

We re-label conventional single trait BayesA and BayesB models as uBayesA and uBayesB, respectively to indicate the emphasis on univariate analysis. We infer upon key hyperparameters in this model using prior specifications and strategies previously outlined by Yang and Tempelman (2012).

5.2.3 Bivariate Ridge regression (bGBLUP)

Outside of some strategies for rescaling, we specify the realized relationship matrix based on the unscaled genotype matrix for SNPs derived as $\mathbf{G}=\mathbf{Z}\mathbf{Z}'$ in a bivariate mixed effects model that we label as bivariate genomic BLUP or bGBLUP. In the bGBLUP model, we specified multivariate normal distributions having null means for each of $\mathbf{u}=[\mathbf{u}_1' \ \mathbf{u}_2']'$ and $\mathbf{e}=[\mathbf{e}_1' \ \mathbf{e}_2']'$ such that $\text{var}(\mathbf{u})=((\mathbf{Z}\mathbf{Z}') \otimes \Sigma_g)$ and

$\text{var}(\mathbf{e}) = (\mathbf{I}_{2 \times 2} \otimes \Sigma_e)$. Based on these specifications we use ASReml 3.0 (Gilmour, Gogel et al. 2009) to provide REML estimates of Σ_g and Σ_e in order to compute the BLUP $\hat{\mathbf{u}}$ of \mathbf{u} and hence $\hat{\mathbf{g}}$ of \mathbf{g} as necessary.

5.2.4 Bivariate Student-*t* (IWBayesA)

A convenient and previously used extension for bivariate trait WGP model is to apply a conjugate inverted Wishart prior on heterogeneous SNP-specific variance-covariance matrices. This specification represents a multivariate extension of BayesA (CALUS and VEERKAMP 2011) that we label as IWBayesA. For the joint effects of SNP k on the two traits, we can specify a bivariate normal density conditionally as follows:

$$\mathbf{g}_{\cdot k} \sim \mathbf{N} \left(\mathbf{0}_{2 \times 1}, \mathbf{G}_k = \begin{bmatrix} \sigma_{g_{1k}}^2 & \sigma_{g_{1k}g_{2k}} \\ \sigma_{g_{1k}g_{2k}} & \sigma_{g_{2k}}^2 \end{bmatrix} \right) \quad [4]$$

where \mathbf{G}_k is the SNP-specific variance-covariance matrix for the two traits and is regarded as a random draw from a conjugate Inverted Wishart prior with degrees of

freedom v_g and scale matrix $\Sigma_g = \begin{bmatrix} \sigma_{g_1}^2 & \sigma_{g_1g_2} \\ \sigma_{g_1g_2} & \sigma_{g_2}^2 \end{bmatrix}$. For a fully hierarchical Bayesian model

as developed in Yang and Tempelman (2012), we inferred upon hyperparameters after specifying a prior $p(v_g)$ on v_g , and also a conjugate Wishart prior $\Sigma_g \sim W(v_0, \mathbf{S}_0)$ on Σ_g .

Note that the uncertainty of IWBayesA is only controlled by one scalar v_g (MUNILLA and CANTET 2012). Furthermore, IWBayesA assumes that every single SNP is pleiotropic. Conceptually, IWBayesA is not much different from that developed (with the same label) for reaction norm modeling in Chapter 4.

5.2.5 Cholesky decomposition specifications (CDBayesA\CDBayesB)

In order to address the potential inflexibility in IWBayesA, we developed an alternative approach based on the square root free Cholesky decomposition (CD) of \mathbf{G}_k . We have previously shown that the CD parameterization can provide greater flexibility for modeling variance-covariance matrices (Bello, Steibel et al. 2010). Based on a particular order for two traits, we can write the SNP effects on the second trait \mathbf{g}_2 as a linear regression of SNP effects on the first trait \mathbf{g}_1 :

$$\mathbf{g}_2 = \mathbf{\Psi}\mathbf{g}_1 + \mathbf{g}_{2|1} \quad [5]$$

Hence we can re-write the general bivariate trait WGP model [2] as:

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{Z} & \mathbf{0} \\ \mathbf{0} & \mathbf{Z} \end{bmatrix} \begin{bmatrix} \mathbf{g}_1 \\ \mathbf{\Psi}\mathbf{g}_1 + \mathbf{g}_{2|1} \end{bmatrix} + \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix} \quad [6]$$

where $\mathbf{g}_{2|1} = \{g_{2|1,k}\}_{k=1}^m$ is the vector of SNP effects on the second trait conditional on the first trait, and $\mathbf{\Psi} = \text{diag}\{\phi_k\}_{k=1}^m$ is a diagonal matrix of SNP-specific *association effects* between SNP effects on the two traits.

Suppose we specify $g_{1k} \sim N(0, \sigma_{g_{1k}}^2) \forall k$ with $\sigma_{g_{1k}}^2 \sim \chi^{-2}(v_1, v_1 s_1^2) \forall k$ for SNP effects and their respective variances on Trait 1. Similarly, we specify $g_{2|1,k} \sim N(0, \sigma_{g_{2|1,k}}^2) \forall k$ with $\sigma_{g_{2|1,k}}^2 \sim \chi^{-2}(v_{2|1}, v_{2|1} s_{2|1}^2) \forall k$ for SNP effects and variances on Trait 2 conditional on Trait 1. We label this model as CDBayesA given its CD-based multivariate extension of BayesA; conceptually, it is very similar to the model of the same name used in Chapter 4 for reaction norm modeling. For the SNP specific

association effects between the two traits, we specify independent normal priors $\phi_k \sim N(\mu_\phi, \sigma_\phi^2) \forall k$. Key hyperparameters, namely the degrees of freedom ($\nu_1, \nu_{2||}$) and scale ($s_1^2, s_{2||}^2$) parameters, can be inferred upon in CDBayesA using prior specifications similar to those in the conventional uBayesA model.

Here, the SNP-specific variance-covariance matrices \mathbf{G}_k are still specified very generally by three parameters, as with IWBayesA, but expressed by an alternative parameterization; i.e.

$$\mathbf{G}_k = \begin{bmatrix} \sigma_{g_{1k}}^2 & \sigma_{g_{1k}}^2 \phi_k \\ \sigma_{g_{1k}}^2 \phi_k & \sigma_{g_{1k}}^2 \phi_k^2 + \sigma_{g_{2||k}}^2 \end{bmatrix}$$

However, unlike IWBayesA whereby the uncertainty on \mathbf{G}_k is essentially controlled by one degrees of freedom parameter, the uncertainty on \mathbf{G}_k in CDBayesA is controlled by three such parameters: the two different degrees of freedom terms ν_1 and $\nu_{2||}$ as well as the variance component σ_ϕ^2 . Nevertheless, CDBayesA, as does IWBayesA, assumes that every SNP has a pleiotropic effect.

In an attempt to provide even greater flexibility than CDBayesA; i.e., to allow not only pleiotropic effects but also non-pleiotropic effects and/or potentially null effects for each SNP on both traits, we developed a variable selection approach analogous to a BayesB type of specification which we naturally label as CDBayesB. Let $\sigma_{g_{1k}}^2$ have a mixture prior of point mass at zero with probability $(1 - \pi_1)$ or randomly drawn from

$\chi^{-2}(v_1, v_1 s_1^2)$ with probability π_1 for $k=1,2,\dots,m$. Similarly, $\sigma_{g_{2|k}}^2$ has a mixture prior of point mass at zero with probability $(1-\pi_{2|})$ or randomly drawn from $\chi^{-2}(v_{2|}, v_{2|} s_{2|}^2)$ with probability $\pi_{2|}$ for $k=1,2,\dots,m$. For the association effects, ϕ_k between the two traits at each SNP $k=1,2,\dots,m$, we specify a mixture prior of point mass at zero with non-association probability $(1-\pi_\phi)$ or randomly drawn from $N(\mu_\phi, \sigma_\phi^2)$ with association probability π_ϕ . Hence, in CDBayesB, we could infer upon SNP effects that are non-zero and trait-specific (i.e., non-pleiotropic). For Trait 1, this would occur when $g_{1k} \neq 0$ with $\phi_k = 0$ and $g_{2|k} = 0$ whereas for Trait 2 this would entail $g_{1k} = 0$ and $g_{2|k} \neq 0$ regardless of the value of ϕ_k . Pleiotropic effects will be inferred if $g_{1k} \neq 0$ and $\phi_k \neq 0$ regardless of the value of $g_{2|k}$ although a value of $g_{2|k} = 0$ would then imply a situation of “perfect” pleiotropy between the two traits (i.e., a SNP-specific genetic correlation equal to ± 1). Using prior specifications similar to those in CDBayesA, we could infer upon the key hyperparameters, i.e. degrees of freedom $(v_1, v_{2|})$ and scale parameters $(s_1^2, s_{2|}^2)$. Furthermore, we could specify informative or diffuse priors $p(\pi_1)$, $p(\pi_{2|})$, $p(\pi_\phi)$, $p(\mu_\phi)$, and $p(\sigma_\phi^2)$ on π_1 , $\pi_{2|}$, π_ϕ , μ_ϕ and σ_ϕ^2 , respectively. This model is somewhat analogous to the same named model for reaction norms in Chapter 4 except that here we specify three mixture rather than two mixture distributions to provide even greater flexibility.

5.2.6 Bayesian inference

For fully Bayesian inference using Markov Chain Monte Carlo (MCMC) methods, we require strategies for drawing random draws from the full conditional densities (FCD) of each unknown parameter (or blocks thereof) under all models. FCD for uBayesA\ uBayesB have been illustrated in our previous work (YANG and TEMPELMAN 2012). In this paper, we present FCD for three bivariate trait WGP models (IWBayesA, CDBayesA, CDBayesB) in Appendix D1. The fourth bivariate WGP model was analyzed using classical REML and BLUP for computational expedience although we determined that these inferences weren't practically different from MCMC based inferences (results not shown).

5.2.7 Simulation studies

We designed a naive small-scale simulation study involving independent markers (i.e., not in LD) using a response surface design (Table D2.1 in Appendix D2) based on five factors that we thought might be particularly important for influencing the WGP accuracies on the trait with the lower heritability between two rather different models: IWBayesA versus CD-BayesB. These five factors included the number (n) of animals, the number (M_1) of QTL controlling Trait 1 ($h^2 = 0.8$), the number (M_2) of QTL controlling Trait 2 ($h^2 = 0.1$), the number (M_{12}) of QTL pleiotropically controlling both traits and the variability ($\sigma_{QTL,\phi}^2$) of the associations between the two traits across QTL. Among all five factors considered, only M_{12} had a significant interaction with model; i.e., the difference between IWBayesA and CDBayesB for WGP accuracy on Trait 2 depended on M_{12} ($P < 0.0001$) as further noted in Table D2.2 of Appendix D2. We used

this knowledge to design a more focused LD simulation study to compare the WGP accuracies for bivariate trait analyses involving six different models (uBayesA, uBayesB, bGBLUP, IWBayesA, CDBayesA, CDBayesB). Two populations were targeted, differing only in the number of pleiotropic QTL influencing both traits ($M_{12}=10$ versus $M_{12}=30$), with all the other specifications being the same as indicated in Table 5.1.

Table 5.1: Summary of two different populations compared in a LD simulation study.

| Factors | Population 1 | Population 2 |
|--|--------------|--------------|
| Constant | | |
| Heritability of Trait 1 | 0.5 | 0.5 |
| Heritability of Trait 2 | 0.1 | 0.1 |
| Residual covariance between two traits | 0 | 0 |
| Number of SNPs | 2000 | 2000 |
| Number of animals | 500 | 500 |
| Mean on association parameters (μ_ϕ) | 0.8 | 0.8 |
| Variance on association parameters (σ_ϕ^2) | 0.05 | 0.05 |
| Number of QTLs for Trait 1(M_1) | 10 | 10 |
| Number of QTLs for Trait 2(M_2) | 10 | 10 |
| Of Interest | | |
| Number of QTLs for both traits (M_{12}) | 10 | 30 |

For each of the two populations or scenarios, we generated 20 replicates based on a constant population size of 100 for 6000 generations of random mating using the `hypr` package in R (TECHNOW 2012). For each replicate, we defined the genome as composed of one chromosome of length 1 Morgan having 20,000 SNP loci; i.e., the number of recombinations for each meiosis event was drawn from a Poisson(1) distribution with crossing over locations drawn from a uniform distribution. In the base population, all loci were monomorphic with polymorphisms created by a recurrent mutation rate of 2.5×10^{-4} per locus per generation for each of the first 6000 generations. After 6000 generations, two additional generations (6001 and 6002) were created with expanded population sizes of 500 animals each. In Generation 6001, we excluded SNP with a minor allele frequency (MAF) < 0.1 . We defined genotype dosages (i.e., genotype matrix) in Generation 6001 and 6002 as counts (0, 1, 2) of the minor allele for all the remaining SNPs. We randomly selected 2000 SNPs plus an additional $M_1 + M_2 + M_{12}$ SNPs to be QTLs in Generation 6001.

Now $M_1=10$, $M_2=10$, $M_{12}=10$ in Population 1 whereas $M_1=10$, $M_2=10$, $M_{12}=30$ in Population 2. We generated QTL effects $\{g_{QTL,1j}\}_{j=1}^{j=M_1+M_{12}}$ for Trait 1 from a reflected gamma distribution with shape=0.4 and scale=2.24. QTL effects $\{g_{QTL,2|j}\}_{j=M_1+1}^{j=M_1+M_{12}+M_2}$ for Trait 2, conditional on Trait 1 were also generated from a reflected gamma distribution with shape=0.4 and scale = 1.34. The association variables between Trait 1 and Trait 2 (i.e. $\{\phi_{QTL,j}\}_{j=M_1+1}^{j=M_1+M_{12}}$) were simulated from a $N(\mu_{QTL,\phi}=0.8, \sigma_{QTL,\phi}^2=0.05)$. The effect for QTL j on Trait 2 was thereby determined as $g_{QTL,2j} = g_{QTL,1j} + \phi_{QTL,j}g_{QTL,2|j}; j=1, 2, \dots, M_1 +$

$M_{12} + M_2$ noting that $\{\phi_{QTL,j}\}_{j=1}^{j=M_2} = 0$, $\{\phi_{QTL,j}\}_{j=M_1+M_{12}+1}^{j=M_1+M_{12}+M_2} = 0$, $\{g_{QTL,1j}\}_{j=M_1+M_{12}+1}^{j=M_1+M_{12}+M_2} = 0$,

and $\{g_{QTL,21j}\}_{j=1}^{j=M_1} = 0$. Hence pleiotropic QTL effects were generated from a complex

bivariate distribution. If we define $\mathbf{Z}_{QTL,1}$ as the subset of the $n \times M_1$ SNP genotypes in \mathbf{Z}

that are designated to be QTL for Trait 1 only and $\mathbf{Z}_{QTL,12}$ as the subset of the $n \times M_{12}$

SNP genotypes in \mathbf{Z} that are designated to be QTL for both traits, the true breeding

values $\{TBV_{1i}\}_{i=1}^{i=n}$ for Trait 1 in each of Generations 6001 and 6002 were generated using

$$\{TBV_{1i}\}_{i=1}^{i=n} = \mathbf{Z}_{QTL,12} \{g_{QTL,1j}\}_{j=1}^{j=M_{12}} + \mathbf{Z}_{QTL,1} \{g_{QTL,1j}\}_{j=M_{12}+1}^{j=M_{12}+M_1} \quad [7a]$$

Similarly, if we define $\mathbf{Z}_{QTL,2}$ as the subset of the $n \times M_2$ SNP genotypes in \mathbf{Z} that

are designated to be QTL for Trait 2 only, the true breeding values $\{TBV_{2i}\}_{i=1}^{i=n}$ for Trait 2

in each of Generations 6001 and 6002 were generated using

$$\{TBV_{2i}\}_{i=1}^{i=n} = \mathbf{Z}_{QTL,12} \{g_{QTL,2j}\}_{j=1}^{j=M_{12}} + \mathbf{Z}_{QTL,2} \{g_{QTL,2j}\}_{j=M_{12}+1}^{j=M_{12}+M_2} \quad [7b]$$

Based on heritabilities $h_1^2 = 0.5$ and $h_2^2 = 0.1$ for Trait 1 and Trait 2, respectively,

we generated the pair of residuals for each of the two traits on each animal from a

bivariate normal distribution with null mean and variance covariance matrix Σ_e , i.e.

$$\Sigma_e = \begin{bmatrix} \text{var}\left(\{TBV_{1i}\}_{i=1}^{i=n}\right)(1-h_1^2)/h_1^2 & 0 \\ 0 & \text{var}\left(\{TBV_{2i}\}_{i=1}^{i=n}\right)(1-h_2^2)/h_2^2 \end{bmatrix} \quad [8]$$

In other words, residuals were specified to be uncorrelated as it has been determined

previously by Jia and Jannink (2012) that the nature of residual correlation between two

traits is inconsequential to the accuracy of WGP prediction in bivariate models. Phenotypic records for two traits were then generated as $y_{1i} = TBV_{1i} + e_{1i}$ and $y_{2i} = TBV_{2i} + e_{2i}$. Prediction accuracies of breeding values for the two traits in Generation 6002 were defined as correlation between $\{TBV_{1i}\}_{i=1}^{i=n}$ and $\{EBV_{1i}\}_{i=1}^{i=n}$ for Trait 1, and correlation between $\{TBV_{2i}\}_{i=1}^{i=n}$ and $\{EBV_{2i}\}_{i=1}^{i=n}$ for Trait 2. The factor M_{12} , of interest, here influences the overall genetic correlation ($\bar{\rho}_{g_1g_2}$) between the two traits, which we determined as the correlation between $\{TBV_{1i}\}_{i=1}^{i=n}$ and $\{TBV_{2i}\}_{i=1}^{i=n}$ in Generation 6001.

5.2.8 Pine data analyses

Resende et al. (2012) provide a data set of loblolly pine phenotypes and genotypes for demonstration of WGP methods which has been previously used by Jia and Jannink (2012). The original data set had genotypes on 4854 SNPs and 926 individuals. After we excluded SNPs with $MAF < 0.05$ and with $P < 10^{-4}$ in HWE test, 2684 SNPs remained. Using de los Campos et al. (2013), we standardized the genotype matrix based $(z_{ij} - 2p_j) / \sqrt{2p_j(1-p_j)}$, where z_{ij} is genotype of j^{th} SNP (minor allele dosage of 0, 1, or 2) on i^{th} animal and p_j is allele frequency of one allele for j^{th} SNP. Although raw phenotypes were not publicly available, the authors provided deregressed EBVs for 17 traits. Following Jia and Jannink (2012), we fitted deregressed EBVs as response variables to compare the various WGP models. We selected two disease resistance traits, i.e. Rust gall volume (RGV) with heritability of 0.12 and presence or absence of rust (RBIN) with heritability of 0.21. After merging deregressed EBVs for the two traits and

SNP genotypes, 807 individuals with deregressed EBVs on both traits and each genotyped with 2684 SNPs remained in the final data set.

Hyperparameters for each of the six models (uBayesA/uBayesB, bGBLUP, IWBAYES, CDBAYES, CDBAYESB) were estimated. Bayesian inference was based on 600,000 MCMC iterations with a burn-in period of 50,000 cycles for uBayesA, uBayesB and IWBAYES. However, we found MCMC samples of hyperparameters using the CDBAYES and CDBAYESB models were mixing very slowly, particularly for hyperparameters like scale ($s_1^2, s_{2||}^2$) and degrees of freedom ($v_1, v_{2||}$). To alleviate this problem, we arbitrarily specified the degrees of freedom ($v_1, v_{2||}$) to be 4 for both traits in the CDBAYES model since its specification didn't influence the WGP accuracy of BV as we found in Chapter 3. We fixed the scale parameters ($s_1^2, s_{2||}^2$) in CDBAYES model unique to each trait to their corresponding REML estimates in a bBLUP model as based on the Cholesky decomposition of Σ_g . Nevertheless, we still estimated the mean and variance of the association parameters (i.e., μ_ϕ and σ_ϕ^2) in CDBAYES model using MCMC. For CDBAYESB model, we fixed the degrees of freedom ($v_1, v_{2||}$) to 5. We also fixed scale parameters ($s_1^2, s_{2||}^2$) and probabilities ($\pi_1, \pi_{2||}$) unique to each trait based on their corresponding estimates using uBayesB. Other hyperparameters (π_ϕ, μ_ϕ and σ_ϕ^2) in CDBAYESB were estimated using MCMC.

To further compare the six models by cross-validation, we randomly split the data 20 different times into a training subset with 726 individuals (90%) and a validation subset with the remaining 81 individuals (10%), thereby leading to 20 cross-validation replicates. In order to investigate the influence of the specification of trait order in

CDBayesA and CDBayesB, we first analyzed the training data by setting RBIN as Trait 1 and RGV as Trait 2, labeling the two models as CDBayesA1 and CDBayesB1. Then, we switched the order of the two traits and labeled the two models as CDBayesA2 and CDBayesB2. For the seven methods, we predicted the deregressed EBVs in the validation dataset based on posterior mean estimates of SNP effects for the two traits from the training dataset. Performance of the seven models, namely cross-validation predictive ability, was evaluated by Pearson correlation between the predicted and fitted deregressed EBVs in the validation data. In cross-validation, we expected to see more differences in predictive ability comparing the seven models on the low heritability trait RGV than on RBIN. For each of the models on trait RGV, we also assessed inferences on effects on the various SNPs to see if there might be any meaningful differences between the various models in this respect.

5.2.9 Priors used for data analyses

In the simulation study, we specified a non-informative prior $p(\nu) \propto (\nu + 1)^{-2}$ and Gelman's prior $s^2 \sim \chi^{-2}(-1, 0)$ for uBayesA as we've done previously (YANG and TEMPELMAN 2012). We specified a non-informative prior $p(\nu) \propto (\nu + 1)^{-2}$, a proper conjugate prior $s^2 \sim \text{Gamma}(0.1, 0.1)$ and $\pi \sim \text{Beta}(\alpha_\pi = 1, \beta_\pi = 8)$ in uBayesB. For IWBayesA, we specified a proper conjugate prior $p(\Sigma_g) \propto W(\nu_0, \Sigma_0)$ where ν_0 is 4 and Σ_0 is a 2 by 2 identity matrix. For CDBayesA, we specified the same non-informative prior on degree of freedom $(\nu_1, \nu_{2||})$ and Gelman's prior on scale parameters $(s_1^2, s_{2||}^2)$ as in uBayesA. Priors $p(\mu_\phi) \propto N(0, 1)$ and $p(\sigma_\phi^2) \propto \chi^{-2}(-1, 0)$ were specified on mean and

variance of association parameters in CDBayesA. For CDBayesB, we specified the same prior on degree of freedom ($\nu_1, \nu_{2||}$), scale parameters ($s_1^2, s_{2||}^2$) and probabilities ($\pi_1, \pi_{2||}$) as in uBayesB. We specified the same priors on μ_ϕ and σ_ϕ^2 as in CD-BayesA. For π_ϕ , we specified proper prior $\pi_\phi \sim \text{Beta}(\alpha_\pi = 1, \beta_\pi = 8)$.

For the analysis of the pine data, we used the same priors on hyperparameters in uBayesA and uBayesB. For IWBayesA, we specified proper conjugate prior $p(\Sigma_g) \propto W(\nu_0, \Sigma_0)$ where ν_0 is 2 and Σ_0 is a 2 by 2 diagonal matrix. The first and second diagonal elements in Σ_0 were specified to be the estimates of scale on RGV and RBIN using uBayesA, i.e. 1.323e-05 and 3.215e-05. For CDBayesA and CDBayesB, we fixed the degrees of freedom and scale parameters to estimates from the corresponding univariate analyses (uBayesA and uBayesB) because of slow mixing as previously noted. We used the same priors on μ_ϕ , σ_ϕ^2 , and π_ϕ for CDBayesA and CDBayesB as specified in the simulation study.

5.3 Results

5.3.1 Simulation Studies

In the simulation study, the overall genetic correlation ($\bar{\rho}_{g_1g_2}$) between two traits was 0.48 and 0.63, respectively, for Populations 1 and 2 over the 20 replicates per population. Population 2 had a much smaller between replicate standard deviation (~ 0.13) for $\bar{\rho}_{g_1g_2}$ compared to Population 1 (~ 0.29).

Figure 5.1 illustrates the average WGP accuracy of predicting breeding values (BV) in Generation 6002 for the two traits over the 20 replicates for each of Populations 1 (Figure 5.1A) and 2 (Figure 5.2A). For Trait 1 ($h^2 = 0.5$) in Population 1, bGBLUP had >5% lower ($P < 0.05$) average accuracy compared to IWBayesA, while IWBayesA had ~2% lower ($P < 0.05$) accuracy compared to the four models (uBayesA\uBayesB\CDBayesA\CDBayesB), including two based on univariate WGP analyses. For Trait 2 ($h^2 = 0.1$) in Population 1, bGBLUP and uBayesA had ~8% and ~3% lower ($P < 0.05$) average accuracy respectively compared to the other four models (BayesB\uBayesA\CDBayesA\CDBayesB). No significant difference was found between uBayesB and the three bivariate models (IWBayesA\CDBayesA\CDBayesB) for Trait 2 in Population 1.

For Trait 1 in Population 2 (Figure 5.1B), we found that both bGBLUP and IWBayesA had ~2% lower ($P < 0.05$) average accuracy than the other four models (uBayesA\uBayesB\CDBayesA\CDBayesB). For Trait 2 in Population 2, the three bivariate trait models (IWBayesA\CDBayesA\CDBayesB) outperformed bGBLUP (~3%), while bGBLUP had ~5% higher ($P < 0.05$) accuracy than the uBayesA and uBayesB. No significant difference was found among the three Bayesian bivariate trait models (IWBayesA\CDBayesA\CDBayesB) for Trait 2 in Population 2.

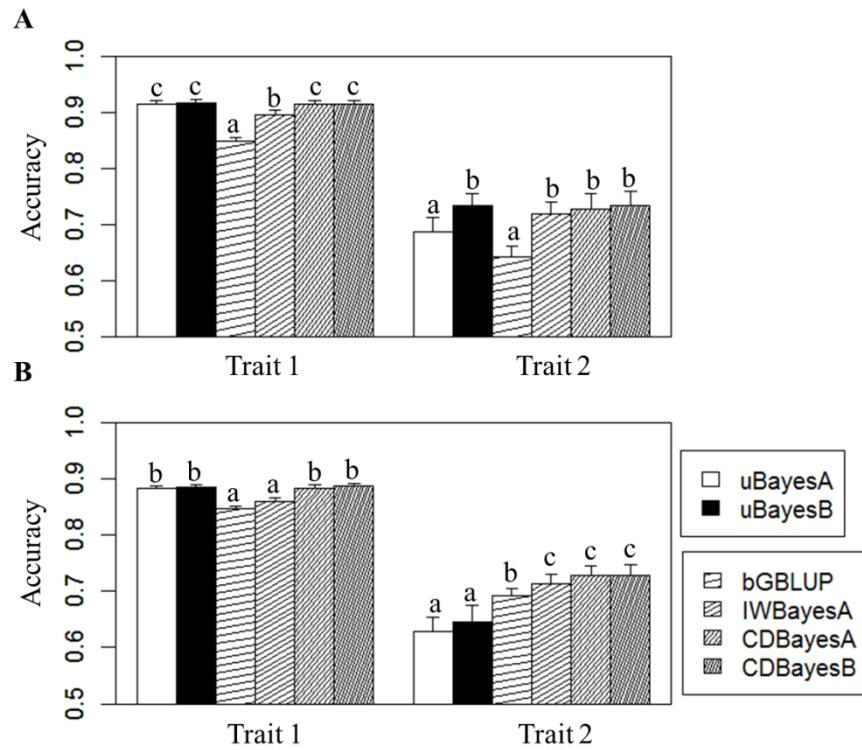


Figure 5.1: Accuracy of breeding value prediction for six methods (uBayesA\ uBayesB\ bGBLUP\ IWBAYES\ CDBAYES\ CDBAYESB) in two scenarios: A) number of QTLs for both traits = 10; B) number of QTLs for both traits = 30. Different letters indicate significant difference with $P < 0.05$.

5.3.2 Pine data analyses

The average predictive abilities for the eight different models over the 20 different replicates in the cross validation are summarized in Figure 5.2.

For the lower heritability trait RGV, we found that CDBayesB1 and CDBayesB2 had ~5% greater ($P<0.05$) predictive accuracy compared to the other six models, whereas bGBLUP, CDBayesA1 and CDBayesA2 had lower ($P<0.05$) predictive accuracies (~6% and ~9%) compared to the other three bivariate trait models. There was no evidence that IWBayesA had different predictive accuracies for RGV compared to uBayesA and uBayesB. However, four of the models (uBayesA\ uBayesB\ IWBayesA\ CDBayesA1) including the two univariate models outperformed bGBLUP (~4% - 6%) for RGV.

For the higher heritability trait RBIN, we found that bGBLUP, CDBayesA1 and CDBayesA2 had lower ($P<0.05$) predictive accuracies (~6% and ~9%) compared to the other three bivariate trait models. Furthermore, there was no evidence of a difference among uBayesA and uBayesB and three of the bivariate trait models (IWBayesA\ CDBayesB1\ CDBayesB2).

Across each of the cross validation replicates, the estimated SNP effects for either trait using either the two specifications for order of traits using CDBayesB agreed rather well (refer to panels A and B of Figure D2.1 in Appendix D2). However, in contrast with CDBayesB, there was less agreement between the two different trait orders using CDBayesA (refer to panels C and D of Figure D2.1 in Appendix D2). To further demonstrate the difference among three methods (uBayesA, IWBayesA and CDBayesB1), Figure 5.3 shows the absolute values of estimated SNP effects on RGV against SNP

index using all of the data. IWBayesA (Figure 5.3B) detected the same number of extreme large SNP effects as uBayesA (Figure 5.3A). However, compared to IWBayesA and uBayesA, there were larger SNP effects inferred in CDBayesB1 (Figure 5.3C) for RGV, which might partially explain the higher predictive accuracy for RGV in cross-validation for that particular model.

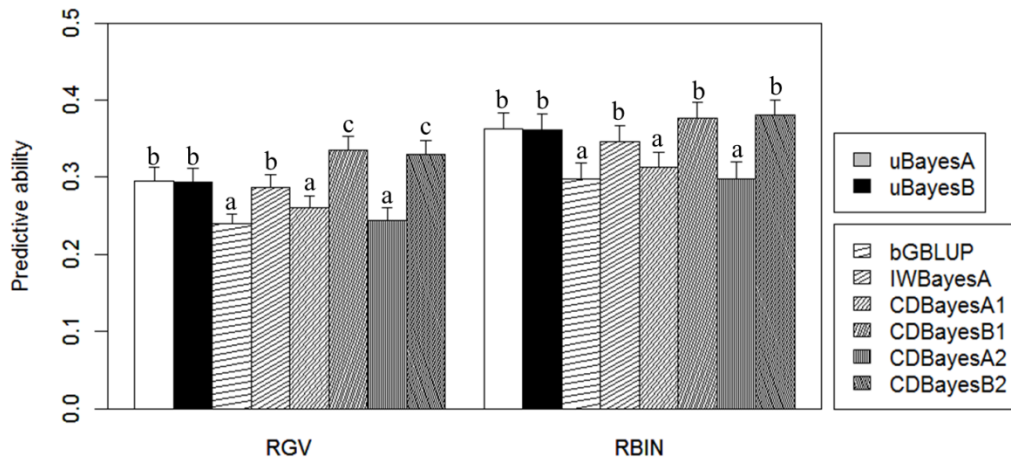


Figure 5.2: Average predictive ability from cross-validation using the loblolly pine data set for eight methods (uBayesA\ uBayesB\ bGBLUP\ IWBayesA\ CDBayesA1\ CDBayesB1\ CDBayesA2\ CDBayesB2), where CDBayesA1 and CDBayesB1 were using RBIN as the first trait and RGV as the second trait; CDBayesA2 and CDBayesB2 were using RGV as the first trait and RBIN as the second trait; Different letters indicate means are different ($P < 0.05$) from each other.

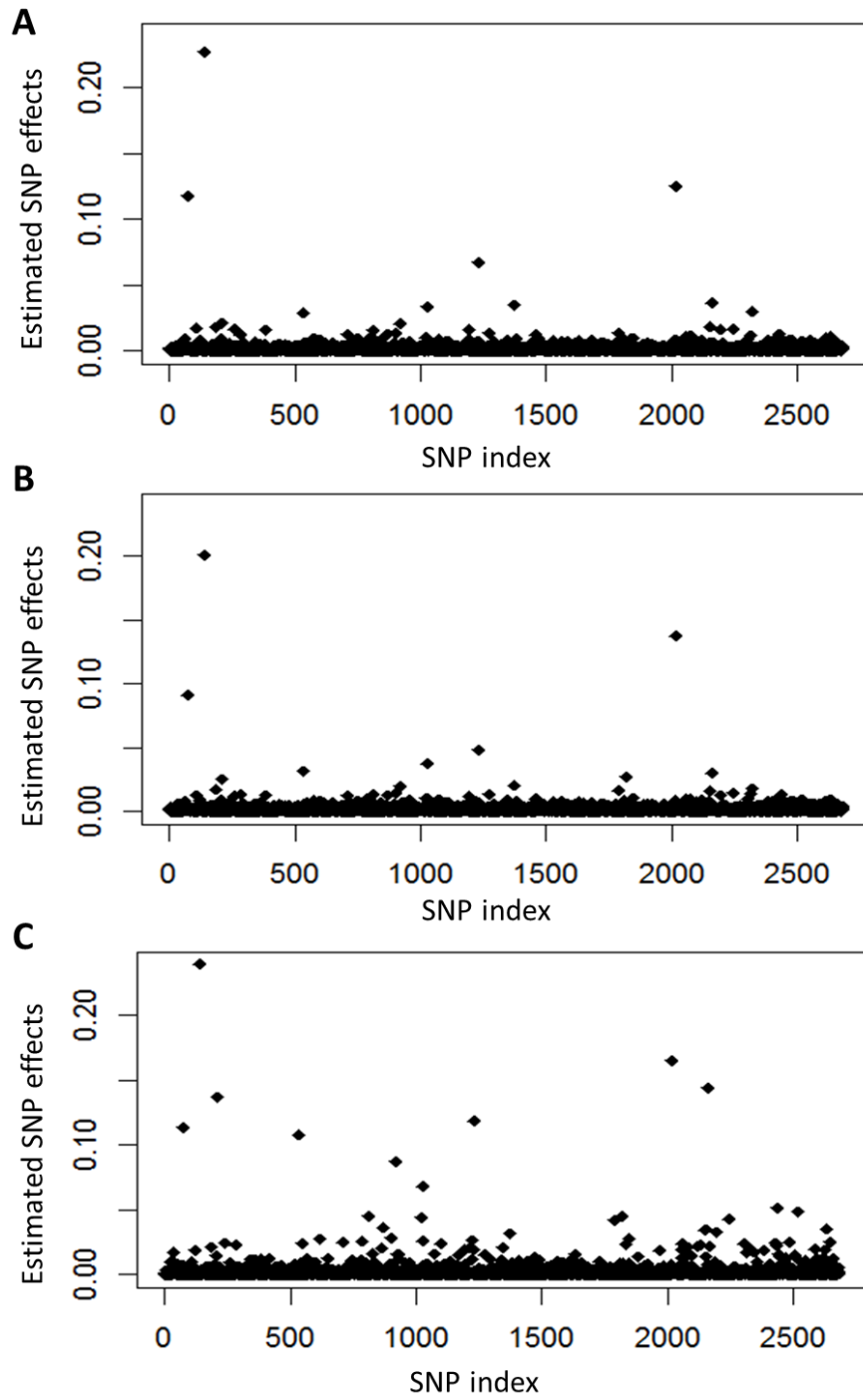


Figure 5.3: Estimated SNP effects for RGV from using 807 individuals and 2684 SNPs in Pine data set using three methods: A) uBayesA; B) IWBayesA; C) CDBayesB1.

5.4 Discussion

Multiple trait extensions to WGP have been developed to improve prediction accuracy by accounting for genetic correlations between traits (CALUS and VEERKAMP 2011; JIA and JANNINK 2012). Many studies have shown the advantage of multiple trait models compared to their univariate counterparts especially for lower heritability traits. For multiple trait models, variable selection methods such as BayesSSVS have shown some advantage in prediction accuracy over models based on no such mixtures and/or normality (CALUS and VEERKAMP 2011). Nevertheless, in Jia and Jannink (2012), a variable selection method BayesC π , based on normality for one of the mixtures, demonstrated advantages over bivariate BayesA, which is similar to our IW-BayesA, and GBLUP. A Gaussian prior on SNP effects might not be an ideal specification for genetic architectures characterized by few large QTL effects. Furthermore, heavy-tailed variable selection methods like BayesB popularized for univariate WGP analyses have not yet been considered in multiple trait analysis. In this study, we developed multiple trait WGP models based on the two univariate methods BayesA and BayesB. We did not, however, pursue a variable selection of IWBayesA, analogous to IWBayesB developed in the previous chapter, as we believed it to be dubious to attempt to fit a model where SNP effects were either both zero or both non-zero for different traits.

In univariate trait WGP analyses, many factors, e.g. number of animals, number of SNP markers, number of QTLs and heritability for the trait, could influence prediction accuracy (MEUWISSEN and GODDARD 2010). According to Jia and Jannink (2012) factors such as number of QTLs, genetic correlation and heritabilities for the two traits could influence the prediction accuracy in bivariate trait WGP analyses,. They pursued “change

one factor at a time” techniques for their simulation study experimental design, which reduced the total number of simulated replicates to be generated and analyzed by various competing WGP models compared to a full factorial design. Conversely, we used a response surface design to quickly pinpoint factors that might be particularly important for influencing differences in accuracy between CDBayesB and IWBayesA for WGP prediction leading us to focus in on M_{12} , the number of pleiotropic loci.

In previously developed bivariate trait WGP simulations as well as distributional assumptions for various modeling, QTLs have been generally assumed to be always pleiotropic, often to the point that the genetic correlation between two traits is uniform throughout the genome (CALUS and VEERKAMP 2011; JIA and JANNINK 2012). Under such scenarios, these investigators generally found greater advantages for bivariate trait compared to univariate trait WGP models. Conversely, we considered a situation whereby QTL may be either pleiotropic or non-pleiotropic in their effects between the two traits. We specifically defined three categories for QTLs, where M_1 and M_2 was the number of non-pleiotropic QTLs for each of the two respective traits, while M_{12} represented the number of pleiotropic QTLs. We further allowed for the fact, in both our simulations and some of our models (e.g., IWBayesA, CDBayesA, and CDBayesB) that the nature of the strength of association (i.e., genetic correlation) might be rather heterogeneous across pleiotropic QTL as association variables (ϕ_j) between two traits for the M_{12} QTL were drawn from a normal distribution.

In a focused LD simulation study, the only difference between two competing scenarios was M_{12} ($M_{12} = 10$ in Scenario 1, $M_{12} = 30$ in Scenario 2). We found that bivariate trait models improved accuracy of WGP compared to univariate BayesB for the

lower heritability trait in Scenario 2 but not in Scenario 1. That is, increasing M_{12} seemed to provide more power to detect pleiotropic effects in bivariate trait models. However, we did not detect any difference in WGP accuracy between CDBayesB and IWBayesA in Scenario 2, even though CDBayesB, unlike IWBayesA, implicitly distinguishes between pleiotropic versus non-pleiotropic QTL.

Although IWBayesA is a convenient choice for bivariate trait WGP modeling and closely mirrors the multivariate BayesA procedure previously developed by Calus and Veerkamp (2011), it is imperative that hyperparameters like degrees of freedom and scale parameters are properly “tuned” or inferred upon rather than set to arbitrary values; otherwise, WGP accuracies can be badly compromised (JIA and JANNINK 2012). In this study, we inferred upon hyperparameters based on the specification of diffuse prior distributions. However, we also recognized in this study that mixing problems can arise with real data applications such that it might be necessary to tune these hyperparameters somewhat on univariate analyses.

In our LD simulation study, we found that IWBayesA has much lower accuracy compared to other competing methods for the higher heritability trait. Conversely, in Jia and Jannink (2012), IWBayesA outperformed uBayesA in their default simulation scenario. The reason for the different results might be how QTL were differently generated between our two studies. IWBayesA assumes that every SNP has a pleiotropic effect, which partly agrees with the specifications in Jia and Jannink’s simulation study. However, we also specified a substantial proportion of QTL in our study to be non-pleiotropic (i.e., trait-specific). This may have resulted in the IWBayesA being rather

inflexible relative to other competing models, including even univariate WGP models, for the higher heritability trait.

In our simulation study, bGBLUP performed poorly compared to IWBAYESB and even to univariate WGP models (uBAYESB) for Trait 1 (high heritability trait) under both Scenarios 1 and 2. The assumption of bGBLUP states that SNP effects for bivariate traits follows a light tail (multivariate normal) distribution, which is often violated when there are only a few QTL that underly both traits. Our simulation study and the analysis of the pine data indicated that bGBLUP had lower WGP accuracy compared to IWBAYESB which effectively assumes that SNP effects follow a heavier tailed multivariate t distribution. This is also consistent with conclusions drawn by Jia and Jannink (2012).

We found CDBAYESB had much higher predictive ability in cross validation on RGV in the pine data analysis. Jia and Jannink (2012) studied the same dataset in their comparisons. However, they didn't find any significant difference in cross-validation performance between bivariate and univariate WGP trait models (except for the situation when they assumed some missing values for one trait). With the ability to differentially infer upon both pleiotropic and non-pleiotropic effects, we believe that CDBAYESB offers more flexibility compared to other competing models, including those previously tested by Jia and Jannink (2012) and Calus and Veerkamp (2011).

CDBAYESA model is a special case of CDBAYESB with probabilities of nonzero effects on both traits and associations set to 1. Unlike IWBAYESB, it is necessary to specify an order for the two traits for both CD models. In LD simulation, we analyzed the

data under the two models, using the same order of traits we simulated. However, it might not be obvious whether the order of traits specified is important in actual applications. We initially used the higher heritability trait (RBIN) as the first trait and the lower heritability trait (RGV) as the second trait for both CD models. After switching orders for these two traits, we found that predictive ability was unaffected for both CDBayesA and CD-BayesB. One possible reason for this is that these two models might be far more flexible than IWBayesA for distinguishing between pleiotropic versus non-pleiotropic QTL.

5.5 Conclusions

Alternative Cholesky-based parameterizations (CDBayesA\CDBayesB) and inverted Wishart specification on VCVs (IWBayesA) for bivariate trait WGP models were investigated for their advantage in prediction accuracy compared to bivariate ridge regression (bGBLUP) and univariate WGP models (uBayesA\uBayesB). With both non-pleiotropic and pleiotropic QTLs specified in two scenarios of LD simulation, the three bivariate trait WGP models had higher accuracy than the two univariate trait models (~8%) when the number of pleiotropic QTLs was relatively large. For the low heritability trait in the two scenarios, the three Bayesian bivariate trait WGP models outperformed bGBLUP ($P<0.05$). However, we didn't find any significant difference among the three Bayesian bivariate trait WGP models in both scenarios. Jointly accounting for pleiotropic and non-pleiotropic SNP effects in CDBayesB is obviously more flexible compared to bivariate models (CDBayesA and IWBayesA) assuming all SNP are pleiotropic. Due to its flexibility, CDBayesB had higher predictive ability (~5%) compared to other competing models regardless of the order on the two traits in application to pine data.

Chapter 6 Discussion, Conclusions and Future Work

This dissertation has focused on extending statistical models and developing computing strategies to better conduct whole genome prediction (WGP) for selection of breeding stock for economically important traits based on high density single nucleotide polymorphisms (SNP) marker panels. The primary intent of this work was to develop greater flexibility of WGP models in a number of potentially different ways. One key enhancement was to model potential spatial correlation between SNP effects due to the presence of QTL (Chapter 2). Another was to allow for potentially different modes of genetic action (i.e., pleiotropic versus trait-specific), whether for reaction norm models that account for a specific form of genotype by environment interaction (Chapter 4) or for bivariate trait analysis (Chapter 5). Additional hybrid models that combine the features of various WGP models in this dissertation (e.g., bivariate antedependence models) could be conceptually derived and tested in future work as well.

Some researchers might be rather critical of these efforts, recognizing that this dissertation has only added further to the “Bayesian alphabet” (Gianola, de los Campos et al. 2009; Gianola 2013) given proposed model labels such as “ante-BayesA” or “CD-BayesB”, for example. This criticism is certainly warranted if key hyperparameters are not properly tuned, since improper tuning would only distort comparisons between the models proposed in this dissertation and more conventional models used in current WGP implementations. Hence this work has been prepared with this issue keenly in mind, presenting fully Bayesian inferential strategies to infer upon these key hyperparameters in every instance whenever possible. In fact, an entire chapter (Chapter 3) addresses

computational efficiencies for alternative strategies and the impact of hyperparameter misspecification in WGP models.

However, we recognize that there is much more work that needs to be done on this front, particularly as the applications in this dissertation were smaller in scale, i.e., with respect to number of genotypes m and number of phenotypes n , compared to many current applications. Even in those cases, some difficulties were encountered. For example, we resorted to inferring upon some key hyperparameters using conventional univariate models before properly tuning the key bivariate hyperparameters for some of the bivariate genomic analyses in Chapter 5.

Some of the models developed in this dissertation did not always perform better than more conventional specifications; indeed, this was contrary to our expectations based on simulated genetic architectures. We realize that all of the simulation studies and applications considered in this dissertation are, by no means, exhaustive; nevertheless we do, for example, note the following. Antedependence specifications (ante-BayesA/ ante-BayesB) seemed to show particular advantages when linkage disequilibrium was substantial (Chapter 2) whereas finely constructed bivariate genomic models such as CD-BayesB which differentially model pleiotropic from non-pleiotropic QTL did particularly well when the genetic architecture was simple (Chapter 5); i.e., low numbers (m_{QTL}) of QTL relative to the number of SNP markers (m). However, on nearly just as many occasions, we did not detect meaningful differences in WGP accuracy between seemingly disparate model specifications. For example, results were often counterintuitive with our reaction norm model work (Chapter 4) in that sometimes IW-BayesA, a model that assumes complete pleiotropy throughout the genome, did better than CD-BayesB which

was constructed to loosen that requirement. Recent work from Wimmer et al. (2013) might be particularly enlightening in that regard; that is, they concluded that Bayesian variable selection models are not likely to confer substantial advantages over simpler GBLUP specifications when the heritability is low, the level of determinedness (n/m), is low, the model complexity (m_{QTL}/n) is high and/or the LD is high. This may partially explain why some of the proposed variable selection methods in this dissertation (e.g. CD-BayesB) did not confer substantial advantages for some of the smaller scale examples considered.

Nevertheless, these rules do not necessarily apply to looking at different distributional forms, e.g., Student t versus normal, with extensions to various bivariate forms, e.g. based on inverted Wishart specifications versus more flexible specifications based on the Cholesky decomposition, especially if model complexity is high. Furthermore, the level of determinedness is increasing so fast in some populations, e.g. Holsteins, that now $n > m$ (LEGARRA and DUCROCQ 2012) such that it might become increasingly more feasible to develop more comprehensive WGP models. This might become particularly true for GWAS types of analyses where it has been noted in this dissertation that inferences on individual SNP effects may be sensitive to model specification. Furthermore although m and pairwise LD will admittedly increase with sequencing technologies and thereby limit the effectiveness of more elaborate WGP model specifications, it is also then more likely that future WGP models might be based on haplotypes of SNP rather than single SNP per se, thereby further distorting the issue of WGP model fit and choice relative to work by, e.g., Wimmer et al. (2013). Hence, this continues to be a promising and exciting area of research.

There are certainly other strong limitations to the work in this dissertation that may further distort the comparisons between the various WGP models, particularly for analyses that involve real data. Firstly, we assumed genetic effects to be strictly additive, such that it is unpredictable what the effects of non-additive gene action might be (Gianola, Wu et al. 2010) on our comparisons. Certainly, there may be other nonlinearities that might not have been accounted for in our work as well. Plasmode-based simulations may represent a more effective way of reassessing the relative performance of WGP models (Vaughan, Divers et al. 2009).

We have already mentioned the computational limitations of some of our proposed models, particularly when hyperparameters need to be estimated. Animal breeders have been reticent, at best, to attempt to infer upon or properly tune these hyperparameters for good reason; it has been rather difficult to do so except, perhaps, based on some method of moments based determinations (de los Campos, Hickey et al. 2013). Although the toolkit in this dissertation was based on MCMC, it might be prudent to pursue other computationally feasible analytical approximations based on, for example, variational Bayes (Logsdon, Hoffman et al. 2010) or expectation-maximization like methods (KARKKAINEN and SILLANPAA 2012). A similar argument also applies to hyperparameter estimation. For example, in ridge regression or GBLUP like models, REML could be used to estimate the key “hyperparameters” like the common SNP variance component and the residual variance for example; certainly something similar could be done for Student t (e.g. Bayes A) implementations as well; e.g. (Pinheiro, Liu et al. 2001). This should be another fruitful area for future research.

APPENDICES

APPENDIX A: Chapter 2

A1 Markov Chain Monte Carlo Implementation Strategy for Ante-BayesA and Ante-BayesB

In order to conduct MCMC, it is necessary to first specify the joint posterior density of all unknown parameters (SORENSEN and GIANOLA 2002). To do this, we interchangeably reparameterize the joint density of the data \mathbf{y} and the random SNP effects, using \mathbf{g} for ante-BayesA and $\boldsymbol{\delta}$ for ante-BayesB in order to exploit algorithmic efficiencies that are unique to either model. For instance with ante-BayesA, we write

$$p(\mathbf{y}, \mathbf{g} | \boldsymbol{\beta}, \mathbf{u}, \boldsymbol{\sigma}_\delta, \mathbf{t}, \sigma_e^2) = p(\mathbf{y} | \boldsymbol{\beta}, \mathbf{u}, \mathbf{g}, \sigma_e^2) p(\mathbf{g} | \boldsymbol{\sigma}_\delta, \mathbf{t}) \quad [\text{A1}]$$

Note the component $p(\mathbf{y} | \boldsymbol{\beta}, \mathbf{u}, \mathbf{g}, \sigma_e^2) = N(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{g} + \mathbf{W}\mathbf{u}, \mathbf{I}\sigma_e^2)$ is based on Equation [1]

whereas $\mathbf{g} \sim p(\mathbf{g} | \boldsymbol{\sigma}_\delta, \mathbf{t}) = N(\mathbf{0}, \boldsymbol{\Sigma}_g)$ with $\boldsymbol{\Sigma}_g = (\mathbf{I} - \mathbf{T})^{-1} \boldsymbol{\Delta} (\mathbf{I} - \mathbf{T})^{-1'}$ are defined by

elements in $\boldsymbol{\sigma}_\delta = [\sigma_{\delta_1}^2 \ \sigma_{\delta_2}^2 \ \sigma_{\delta_3}^2 \ \dots \ \sigma_{\delta_m}^2]$ specified along the diagonal of $\boldsymbol{\Delta}$, and by

$\mathbf{t} = [t_{2,1}, t_{3,2}, \dots, t_{m,m-1}]'$ specified just below the diagonal elements in \mathbf{T} as previously

indicated. For ante-BayesB, we reparameterize [A1] differently:

$$p(\mathbf{y}, \boldsymbol{\delta} | \boldsymbol{\beta}, \mathbf{u}, \boldsymbol{\sigma}_\delta, \mathbf{t}, \sigma_e^2) = p(\mathbf{y} | \boldsymbol{\beta}, \boldsymbol{\delta}, \mathbf{t}, \sigma_e^2) p(\boldsymbol{\delta} | \boldsymbol{\sigma}_\delta) \quad [\text{A2}]$$

recognizing that $\boldsymbol{\delta} = (\mathbf{I} - \mathbf{T})\mathbf{g}$ such that [A2] represents a linear transformation of [A1].

That is, the first component of [A2] is based on

$$p(\mathbf{y} | \boldsymbol{\beta}, \boldsymbol{\delta}, \mathbf{t}, \sigma_e^2) = N(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}(\mathbf{I} - \mathbf{T})^{-1} \boldsymbol{\delta} + \mathbf{W}\mathbf{u}, \mathbf{I}\sigma_e^2) \text{ whereas } p(\boldsymbol{\delta} | \boldsymbol{\sigma}_\delta) = \prod_{j=1}^m N(0, \sigma_{\delta_j}^2).$$

We'll subsequently represent [A1] and [A2] together as $p(\mathbf{y}, \mathbf{g}(\boldsymbol{\delta}) | \boldsymbol{\beta}, \mathbf{u}, \boldsymbol{\sigma}_\delta, \mathbf{t}, \sigma_e^2, \mu_t, \sigma_t^2)$

to recognize the interchangeability between \mathbf{g} and δ when conditioning on \mathbf{t} . The joint posterior density of all unknown parameters can be written as products of specifications provided previously:

$$\begin{aligned}
& p(\boldsymbol{\beta}, \mathbf{g}, \mathbf{u}, \sigma_e^2, \sigma_\delta^2, \mathbf{t}, \mu_t, \sigma_t^2, \nu_\delta, s_\delta^2 | \mathbf{y}) \propto \\
& p(\mathbf{y}, \mathbf{g}(\delta) | \boldsymbol{\beta}, \mathbf{u}, \sigma_e^2, \sigma_\delta^2, \mathbf{t}, \mu_t, \sigma_t^2) p(\boldsymbol{\beta}) \left(\prod_{j=2}^m p(t_{j,j-1} | \mu_t, \sigma_t^2) \right) \\
& \left(\prod_{j=1}^m p(\sigma_{\delta_j}^2 | \nu_\delta, s_\delta^2, \pi_\delta) \right) p(\sigma_u^2 | \nu_u, s_u^2), p(\sigma_e^2 | \nu_e, s_e^2) p(\nu_\delta) \\
& p(s_\delta^2 | \alpha_s, \beta_s) p(\mu_t | \mu_{t0}, s_{t0}^2) p(\sigma_t^2 | \nu_t, s_t^2) p(\pi_\delta | \alpha_\pi, \beta_\pi)
\end{aligned} \tag{A3}$$

From the paper, $p(\boldsymbol{\beta}) = N(\boldsymbol{\beta}, \mathbf{V}_\beta)$, $p(t_{j,j-1} | \mu_t, \sigma_t^2) = N(\mu_t, \sigma_t^2)$, $p(\sigma_u^2 | \nu_u, s_u^2) = \chi^{-2}(\nu_u, \nu_u s_u^2)$,

$p(\sigma_e^2 | \nu_e, s_e^2) = \chi^{-2}(\nu_e, \nu_e s_e^2)$, $p(s_\delta^2 | \alpha_s, \beta_s) = \text{Gamma}(\alpha_s, \beta_s)$, $p(\mu_t | \mu_{t0}, s_{t0}^2) = N(\mu_{t0}, s_{t0}^2)$,

$p(\sigma_t^2 | \nu_t, s_t^2) = \chi^{-2}(\nu_t, \nu_t s_t^2)$, and $p(\pi_\delta | \alpha_\pi, \beta_\pi) = \text{Beta}(\alpha_\pi, \beta_\pi)$. Furthermore,

$p(\sigma_{\delta_j}^2 | \nu_\delta, s_\delta^2, \pi_\delta)$ is a mixture analogous to Equation [2] for ante-BayesB whereas

$p(\sigma_{\delta_j}^2 | \nu_\delta, s_\delta^2, \pi_\delta = 1) = \chi^{-2}(\nu_\delta, \nu_\delta s_\delta^2)$ for ante-BayesA as described in the paper. For some

parameters, we subsequently derive and present FCD separately for ante-BayesA ($\pi_\delta = 1$)

from ante-BayesB ($\pi_\delta < 1$) as some MCMC sampling strategies appear to be simpler or

more computationally efficient for one or the other model. Now MCMC requires

random draws from the full conditional densities of each unknown parameter (or blocks thereof) conditional on all other parameters and the data (SORENSEN and GIANOLA 2002).

These full conditional densities are provided below for various classes of these unknown parameters.

To sample all fixed and random effects in ante-BayesA, write $\boldsymbol{\theta} = [\boldsymbol{\beta}' \quad \mathbf{g}' \quad \mathbf{u}']'$ as the $(p+m+q)$ vector of fixed and random effects, $\mathbf{Q} = [\mathbf{X} \vdots \mathbf{Z} \vdots \mathbf{W}]$ as the $n \times (p+m+q)$ overall model incidence matrix with $\boldsymbol{\Sigma}^- = \text{diag}(\mathbf{V}_{\boldsymbol{\beta}}^{-1} \quad \boldsymbol{\Sigma}_{\mathbf{g}}^{-1} \quad \mathbf{A}^{-1}\sigma_u^{-2})$ as a block diagonal matrix with the corresponding listed components as the various blocks. It can be readily demonstrated (SORENSEN and GIANOLA 2002) that the FCD of $\boldsymbol{\theta}$ is

$$\boldsymbol{\theta} | \mathbf{y}, ELSE \sim N(\hat{\boldsymbol{\theta}}, \mathbf{C}) \quad [\text{A4}]$$

where *ELSE* denotes all other parameters in [A3] other than $\boldsymbol{\theta}$ and

$$\hat{\boldsymbol{\theta}} = \mathbf{CQ}'\mathbf{y} + [\boldsymbol{\beta}_0' \mathbf{V}_{\boldsymbol{\beta}}^{-1} \quad \mathbf{0}_{1 \times (m+q)}]'$$
 for $\mathbf{C} = (\mathbf{Q}'\mathbf{Q} + \boldsymbol{\Sigma}^-)^{-1} \sigma_e^2$. Note that with a typical “flat”

prior for $\boldsymbol{\beta}$ is defined by $\mathbf{V}_{\boldsymbol{\beta}}^{-1} = \mathbf{0}$ such that $\hat{\boldsymbol{\theta}} = \mathbf{CQ}'\mathbf{y}$. Also, note that univariate or

multivariate block FCD subsets of $\boldsymbol{\theta}$ could also be partitioned and sampled using [A4]

based on results from Wang and Gianola (1994). The structure of $\boldsymbol{\Sigma}_{\mathbf{g}}^{-1} = \{\boldsymbol{\Sigma}_{\mathbf{g}}^{jj'}\}$

contained within $\boldsymbol{\Sigma}^-$ is a simple tri-diagonal matrix: using ZIMMERMAN and NÚÑEZ-

ANTÓN (2010), the diagonal elements are $\boldsymbol{\Sigma}_{\mathbf{g}}^{jj} = \sigma_{\delta_j}^{-2} + t_{j+1,j}^2 \sigma_{\delta_{j+1}}^{-2}$ for $j = 1, 2, \dots, m-1$ with

$\boldsymbol{\Sigma}_{\mathbf{g}}^{mm} = \sigma_{\delta_m}^{-2}$ whereas the elements adjacent to the diagonal are

$$\boldsymbol{\Sigma}_{\mathbf{g}}^{j,j+1} = \boldsymbol{\Sigma}_{\mathbf{g}}^{j+1,j} = -t_{j+1,j} \sigma_{\delta_{j+1}}^{-2}.$$

To sample marker-specific variances in ante-BayesA: Consider now the FCD for $\sigma_{\delta_j}^2, j=1, 2, \dots, m$:

$$p(\sigma_{\delta_j}^2 | \mathbf{y}, ELSE) \propto p(\mathbf{g} | t_{21}, t_{32}, \dots, t_{m,m-1}, \sigma_{\delta_1}^2, \sigma_{\delta_2}^2, \sigma_{\delta_3}^2, \dots, \sigma_{\delta_m}^2) p(\sigma_{\delta_j}^2 | \nu_{\delta}, s_{\delta}^2) \quad [\text{A5}]$$

We use Chan and Jeliazkov (2009, pg 461) to simplify the first component of [A5] as follows:

$$\begin{aligned}
& p(\mathbf{g} | t_{21}, t_{32}, \dots, t_{G,G-1}, \sigma_{\delta_1}^2, \sigma_{\delta_2}^2, \sigma_{\delta_3}^2, \dots, \sigma_{\delta_G}^2) \\
& \propto |\mathbf{G}^{-1}|^{1/2} \exp\left(-\frac{1}{2} \mathbf{g}' \Sigma_{\mathbf{g}}^{-1} \mathbf{g}\right) = |(\mathbf{I} - \mathbf{T})' \Lambda^{-1} (\mathbf{I} - \mathbf{T})|^{1/2} \exp\left(-\frac{1}{2} \mathbf{g}' (\mathbf{I} - \mathbf{T})' \Lambda^{-1} (\mathbf{I} - \mathbf{T}) \mathbf{g}\right) \\
& \propto |\Lambda^{-1}|^{1/2} \exp\left(-\frac{1}{2} \boldsymbol{\delta}' \Lambda^{-1} \boldsymbol{\delta}\right) \propto \prod_{j=1}^G (\sigma_{\delta_j}^2)^{-1/2} \exp\left(-\frac{1}{2} \frac{\delta_j^2}{\sigma_{\delta_j}^2}\right)
\end{aligned}
\tag{A6}$$

Using the component in [A6] pertaining to $\sigma_{\delta_j}^2$ in [A5] and

$$\begin{aligned}
p(\sigma_{\delta_j}^2 | \nu_{\delta}, s_{\delta}^2) & \propto \sigma_{\delta_j}^2^{-\left(\frac{\nu_{\delta}}{2} + 1\right)} e^{-\frac{\nu_{\delta} s_{\delta}^2}{2\sigma_{\delta_j}^2}} \text{ then} \\
p(\sigma_{\delta_j}^2 | \mathbf{y}, ELSE) & \propto (\sigma_{\delta_j}^2)^{-1/2} \exp\left(-\frac{1}{2} \frac{\delta_j^2}{\sigma_{\delta_j}^2}\right) \frac{\left(\frac{\nu_{\delta} s_{\delta}^2}{2}\right)^{\frac{\nu_{\delta}}{2}}}{\Gamma\left(\frac{\nu_{\delta}}{2}\right)} \sigma_{\delta_j}^2^{-\left(\frac{\nu_{\delta}}{2} + 1\right)} e^{-\frac{\nu_{\delta} s_{\delta}^2}{2\sigma_{\delta_j}^2}} \\
& \propto (\sigma_{\delta_j}^2)^{-\left(\frac{\nu_{\delta} + 1}{2} + 1\right)} \exp\left(-\frac{1}{2} \frac{(\delta_j^2 + \nu_{\delta} s_{\delta}^2)}{\sigma_{\delta_j}^2}\right)
\end{aligned}
\tag{A7}$$

i.e. $p(\sigma_{\delta_j}^2 | \mathbf{y}, ELSE) = \chi^{-2}(\nu_{\delta} + 1, \delta_j^2 + \nu_{\delta} s_{\delta}^2)$. As a sidenote, elements of $\boldsymbol{\delta}$ can be

recursively derived from \mathbf{g} :

$$\boldsymbol{\delta} = (\mathbf{I} - \mathbf{T}) \mathbf{g} = \begin{bmatrix} 1 & 0 & \dots & 0 & 0 \\ -t_{21} & 1 & \dots & 0 & 0 \\ 0 & -t_{32} & & 0 & 0 \\ \vdots & & \ddots & 1 & 0 \\ 0 & 0 & & -t_{m,m-1} & 1 \end{bmatrix} \begin{bmatrix} g_1 \\ g_2 \\ g_3 \\ \vdots \\ g_m \end{bmatrix} = \begin{bmatrix} g_1 \\ g_2 - t_{21} g_1 \\ g_3 - t_{32} g_2 \\ \vdots \\ g_m - t_{m,m-1} g_{m-1} \end{bmatrix}
\tag{A8}$$

To sample fixed and random effects other than SNP effects in ante-BayesB, here we deem it computationally tractable to sample the rest of the location parameters separately from \mathbf{g} . We again use Equation [A4] except that now we define $\boldsymbol{\theta} = [\boldsymbol{\beta}' \quad \mathbf{u}']'$ as a $(p+q) \times 1$ vector of fixed and random polygenic effects with $\mathbf{Q} = [\mathbf{X} \quad \mathbf{W}]$ being the corresponding $n \times (p+q)$ submodel incidence matrix and $\boldsymbol{\Sigma}^- = \text{diag}(\mathbf{0}_{p \times p} \quad \mathbf{A}^{-1} \sigma_u^{-2})$ being the corresponding block diagonal matrix. We then sample using Equation [A4] and $\hat{\boldsymbol{\theta}} = \mathbf{CQ}'(\mathbf{y} - \mathbf{Zg}) + [\boldsymbol{\beta}_0' \mathbf{V}_{\beta}^{-1} \quad \mathbf{0}_{1 \times q}]'$ for $\mathbf{C} = (\mathbf{Q}'\mathbf{Q} + \boldsymbol{\Sigma}^-)^{-1} \sigma_e^2$.

To sample random SNP effects and variances in ante-BayesB, we consider the collapsed sampling strategy (Liu 1994) for jointly sampling $\sigma_{\delta_j}^2$ and δ_j as previously adapted for Bayes B in Meuwissen et al. (2001). Consider the previously described mixture prior on the conditional variances

$$p(\sigma_{\delta_j}^2 | v_{\delta}, s_{\delta}^2, \pi_{\delta}) = \begin{cases} 0 & \text{with probability } \pi_{\delta} \\ \chi^{-2}(v_{\delta}, v_{\delta} s_{\delta}^2) & \text{with probability } 1 - \pi_{\delta} \end{cases} \quad [\text{A9}]$$

We jointly sample $\sigma_{\delta_j}^2$ and δ_j from $p(\sigma_{\delta_j}^2, \delta_j | ELSE, \mathbf{y})$, by sampling first from

$p(\sigma_{\delta_j}^2 | \mathbf{y}, ELSE \text{ except } \delta_j)$ and then from $p(\delta_j | ELSE, \mathbf{y})$. The first component of [A2]

implies the following linear model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{H}\boldsymbol{\delta} + \mathbf{W}\mathbf{u} + \mathbf{e} \quad [\text{A10}]$$

where $\mathbf{H} = \mathbf{Z}(\mathbf{I} - \mathbf{T})^{-1}$. Let's further partition \mathbf{H} into the j th column, \mathbf{h}_j , and other remaining columns \mathbf{H}_{-j} ; similarly, we represent $\boldsymbol{\delta}_{-j}$ as all elements of $\boldsymbol{\delta}$ other than δ_j .

Then we further rewrite [A10] as follows:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{H}_{-j}\boldsymbol{\delta}_{-j} + \mathbf{h}_j\delta_j + \mathbf{W}\mathbf{u} + \mathbf{e} \quad [\text{A11}]$$

It can be readily demonstrated, following similar developments for BayesB provided by Meuwissen et al. (2001), that:

$$\begin{aligned} p(\sigma_{\delta_j}^2 | \mathbf{y}, ELSE \text{ except } \delta_j) &= \int_{\delta_j} p(\sigma_{\delta_j}^2, \delta_j | \mathbf{y}, ELSE) \\ &\propto \int_{\delta_j} p(\mathbf{y} | \boldsymbol{\beta}, \boldsymbol{\delta}, \mathbf{u}, \mathbf{t}, \sigma_e^2) p(\delta_j | \sigma_{\delta_j}^2) p(\sigma_{\delta_j}^2 | \nu_{\delta}, s_{\delta}^2, \pi) d\delta_j \\ &\propto p(\sigma_{\delta_j}^2 | \nu_{\delta}, s_{\delta}^2, \pi) \int_{\delta_j} \exp\left(-\frac{1}{2\sigma_e^2}(\mathbf{y}_j^* - \mathbf{h}_j\delta_j)'(\mathbf{y}_j^* - \mathbf{h}_j\delta_j)\right) \exp\left(-\frac{1}{2\sigma_{\delta_j}^2}\delta_j^2\right) d\delta_j \\ &\propto p(\sigma_{\delta_j}^2 | \nu_{\delta}, s_{\delta}^2, \pi) |\mathbf{V}_j|^{-1/2} \exp\left(-\frac{1}{2}\mathbf{y}_j^{*'}\mathbf{V}_j^{-1}\mathbf{y}_j^*\right) \end{aligned} \quad [\text{A12}]$$

where $\mathbf{y}_j^* = \mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{H}_{-j}\boldsymbol{\delta}_{-j} - \mathbf{W}\mathbf{u}$ and $\mathbf{V}_j = \mathbf{h}_j\mathbf{h}_j'\sigma_{\delta_j}^2 + \mathbf{I}\sigma_e^2$.

Since [A12] is not a recognizable distributional form, a Metropolis Hastings step is required. We adapt the independence chains implementation (CHIB and GREENBERG 1995) as also adapted by Meuwissen et al. (2001) using the prior $p(\sigma_{\delta_j}^2 | \nu_{\delta}, s_{\delta}^2, \pi_{\delta})$ as the candidate density. That is, at MCMC cycle $[k]$, one samples a candidate, say, $\sigma_{\delta_j}^{2[*]}$, from $p(\sigma_{\delta_j}^2 | \nu_{\delta}, s_{\delta}^2, \pi_{\delta})$ conditioned upon the updated values for $\nu_{\delta}, s_{\delta}^2$ and π_{δ} . One

accepts $\sigma_{\delta_{j[k]}}^2 = \sigma_{\delta_{j^*}}^2$ as the value for in cycle $[k]$ with probability based on the Metropolis-

Hastings acceptance ratio $q(\sigma_{\delta_{j[k-1]}}^2 \rightarrow \sigma_{\delta_{j^*}}^2)$:

$$q(\sigma_{\delta_{j[k-1]}}^2 \rightarrow \sigma_{\delta_{j^*}}^2) = \begin{cases} \min \left(\frac{p(\sigma_{\delta_{j^*}}^2 | \mathbf{y}, ELSE \text{ except } \delta_j) p(\sigma_{\delta_{j[k-1]}}^2 | \nu_\delta, s_\delta^2, \pi)}{p(\sigma_{\delta_{j[k-1]}}^2 | \mathbf{y}, ELSE \text{ except } \delta_j) p(\sigma_{\delta_{j^*}}^2 | \nu_\delta, s_\delta^2, \pi)}, 1 \right); \\ 1, otherwise \end{cases}$$

[A13]

If the proposal $\sigma_{\delta_{j^*}}^2$ is rejected, then set $\sigma_{\delta_{j[k]}}^2 = \sigma_{\delta_{j[k-1]}}^2$; i.e., the value of $\sigma_{\delta_j}^2$ in the previous MCMC cycle. It can be demonstrated that using Meuwissen et al. (2001) that [A13] is further equal to:

$$q(\sigma_{\delta_{j[k-1]}}^2 \rightarrow \sigma_{\delta_{j^*}}^2) = \begin{cases} \min \left(\frac{|\mathbf{V}_j^*|^{-1/2} \exp\left(-\frac{1}{2} \mathbf{y}^*{}' \mathbf{V}_j^{*-1} \mathbf{y}^*\right)}{|\mathbf{V}_j^{[t-1]}|^{-1/2} \exp\left(-\frac{1}{2} \mathbf{y}^*{}' (\mathbf{V}_j^{[t-1]})^{-1} \mathbf{y}^*\right)}, 1 \right) \\ 1, otherwise \end{cases} \quad [A14]$$

Note that neither the determinant $|\mathbf{V}_j|$ nor the inverse \mathbf{V}_j^{-1} are trivial computations since m is typically large. Adapting a development from Rohan Fernando (personal communication) for BayesB, it can be readily shown that [A14] further simplifies:

$$q\left(\sigma_{\delta_{[k-1]}}^2 \rightarrow \sigma_{\delta_j^*}^2\right) = \begin{cases} \min \left(\frac{\left(v_j^*\right)^{-1/2} \exp\left(-\frac{1}{2} \frac{\left(\mathbf{h}_j' \mathbf{y}_j^*\right)^2}{v_j^*}\right)}{\left(v_j^{[t+1]}\right)^{-1/2} \exp\left(-\frac{1}{2} \frac{\left(\mathbf{h}_j' \mathbf{y}_j^*\right)^2}{v_j^{[t+1]}}\right)}, 1 \right. \\ \left. 1, \text{otherwise} \right. \end{cases} \quad [\text{A15}]$$

where $v_j^* = \left(\mathbf{h}_j' \mathbf{h}_j\right)^2 \sigma_{\delta_j^*}^2 + \left(\mathbf{h}_j' \mathbf{h}_j\right) \sigma_e^2$ and $v_j^{[k-1]} = \left(\mathbf{h}_j' \mathbf{h}_j\right)^2 \sigma_{\delta_{j[k-1]}}^2 + \left(\mathbf{h}_j' \mathbf{h}_j\right) \sigma_e^2$. Once $\sigma_{\delta_j}^2$ is sampled, one could immediately draw δ_j from $p\left(\delta_j \mid ELSE, \mathbf{y}\right)$ readily seen to be

$$p\left(\delta_j \mid ELSE, \mathbf{y}\right) = N\left(\frac{\mathbf{h}_j' \mathbf{y}_j^*}{\mathbf{h}_j' \mathbf{h}_j + \sigma_{\delta_j}^{-2}}, \frac{\sigma_e^2}{\mathbf{h}_j' \mathbf{h}_j + \sigma_{\delta_j}^{-2}}\right) \quad [\text{A16}]$$

in order to complete the joint collapsed sampler draw from $p\left(\sigma_{\delta_j}^2, \delta_j \mid ELSE, \mathbf{y}\right)$. One could demonstrate the following backward recursive relationship $\mathbf{h}_{j-1} = \mathbf{t}_{j,j-1} \mathbf{h}_j + \mathbf{z}_{j-1}$, $j = m, m-2, \dots, 2$ with \mathbf{z}_j denoting column j of \mathbf{Z} and $\mathbf{h}_m = \mathbf{z}_m$. Hence for computational tractability, one could use this relationship in sampling pairs from $p\left(\sigma_{\delta_j}^2, \delta_j \mid ELSE, \mathbf{y}\right)$ starting with $j = m$ and working recursively backwards to $j=1$.

To sample proportion of SNP markers associated with zero-effects in Ante-BayesB, the FCD of π_{δ} is based on the following:

$$p\left(\pi_{\delta} \mid \mathbf{y}, ELSE\right) \propto \prod_{j=1}^m p\left(\sigma_{\delta_j}^2 \mid v_{\delta}, s_{\delta}^2, \pi_{\delta}\right) p\left(\pi_{\delta} \mid \alpha_{\delta}, \beta_{\delta}\right) \quad [\text{A17}]$$

where $p(\boldsymbol{\pi}_\delta | \boldsymbol{\alpha}_\delta, \boldsymbol{\beta}_\delta) = \text{Beta}(\boldsymbol{\alpha}_\delta, \boldsymbol{\beta}_\delta)$. Let $m_1 = \sum_{j=1}^m I(\sigma_{\delta_j}^2 = 0)$ denote the number of zero-valued elements sampled in $\boldsymbol{\sigma}_\delta$ for a particular MCMC cycle where $I(\cdot)$ denotes the indicator function. Then it can be readily demonstrated that Equation [A17] is simply $p(\boldsymbol{\pi}_\delta | \mathbf{y}, ELSE) = \text{Beta}(\boldsymbol{\alpha}_\delta + m_1, \boldsymbol{\beta}_\delta + m - m_1)$.

To sample antedependence parameters and their corresponding hyperparameters, consider now deriving the joint FCD of $\mathbf{t} = [t_{2,1}, t_{3,2}, \dots, t_{m,m-1}]'$:

$$p(\mathbf{t} | \mathbf{y}, ELSE) \propto p(\mathbf{g} | t_{2,1}, t_{3,2}, \dots, t_{m,m-1}, \sigma_{\delta_1}^2, \sigma_{\delta_2}^2, \sigma_{\delta_3}^2, \dots, \sigma_{\delta_m}^2) \left(\prod_{j=2}^m p(t_{j,j-1} | \mu_t, \sigma_t^2) \right) \quad [A18]$$

Borrowing developments, again from Chan and Jeliakov (2009, pg 462), the first component of [A18] can be rewritten as:

$$\begin{aligned} & p(\mathbf{g} | t_{21}, t_{32}, \dots, t_{m,m-1}, \sigma_{\delta_1}^2, \sigma_{\delta_2}^2, \sigma_{\delta_3}^2, \dots, \sigma_{\delta_m}^2) \\ & \propto |(\mathbf{I} - \mathbf{T})' \boldsymbol{\Lambda}^{-1} (\mathbf{I} - \mathbf{T})|^{1/2} \exp\left(-\frac{1}{2} \mathbf{g}' (\mathbf{I} - \mathbf{T})' \boldsymbol{\Lambda}^{-1} (\mathbf{I} - \mathbf{T}) \mathbf{g}\right) \\ & \propto \exp\left(-\frac{1}{2} \frac{(g_2 - t_{21}g_1)^2}{\sigma_{\delta_2}^2}\right) \exp\left(-\frac{1}{2} \frac{(g_3 - t_{32}g_2)^2}{\sigma_{\delta_3}^2}\right) \dots \exp\left(-\frac{1}{2} \frac{(g_m - t_{m,m-1}g_{m-1})^2}{\sigma_{\delta_m}^2}\right) \\ & \propto \exp\left(-\frac{1}{2} (\mathbf{g}_{(-1)} - \boldsymbol{\Psi} \mathbf{t})' \boldsymbol{\Lambda}_{(-1)}^{-1} (\mathbf{g}_{(-1)} - \boldsymbol{\Psi} \mathbf{t})\right) \end{aligned}$$

[A19]

saving only terms that are functions of \mathbf{t} with $\Psi = \text{diag}(g_1, g_2, \dots, g_{m-1})$ being a diagonal

$m-1 \times m-1$ matrix with the listed elements, $\mathbf{g}_{(-1)} = [g_2 \ g_3 \ \dots \ g_m]'$, and

$\Delta_{(-1)} = \text{diag}(\sigma_{\delta_2}^2, \sigma_{\delta_3}^2, \dots, \sigma_{\delta_m}^2)$ being a diagonal $m-1 \times m-1$ matrix with the listed elements.

Hence, [A18] can be rewritten as follows:

$$\begin{aligned} p(\mathbf{t} | \mathbf{y}, ELSE) &\propto p(\mathbf{g}_{(-1)} | \mathbf{t}, \Delta_{(-1)}) p(\mathbf{t} | \mathbf{1}\mu_t, \mathbf{I}\sigma_t^2) \\ &\propto \left(\exp\left(-\frac{1}{2}(\mathbf{g}_{(-1)} - \Psi\mathbf{t})' \Delta_{(-1)}^{-1} (\mathbf{g}_{(-1)} - \Psi\mathbf{t})\right) \right) \exp\left(-\frac{1}{2\sigma_t^2}(\mathbf{t} - \mathbf{1}\mu_t)'(\mathbf{t} - \mathbf{1}\mu_t)\right) \\ &\propto \left(\exp\left(-\frac{1}{2}(\mathbf{t} - \hat{\mathbf{t}})' \Sigma_t^{-1} (\mathbf{t} - \hat{\mathbf{t}})\right) \right) \end{aligned} \quad [\text{A20}]$$

where

$$\hat{\Sigma}_t = \left(\Psi' \Delta_{(-1)}^{-1} \Psi + \mathbf{I}\sigma_t^{-2} \right)^{-1} \quad [\text{A21}]$$

and

$$\hat{\mathbf{t}} = \left(\Psi' \Delta_{(-1)}^{-1} \Psi + \mathbf{I}\sigma_t^{-2} \right)^{-1} \left(\Psi' \Delta_{(-1)}^{-1} \mathbf{g}_{(-1)} + \mathbf{1}\sigma_t^{-2} \mu_t \right) \quad [\text{A22}]$$

Note that $\Psi' \Delta_{(-1)}^{-1} \Psi + \mathbf{I}\sigma_t^{-2}$ is diagonal with elements $(g_j)^2 \sigma_{\delta_{j+1}}^{-2} + \sigma_t^{-2}$, $j = 1, 2, \dots, m-1$,

whereas element j of $\Psi' \Delta_{(-1)}^{-1} \mathbf{g}_{(-1)} + \mathbf{1}\sigma_t^{-2} \mu_t$ is $g_j g_{j+1} \sigma_{\delta_{j+1}}^{-2} + \sigma_t^{-2} \mu_t$, $j = 1, 2, \dots, m-1$. In other

words, the FCD of $t_{j+1,j}$ is $t_{j+1,j} | ELSE, \mathbf{y} \sim N(\hat{t}_{j+1,j}, \hat{\sigma}_{t(j+1,j)}^2)$ where

$$\hat{t}_{j+1,j} = \frac{g_j g_{j+1} \sigma_{\delta_{j+1}}^{-2} + \sigma_t^{-2} \mu_t}{(g_j)^2 \sigma_{\delta_{j+1}}^{-2} + \sigma_t^{-2}} \quad [\text{A23}]$$

and

$$\hat{\sigma}_{t(j+1,j)}^2 = \left((g_j)^2 \sigma_{\delta_{j+1}}^{-2} + \sigma_t^{-2} \right)^{-1} \quad [\text{A24}]$$

Note further that $\hat{t}_{j+1,j}$ can be written as a weighted average:

$$\hat{t}_{j+1,j} = \frac{g_j g_{j+1} \sigma_{\delta_{j+1}}^{-2} + \sigma_t^{-2} \mu_t}{\sigma_{\delta_{j+1}}^{-2} (g_j)^2 + \sigma_t^{-2}} = \frac{\sigma_{\delta_{j+1}}^{-2} (g_j)^2}{\sigma_{\delta_{j+1}}^{-2} (g_j)^2 + \sigma_t^{-2}} \frac{g_{j+1}}{g_j} + \frac{\sigma_t^{-2}}{\sigma_{\delta_{j+1}}^{-2} (g_j)^2 + \sigma_t^{-2}} \mu_t \quad [\text{A25}]$$

Now with $g_j = 0$, as one might anticipate occasionally with ante-BayesB with markers

defined at the beginning of a linkage group, $\hat{t}_{j+1,j} = \mu_t$ and $\hat{\sigma}_{t(j+1,j)}^2 = \sigma_t^2$ such that one

draws $t_{j+1,j}$ from its prior density based on updated values of μ_t and σ_t^2 . For the much

more common situation in ante-BayesB (assuming large π_δ) where $g_j \neq 0$ but $\sigma_{\delta_{j+1}}^2 = 0$,

the FCD of $t_{j+1,j}$ can be shown to be a normal with mean $\hat{t}_{j+1,j} = \frac{g_{j+1}}{g_j}$ and variance

$$\hat{\sigma}_{t(j+1,j)}^2 = \sigma_t^2.$$

With $p(\mu_t)$ specified to be normal with prior mean μ_{t0} and prior variance s_{t0}^2 then

Gibbs sampling can be used for the corresponding parameters.

$$p(\mu_t, | \mathbf{y}, \text{ELSE}) = N(\tilde{\mu}_t, \tilde{\sigma}_t^2) \quad [\text{A26}]$$

where

$$\tilde{\mu}_t = \frac{\frac{m-1}{\sigma_t^2} \bar{t} + \frac{1}{s_{t0}^2} \mu_{t0}}{\frac{1}{s_{t0}^2} + \frac{m-1}{\sigma_t^2}} \quad [\text{A27}]$$

$$\text{for } \bar{t} = \frac{\sum_{j=2}^m t_{j,j-1}}{m-1} \text{ and}$$

$$\tilde{\sigma}_t^2 = \left(\frac{1}{s_{t0}^2} + \frac{m-1}{\sigma_t^2} \right)^{-1} \quad [\text{A28}]$$

The FCD of σ_t^2 given that the prior $p(\sigma_t^2 | \nu_t, s_t^2)$ is scaled inverted chi-square with known hyperparameters ν_t and s_t^2 can be derived as follows:

$$\begin{aligned} & p(\sigma_t^2, | \mathbf{y}, ELSE) \\ & \propto \left(\prod_{j=2}^m p(t_{j,j-1} | \mu_t, \sigma_t^2) \right) p(\sigma_t^2 | \nu_t, s_t^2) \\ & \propto \left((2\pi\sigma_t^2)^{-(m-1)/2} \exp\left(-\frac{1}{2\sigma_t^2} \sum_{j=2}^m (t_{j,j-1} - \mu_t)^2\right) \right) \sigma_t^{2-\left(\frac{\nu_t}{2}+1\right)} e^{-\frac{\nu_t s_t^2}{2\sigma_t^2}} \\ & \propto \left((\sigma_t^2)^{-\left(\frac{\nu_t+m-1}{2}+1\right)} \exp\left(-\frac{1}{2\sigma_t^2} \left(\sum_{j=2}^m (t_{j,j-1} - \mu_t)^2 + \nu_t s_t^2 \right) \right) \right) \end{aligned} \quad [\text{A29}]$$

That is, $p(\sigma_t^2, | \mathbf{y}, ELSE) = \chi^{-2} \left(m + \nu_t, \sum_{j=2}^m (t_{j,j-1} - \mu_t)^2 + \nu_t s_t^2 \right)$. Note that we advocate

the non-informative specifications $\nu_t = -1$ and $s_t^2 = 0$ in the paper.

To sample the scale parameter for the random SNP effects, borrowing results from Yi and Xu (2008), the FCD for s_δ^2 based on the specification of a conjugate prior

$p(s_\delta^2 | \alpha_s, \beta_s) = \text{Gamma}(\alpha_s, \beta_s)$ can be written as follows:

$$\begin{aligned}
p(s_\delta^2 | \mathbf{y}, ELSE) &\propto \left(\prod_{j=1}^m I(\sigma_{\delta_j}^2 > 0) p(\sigma_{\delta_j}^2 | \nu_\delta, s_\delta^2) \right) p(s_\delta^2 | \alpha_s, \beta_s) \\
&= \left(\prod_{j=1}^m I(\sigma_{\delta_j}^2 > 0) \frac{\left(\frac{s_\delta^2 \nu_\delta}{2} \right)^{\frac{\nu_\delta}{2}}}{\Gamma\left(\frac{\nu_\delta}{2}\right)} \sigma_{\delta_j}^{-\left(\frac{\nu_\delta}{2}+1\right)} e^{-\frac{s_\delta^2 \nu_\delta}{2\sigma_{\delta_j}^2}} \right) \frac{(\beta_s)^{\alpha_s}}{\Gamma(\alpha_s)} (s_\delta^2)^{\alpha_s-1} e^{-\beta_s s_\delta^2} \\
&\propto (s_\delta^2)^{\alpha_s + \frac{m_1 \nu_\delta + 1}{2} - 1} \exp\left(-s_\delta^2 \left(\frac{\nu_\delta}{2} \sum_{j=1}^m I(\sigma_{\delta_j}^2 > 0) \sigma_{\delta_j}^{-2} + \beta_s \right)\right)
\end{aligned} \tag{A30}$$

i.e., a Gamma distribution with parameters $\alpha_s + \frac{m_1 \nu_\delta + 1}{2}$ and $\frac{\nu_\delta}{2} \sum_{j=1}^m I(\sigma_{\delta_j}^2 > 0) \sigma_{\delta_j}^{-2} + \beta_s$.

To sample the degrees of freedom parameter for the random SNP effects, simple Metropolis updates could be used for sampling ν_δ . For an arbitrary prior $p(\nu_\delta)$, the corresponding FCD is as follows:

$$\begin{aligned}
p(\nu_\delta | ELSE) &\propto \left(\prod_{j=1}^m I(\sigma_{\delta_j}^2 > 0) p(\sigma_{\delta_j}^2 | \nu_\delta, s_\delta^2) \right) p(\nu_\delta) \\
&= \left(\prod_{j=1}^m I(\sigma_{\delta_j}^2 > 0) \frac{\left(\frac{\nu_\delta s_\delta^2}{2} \right)^{\frac{\nu_\delta}{2}}}{\Gamma\left(\frac{\nu_\delta}{2}\right)} \sigma_{\delta_j}^{-\left(\frac{\nu_\delta}{2}+1\right)} e^{-\frac{\nu_\delta s_\delta^2}{2\sigma_{\delta_j}^2}} \right) p(\nu_\delta)
\end{aligned} \tag{A31}$$

Details on how to ν_δ can be based on a random walk Metropolis Hastings step; we have provided details on this in other non-genomic applications involving the sampling of degrees of freedom parameters (Kizilkaya and Tempelman 2005; Bello, Steibel et al. 2010).

To sample the residual variance, given a specified scaled inverted chi-square prior $p(\sigma_e^2 | \alpha_e, s_e^2) = \chi^{-2}(\alpha_e, \alpha_e s_e^2)$, the corresponding FCD of σ_e^2 can be written as follows:

$$\begin{aligned} p(\sigma_e^2 | \mathbf{y}, ELSE) & \propto \left((2\pi\sigma_e^2)^{-n/2} \right) \exp \left(-\frac{1}{2\sigma_e^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{g} - \mathbf{W}\mathbf{u})' (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{g} - \mathbf{W}\mathbf{u}) \right) \sigma_e^{-2 \left(\frac{\nu_e}{2} + 1 \right)} e^{-\frac{\nu_e s_e^2}{2\sigma_e^2}} \\ & \propto \sigma_e^{-2 \left(\frac{\nu_e + n}{2} + 1 \right)} \exp \left(-\frac{1}{2\sigma_e^2} ((\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{g} - \mathbf{W}\mathbf{u})' (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{g} - \mathbf{W}\mathbf{u}) + \nu_e s_e^2) \right) \end{aligned}$$

[A32]

In other words, [A32] is $\chi^{-2}(\nu_e + n, (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{g} - \mathbf{W}\mathbf{u})' (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{g} - \mathbf{W}\mathbf{u}) + \nu_e s_e^2)$. Note that we advocate the non-informative specifications $\nu_e = -1$ and $s_e^2 = 0$ in the paper.

To sample the polygenic variance, given a conjugate scaled inverted-chi square prior $p(\sigma_u^2 | \alpha_u, s_u^2) = \chi^{-2}(\alpha_u, \alpha_u s_u^2)$, the FCD of σ_u^2 is classically given as follows:

$$p(\sigma_u^2 | \mathbf{y}, ELSE) \propto (\sigma_u^2)^{-\left(\frac{\nu_u + q}{2} + 1\right)} \exp \left(-\frac{1}{2} \frac{(\mathbf{u}' \mathbf{A}^{-1} \mathbf{u} + \nu_u s_u^2)}{\sigma_u^2} \right) \quad [A33]$$

In other words, [A33] is $\chi^{-2}(\nu_u + q, \mathbf{u}'\mathbf{A}^{-1}\mathbf{u} + \nu_u s_u^2)$. Note that we advocate the non-informative specifications $\nu_u = -1$ and $s_u^2 = 0$ in the paper.

A2 Supplementary Figures and Tables

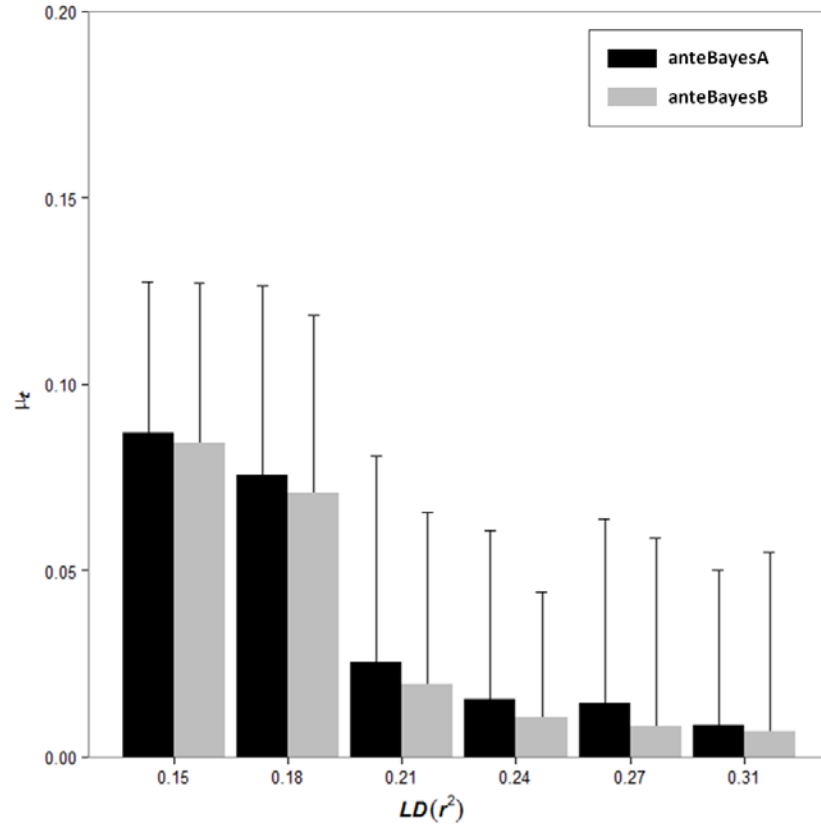


Figure A2.1: Average posterior means of μ_i and empirical standard errors across 20 replicates for each of six different LD levels using ante-BayesA and ante-BayesB. No significant differences ($P>0.01$) were determined between the competing procedures with each other or from zero at each LD level.

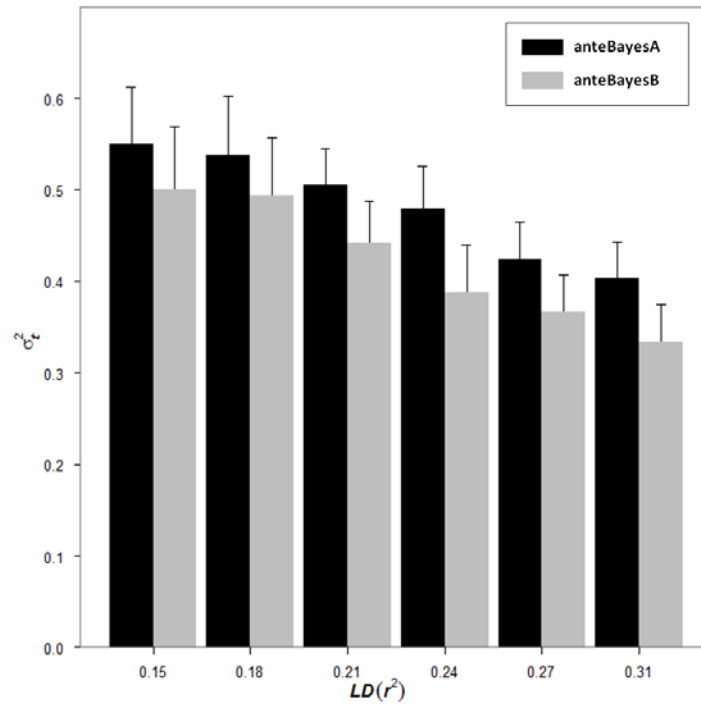


Figure A2.2: Average posterior means of σ_i^2 and empirical standard errors across 20 replicates for each of six different LD levels using ante-BayesA and ante-BayesB. No significant differences ($P > .01$) were determined between the two sets of competing procedures at each LD level.

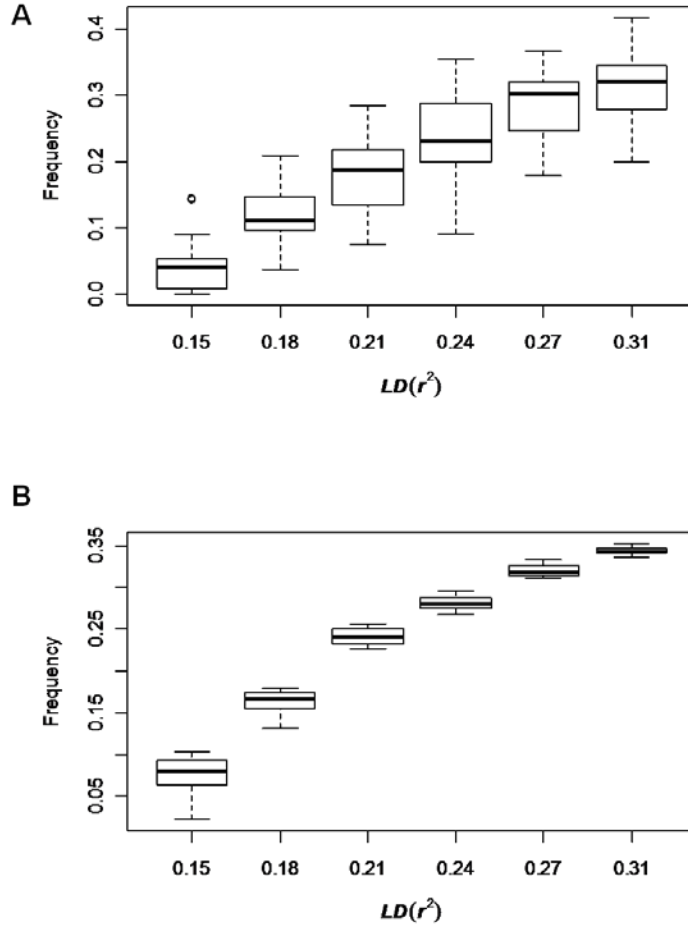


Figure A2.3: Box-plot of proportions of the absolute posterior means of elements of $\{t_{j,j-1}\}_{j=2}^m$ divided by their respective posterior standard deviations that exceeded 2 across all 20 replicates for each of six different levels of LD using ante-BayesA (A) and ante-BayesB (B).

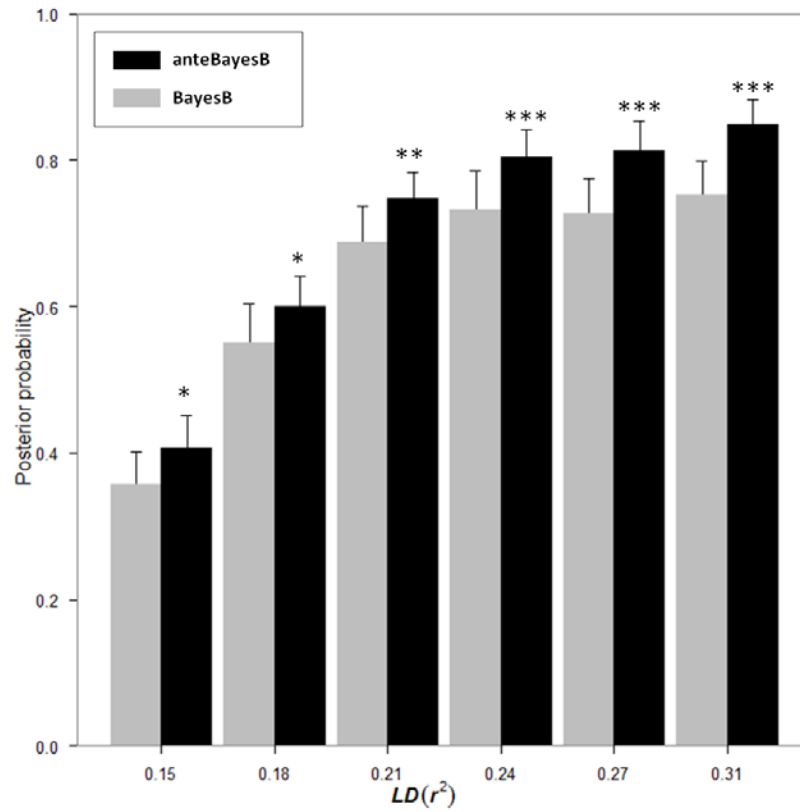
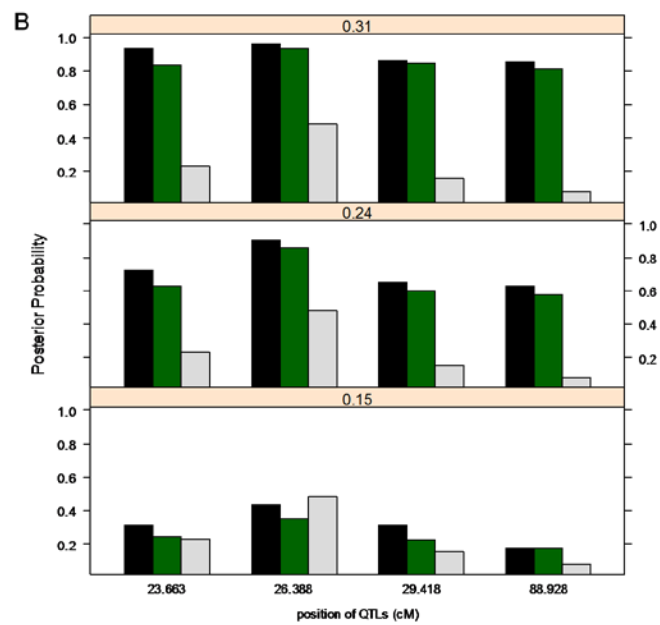
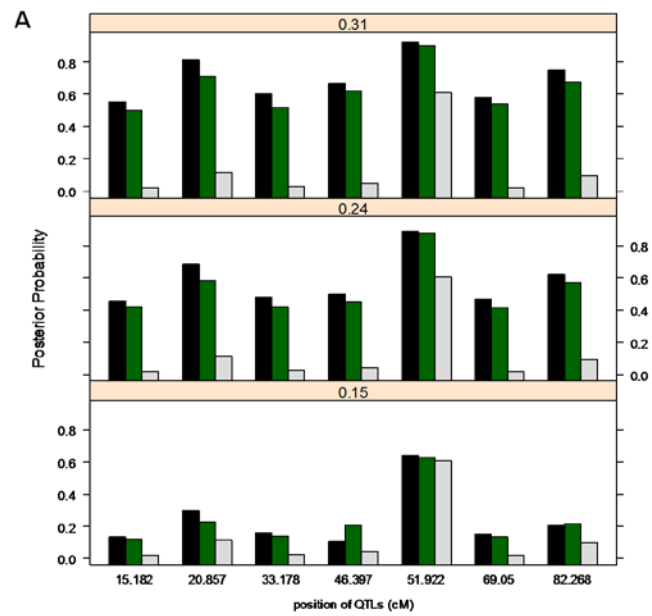


Figure A2.4: Average posterior probabilities of association for the top QTL within each of 20 replicates using BayesB and ante-BayesB for each of six different LD levels. LD-specific differences between the two methods declared significant by *($P < 0.01$), **($P < 0.001$), or ***($P < 0.0001$).



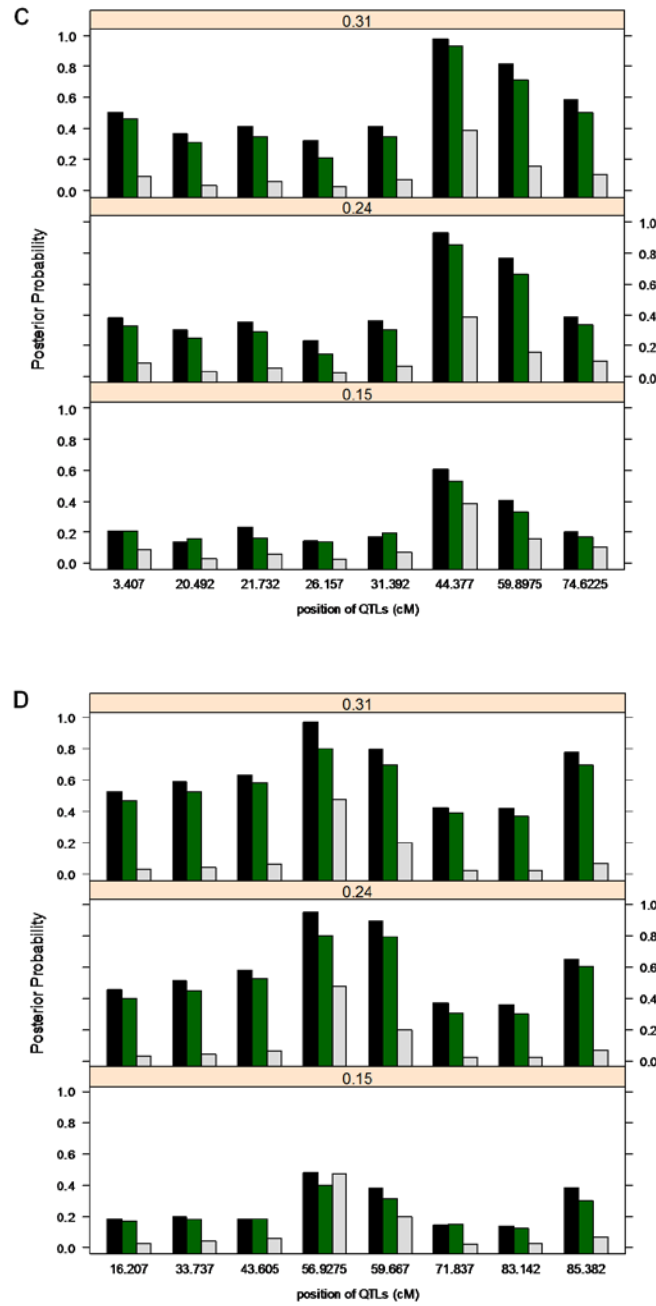


Figure A2.5: Bar plots of posterior probabilities of association of either or both of two bracketing SNP to each of the six largest QTL effects within each of the first four replicates (A,B,C,D) at the highest ($r^2=0.31$), medium ($r^2=0.24$) and lowest ($r^2=0.15$) average LD levels. Posterior probabilities using BayesB and ante-BayesB are represented by green and black bars, respectively, whereas gray bars represent the proportion of the genetic variance accounted for by the corresponding QTL. QTL location is labeled on x-axis for each replicate.

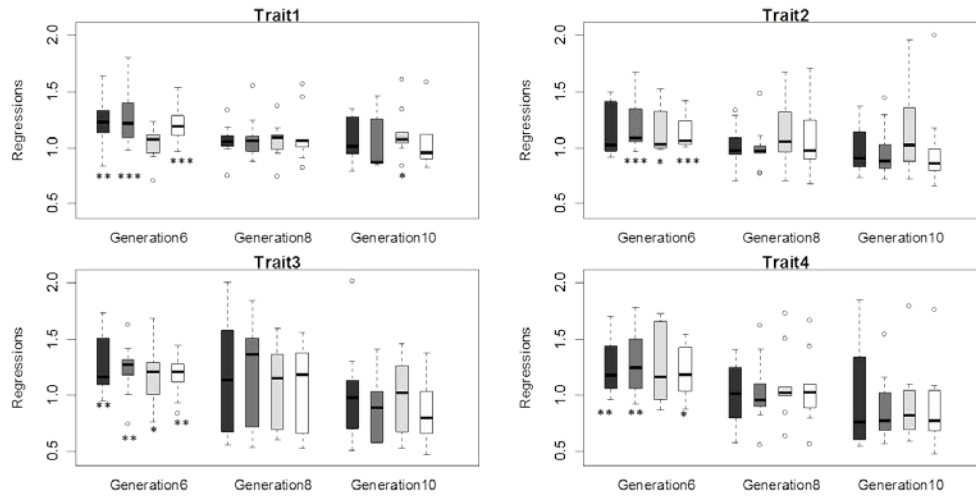


Figure A2.6: Boxplots of estimated slopes for within-replicate regressions of true breeding values on estimated breeding values across 9 replicates for four traits in Generations 6, 8 and 10 from benchmark data of Hickey and Gorjanc (2011) using anteBayesB (black), BayesB (dark gray), anteBayesA (light gray) and BayesA (white). Differences from unity indicated as significant by $*$ ($0.05 < P < .10$), $**$ ($0.01 < P < .05$) or $***$ ($P < .01$).

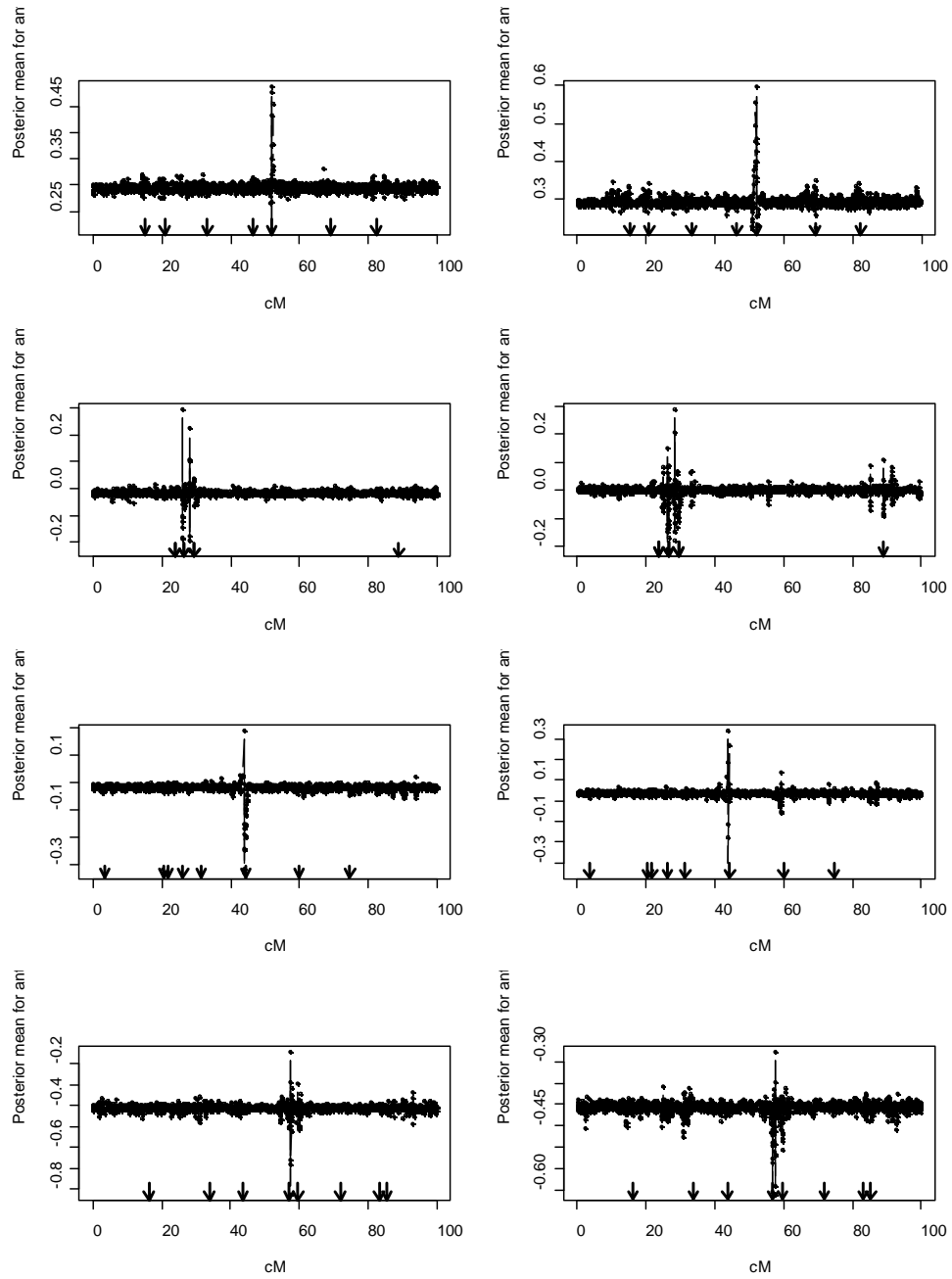


FIGURE A2.7: Posterior means of antedependence parameters $\{t_{j,j-1}\}_{j=2}^m$ versus corresponding SNP bracket location based on anteBayesA (left column) and anteBayesB (right column) for each of the first four replicates (rows 1 through 4) based on analyses using highest average marker density ($r^2 = 0.31$). Arrows denote the position for any QTL that accounted for greater than 2% of the total variance in each replicate.

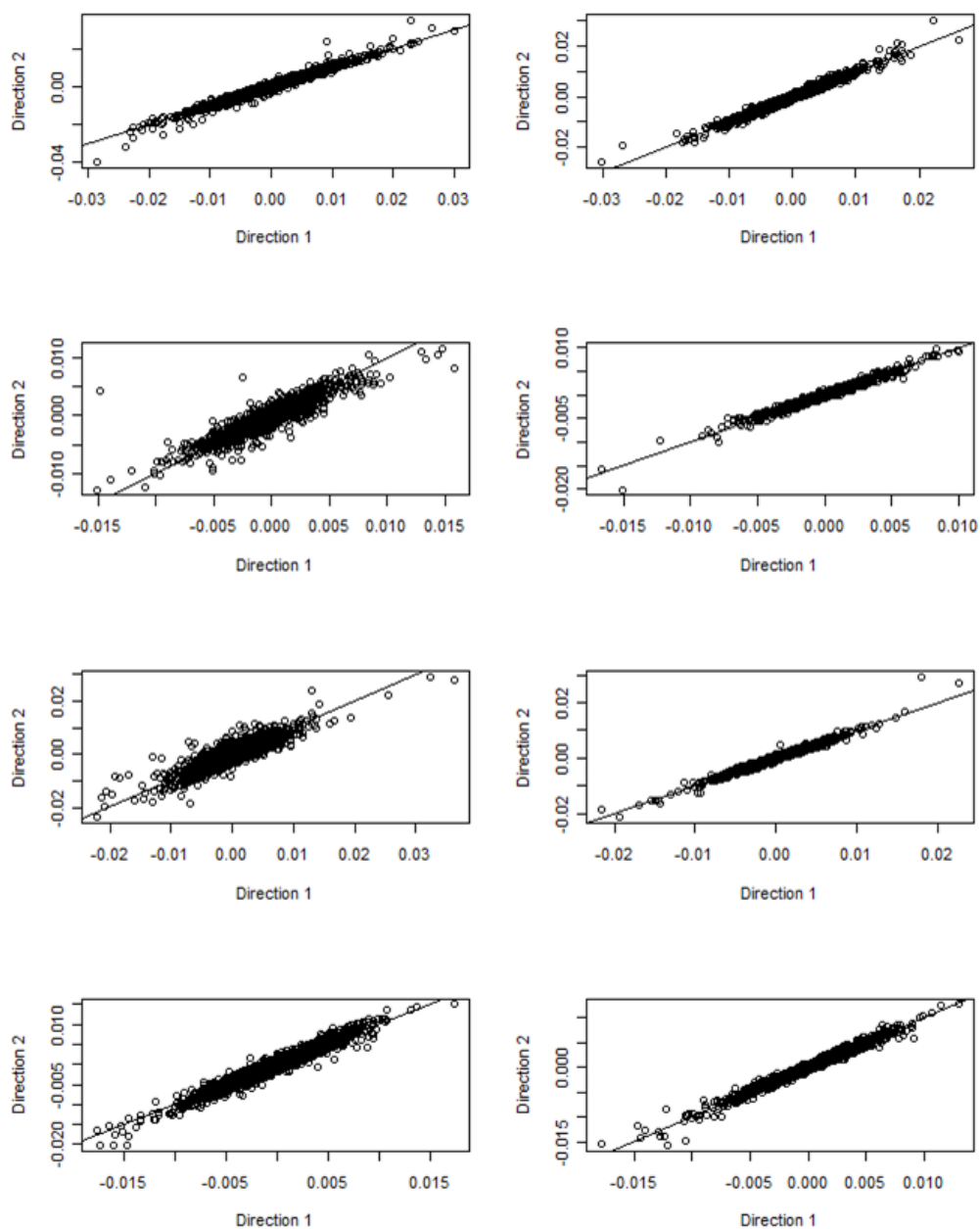


FIGURE A2.8: Posterior means of \mathbf{g} using ante-BayesA (left-column) and ante-BayesB (right-column) based on specifying antedependence in one direction along the chromosome against corresponding posterior means based on the same analyses but specifying antedependence in the opposite direction for each of the first four replicates (rows 1 through 4) and the highest average marker density ($r^2 = 0.31$). Reference lines of intercept 0 and slope 1 are superimposed.

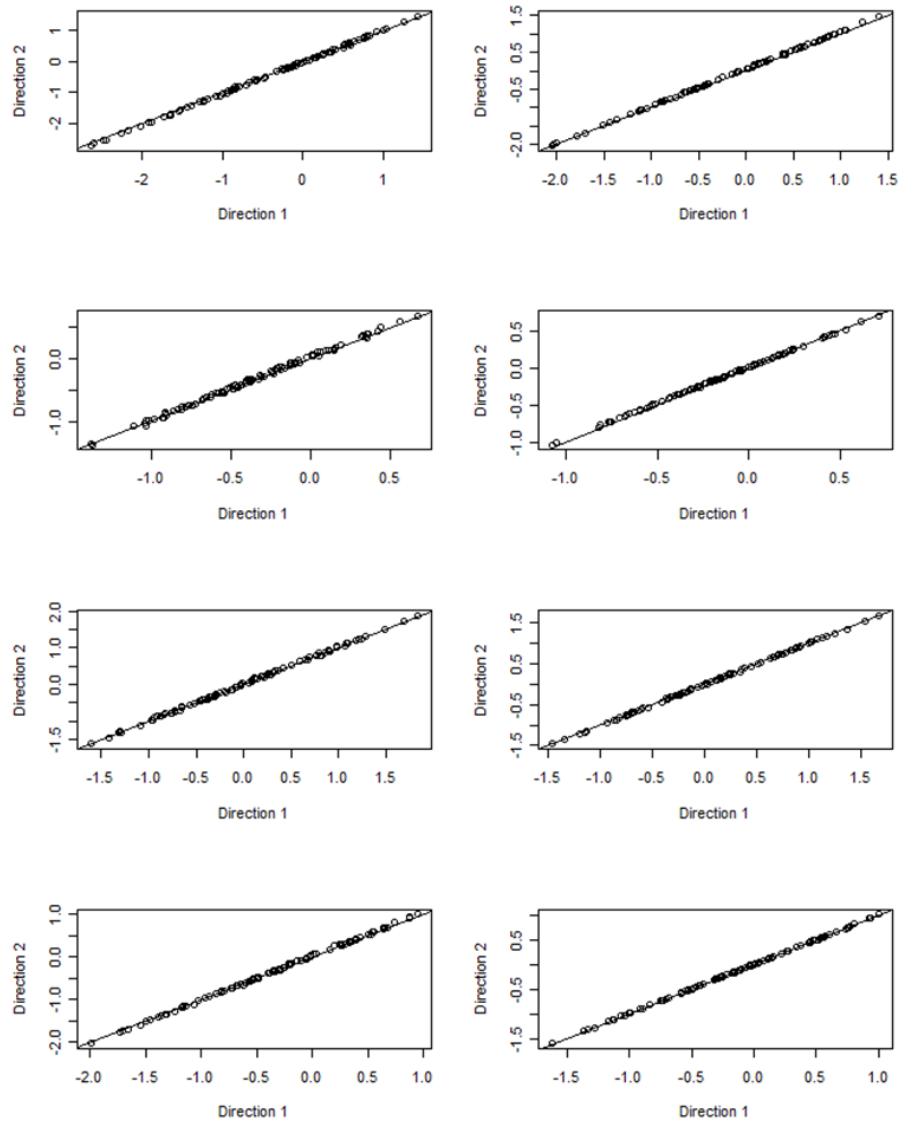


FIGURE A2.9: Posterior means of EBV using ante-BayesA (left-column) and ante-BayesB (right-column) based on specifying antedependence in one direction along the chromosome against corresponding posterior means based on the same analyses but specifying antedependence in the opposite direction for each of the first four replicates (rows 1 through 4) and the highest average marker density ($r^2 = 0.31$). Reference lines of intercept 0 and slope 1 are superimposed.

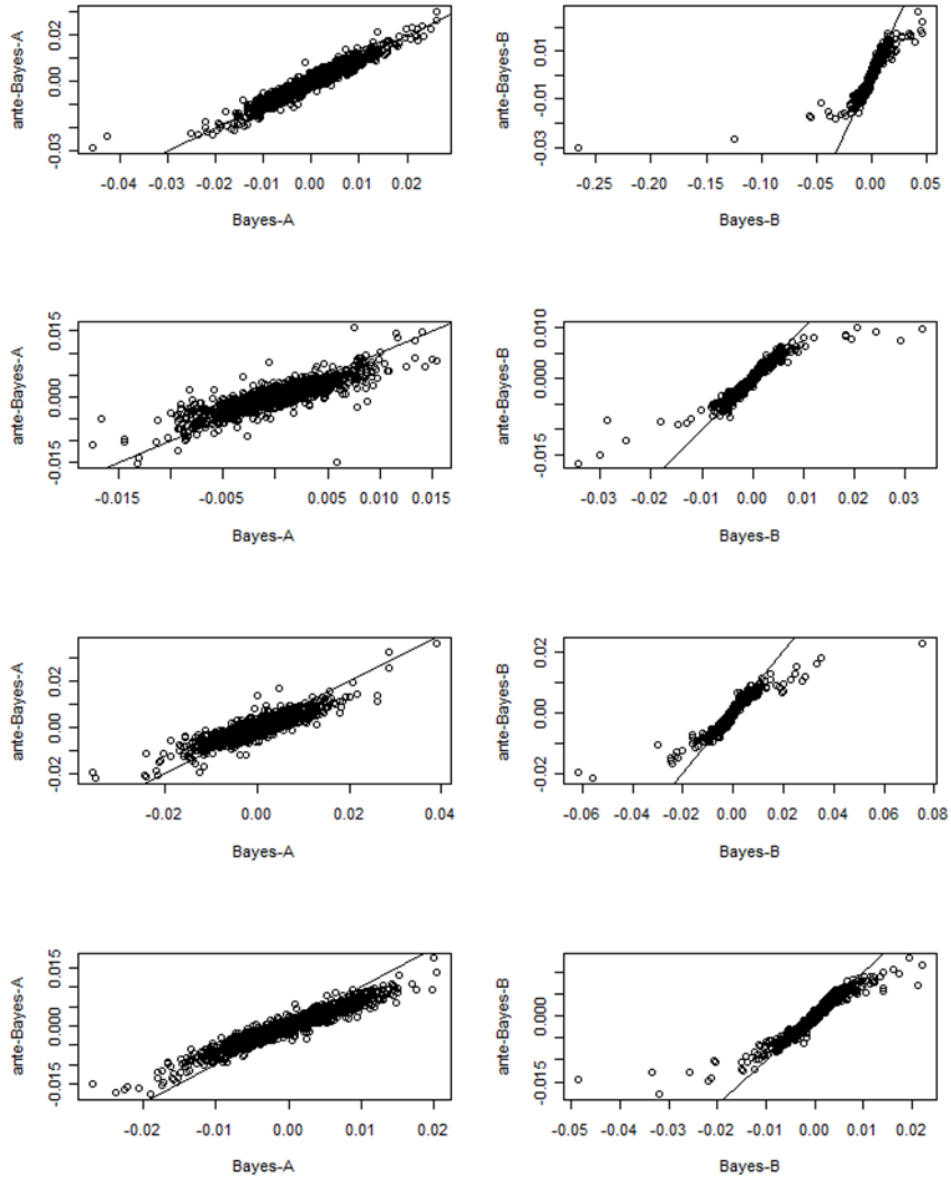


FIGURE A2.10. Posterior means of \mathbf{g} based on ante-BayesA versus BayesA (left-column) and ante-BayesB versus BayesB (right-column) for each of the first four replicates (rows 1 through 4) and the highest marker density ($r^2 = 0.31$). Reference lines of intercept 0 and slope 1 are superimposed.

Table A2.1: Average posterior means (PMEAN), posterior standard deviations (PSD), posterior medians (PMED), and effective sample size (ESS) for residual variance (σ_e^2), cage variance (σ_c^2), polygenic variance (σ_u^2) and key hyperparameters (ν_g , s_g^2 , and π_g) based on BayesA and BayesB analyses of training data subsets derived from 20 different partitions of the heterogeneous stock mice dataset. Empirical standard deviations across the 20 partitions are provided in parentheses.

| Parameter | PMEAN | PSD | PMED | ESS |
|---------------|----------------|----------------|----------------|-------------|
| BayesA | | | | |
| σ_e^2 | 0.32(0.05) | 0.15(0.003) | 0.31(0.05) | 6126(283) |
| σ_c^2 | 2.08(0.13) | 0.31(0.014) | 2.05(0.12) | 6048(247) |
| σ_u^2 | 3.21(0.15) | 0.49(0.018) | 3.18(0.13) | 5728(213) |
| ν_g | 16.12(1.83) | 23.11(2.72) | 7.62(0.56) | 285(20) |
| s_g^2 | 0.002(0.00014) | 0.0007(0.0001) | 0.002(0.00014) | 266(17) |
| BayesB | | | | |
| σ_e^2 | 0.34(0.04) | 0.13(0.002) | 0.34(0.04) | 12745(1213) |
| σ_c^2 | 2.03(0.12) | 0.35(0.011) | 2.04(0.13) | 11642(1086) |
| σ_u^2 | 3.25(0.17) | 0.53(0.021) | 3.25(0.16) | 9892(924) |
| ν_g | 19.31(1.03) | 24.14(1.09) | 9.47(0.68) | 536(65) |
| s_g^2 | 0.02(0.003) | 0.0006(0.003) | 0.02(0.0004) | 493(48) |
| π_g | 0.81(0.03) | 0.01(0.004) | 0.81(0.03) | 517(61) |

Table A2.2: Average posterior means (PMEAN), posterior standard deviations (PSD), posterior medians (PMED), and effective sample size (ESS) for residual variance (σ_e^2), cage variance (σ_c^2), polygenic variance (σ_u^2) and key hyperparameters (ν_δ , s_δ^2 , π_δ , μ_t and σ_t^2) based on ante-BayesA and ante-BayesB analyses of training data subsets derived from 20 different partitions of the heterogeneous stock mice dataset. Empirical standard deviations across the 20 partitions are provided in parentheses.

| Parameter | PMEAN | PSD | PMED | ESS |
|--------------------|---------------|-----------------|---------------|-----------|
| Ante-BayesA | | | | |
| σ_e^2 | 0.32(0.04) | 0.14(0.002) | 0.32(0.03) | 2517(134) |
| σ_c^2 | 2.01(0.13) | 0.34(0.012) | 2.02(0.13) | 2483(128) |
| σ_u^2 | 3.22(0.14) | 0.53(0.03) | 3.21(0.14) | 2357(122) |
| ν_δ | 15.54(1.32) | 21.23(1.54) | 7.27(0.46) | 185(26) |
| s_δ^2 | 0.001(0.0004) | 0.0006(0.00005) | 0.001(0.0004) | 144(20) |
| μ_t | 0.030(0.002) | 0.0139(0.004) | 0.030(0.001) | 2561(52) |
| σ_t^2 | 0.037(0.004) | 0.010(0.0003) | 0.034(0.004) | 986(43) |
| Ante-BayesB | | | | |
| σ_e^2 | 0.34(0.04) | 0.15(0.004) | 0.33(0.04) | 9512(765) |
| σ_c^2 | 2.02(0.12) | 0.39(0.017) | 2.01(0.10) | 9134(726) |
| σ_u^2 | 3.23(0.15) | 0.52(0.03) | 3.22(0.14) | 9038(689) |
| ν_δ | 18.21(1.04) | 22.46(1.22) | 9.26(0.74) | 376(33) |
| s_δ^2 | 0.14(0.05) | 0.32(0.18) | 0.07(0.15) | 259(21) |
| π_δ | 0.80(0.04) | 0.01(0.004) | 0.79(0.03) | 343(32) |
| μ_t | 0.02(0.003) | 0.011(0.002) | 0.02(0.004) | 1480(49) |
| σ_t^2 | 0.032(3e-3) | 0.05(2e-4) | 0.031(3e-3) | 638(41) |

Table A2.3: Average posterior means for residual variance (σ_e^2), polygenic variance (σ_u^2), key hyperparameters (ν_g, s_g^2 , and π_g) based on BayesA and BayesB analyses and key hyperparameters ($\nu_\delta, s_\delta^2, \pi_g, \mu_t$ and σ_t^2) based on ante-BayesB and ante-BayesA analyses for 4 different traits using Hickey and Gorjanc (2011) benchmark data. Empirical standard deviations across 9 replicates are provided in parentheses.

| | Trait 1 | Trait 2 | Trait 3 | Trait 4 |
|-------------------|------------------|------------------|------------------|------------------|
| anteBayesB | | | | |
| σ_e^2 | 0.834(0.019) | 1.348(0.125) | 0.593(0.015) | 0.939(0.038) |
| σ_u^2 | 0.131(0.017) | 0.161(0.033) | 0.101(0.010) | 0.144(0.020) |
| ν_δ | 28.043(1.841) | 15.707(2.430) | 22.941(2.158) | 17.792(2.953) |
| s_δ^2 | 7.13e-4(4.63e-5) | 1.34e-3(1.74e-4) | 4.40e-4(5.30e-5) | 9.14e-4(8.21e-5) |
| π_δ | 0.789(0.016) | 0.834(0.015) | 0.771(0.017) | 0.826(0.019) |
| μ_t | 0.032(0.011) | 0.033(0.009) | 0.011(0.015) | 0.013(0.016) |
| σ_t^2 | 0.038(0.002) | 0.034(0.001) | 0.031(0.006) | 0.046(0.009) |
| BayesB | | | | |
| σ_e^2 | 0.833(0.020) | 1.339(0.123) | 0.591(0.015) | 0.937(0.040) |
| σ_u^2 | 0.108(0.017) | 0.141(0.034) | 0.085(0.010) | 0.114(0.012) |
| ν_g | 31.643(1.623) | 23.475(3.750) | 33.536(2.501) | 27.203(3.674) |
| s_g^2 | 8.23e-4(5.60e-5) | 1.97e-3(3.21e-4) | 5.31e-4(6.75e-5) | 4.91e-3(3.71e-3) |
| π_g | 0.826(0.011) | 0.861(0.015) | 0.823(0.017) | 0.871(0.020) |
| anteBayesA | | | | |
| σ_e^2 | 0.827(0.020) | 1.339(0.124) | 0.590(0.015) | 0.935(0.040) |
| σ_u^2 | 0.157(0.030) | 0.208(0.051) | 0.143(0.018) | 0.169(0.030) |
| ν_δ | 22.556(1.192) | 12.501(2.840) | 20.909(1.194) | 14.941(3.003) |
| s_δ^2 | 6.40e-5(1.45e-5) | 6.84e-5(1.91e-5) | 2.22e-5(1.06e-5) | 4.85e-5(1.81e-5) |
| μ_t | 0.021(0.013) | 0.007(0.020) | 0.002(0.016) | 0.021(0.025) |
| σ_t^2 | 0.035(0.003) | 0.022(0.005) | 0.032(0.002) | 0.027(0.004) |
| BayesA | | | | |
| σ_e^2 | 0.829(0.019) | 1.338(0.124) | 0.592(0.015) | 0.939(0.040) |
| σ_u^2 | 0.099(0.015) | 0.123(0.034) | 0.071(0.009) | 0.097(0.012) |
| ν_g | 23.871(2.033) | 14.639(3.147) | 23.032(2.832) | 15.882(3.812) |
| s_g^2 | 1.05e-4(8.70e-6) | 1.22e-4(1.68e-5) | 6.33e-5(7.23e-6) | 8.30e-5(1.61e-5) |

Table A2.4: Average effective sample size for residual variance (σ_e^2), polygenic variance (σ_u^2), key hyperparameters (ν_g , s_g^2 , and π_g) based on BayesA and BayesB analyses and key hyperparameters (ν_δ , s_δ^2 , π_δ , μ_t and σ_t^2) based on ante-BayesB and ante-BayesA analyses for 4 different traits using Hickey and Gorjanc (2011) benchmark data. Empirical standard deviations across 9 replicates are provided in parentheses.

| | Trait 1 | Trait 2 | Trait 3 | Trait 4 |
|-------------------|------------|------------|------------|------------|
| anteBayesB | | | | |
| σ_e^2 | 4910(879) | 4883(757) | 6139(1065) | 4163(599) |
| σ_u^2 | 1423(317) | 905(131) | 1164(174) | 961(177) |
| ν_δ | 136(12) | 157(25) | 125(10) | 146(24) |
| s_δ^2 | 108(7) | 111(9) | 113(12) | 121(19) |
| π_δ | 216(19) | 236(24) | 210(20) | 249(25) |
| μ_t | 168(29) | 184(35) | 150(26) | 132(27) |
| σ_t^2 | 241(46) | 302(63) | 207(45) | 225(53) |
| BayesB | | | | |
| σ_e^2 | 3058(541) | 4814(1181) | 3462(555) | 3354(344) |
| σ_u^2 | 1056(238) | 968(284) | 993(154) | 836(85) |
| ν_g | 137(6) | 162(18) | 127(5) | 128(13) |
| s_g^2 | 138(4) | 122(28) | 119(16) | 126(16) |
| π_g | 264(36) | 347(41) | 297(26) | 275(39) |
| anteBayesA | | | | |
| σ_e^2 | 4887(1131) | 6352(1900) | 9318(1823) | 5667(1638) |
| σ_u^2 | 426(64) | 1429(1026) | 782(225) | 810(321) |
| ν_δ | 105(4) | 164(48) | 101(8) | 122(25) |
| s_δ^2 | 114(9) | 120(34) | 107(20) | 118(24) |
| μ_t | 137(10) | 110(4) | 101(18) | 107(10) |
| σ_t^2 | 191(25) | 273(66) | 201(25) | 259(51) |
| BayesA | | | | |
| σ_e^2 | 2794(342) | 4212(1053) | 2986(402) | 2541(200) |
| σ_u^2 | 677(93) | 566(111) | 642(77) | 494(38) |
| ν_g | 114(6) | 211(69) | 123(4) | 229(65) |
| s_g^2 | 150(14) | 187(58) | 145(11) | 183(42) |

APPENDIX B: Chapter 3

B1 Description of three sampling strategies

B1.1 Sampling strategy for DFMH

To sample the degrees of freedom parameter for the random SNP effects, we used a proper prior $p(\nu) \propto (\nu+1)^{-2}$ corresponding to a Uniform(0,1) prior on $(\nu+1)^{-1}$. The FCD for ν is as follows:

$$p(\nu | ELSE) \propto \left(\prod_{j=1}^m I(\sigma_{g_j}^2 > 0) p(\sigma_{g_j}^2 | \nu, s^2) \right) p(\nu)$$

$$= \left(\prod_{j=1}^m I(\sigma_{g_j}^2 > 0) \frac{\left(\frac{\nu s^2}{2} \right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)} \sigma_{g_j}^2^{-\left(\frac{\nu}{2}+1\right)} e^{-\frac{\nu s^2}{2\sigma_{g_j}^2}} \right) \frac{1}{(1+\nu)^2}$$

As this FCD is not recognizable, we could use a random walk normal MH step on

$\xi = \log(\nu)$. Note that the Jacobian from ν to ξ is $\exp(\xi)$. The corresponding FCD for ξ is as follows:

$$p(\xi | ELSE)$$

$$\propto \left(\frac{(exp(\xi)s^2 / 2)^{exp(\xi)/2}}{\Gamma(exp(\xi) / 2)} \right)^{m_1} \left(\prod_{j=1}^m I(\sigma_{g_j}^2 > 0) \sigma_{g_j}^2^{-\left(\frac{exp(\xi)}{2}+1\right)} e^{-\frac{exp(\xi)s^2}{2\sigma_{g_j}^2}} \right) \frac{1}{(1+exp(\xi))^2} exp(\xi)$$

Where $m_1 = \sum_{j=1}^m I(\sigma_{g_j}^2 > 0)$, hence

$$\begin{aligned}
& \log p(\xi | ELSE) \\
&= m_1 \left(\frac{\exp(\xi)}{2} (\xi + \log(s^2) - \log(2)) - \log \Gamma \left(\frac{\exp(\xi)}{2} \right) \right) + \\
& \sum_{j=1}^m I(\sigma_{g_j}^2 > 0) \left(- \left(\frac{\exp(\xi)}{2} + 1 \right) \log(\sigma_{g_j}^2) - \frac{\exp(\xi)s^2}{2\sigma_{g_j}^2} \right) - 2 \log(1 + \exp(\xi)) + \xi
\end{aligned}$$

Suppose the value of ξ in the current cycle i is $\xi^{[i]}$, we could propose a random walk value for $\xi^{[i+1]}$ in the next cycle from a Gaussian distribution:

$$p(\xi^*) = \frac{1}{\sqrt{2\pi}c_v} \exp \left(- \frac{(\xi^* - \xi^{[i]})^2}{2c_v^2} \right)$$

That is equivalent to generate a random variable, say δ from $N(0, c_v^2)$ and add it to $\xi^{[i]}$

to propose $\xi^* = \xi^{[i]} + \delta$. To determine the odds ratio $\alpha = \frac{p(\xi^* | ELSE)}{p(\xi^{[i]} | ELSE)}$, we evaluated

this ratio as:

$$\alpha = \exp \left(\log p(\xi^* | ELSE) - \log p(\xi^{[i]} | ELSE) \right)$$

To implement this Metropolis sampling strategy, we first generated u from a

Uniform(0,1) distribution. Then, 1) If $\alpha > 1$, accept $\xi^{[i+1]} = \xi^*$; 2) If $\alpha > u$, accept

$\xi^{[i+1]} = \xi^*$; 3) If $\alpha < u$, then set $\xi^{[i+1]} = \xi^{[i]}$. The following tuning procedure is to

determine c_v^2 :

1) For the last 10 cycles, the rate of acceptance is greater than 80%, increase c_v^2 by a factor of 1.2.

2) For the last 10 cycles, the rate of acceptance is less than 20%, decrease c_v^2 by a factor of 0.7.

3) After the burn-in, keep c_v^2 constant and monitor subsequent acceptance rates to ensure that they fall within 25 to 75%.

To sample the scale parameter for the random SNP effects, borrowing results from Yi and Xu (2008), the FCD for s^2 based on the specification of a conjugate prior $p(s^2 | \alpha_s, \beta_s) = \text{Gamma}(\alpha_s, \beta_s)$ can be written as follows:

$$\begin{aligned}
 p(s^2 | ELSE) &\propto \left(\prod_{j=1}^m I(\sigma_{g_j}^2 > 0) p(\sigma_{g_j}^2 | \nu, s^2) \right) p(s^2 | \alpha_s, \beta_s) \\
 &= \left(\prod_{j=1}^m I(\sigma_{g_j}^2 > 0) \frac{\left(\frac{s^2 \nu}{2} \right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)} \sigma_{g_j}^{-2} e^{-\frac{s^2 \nu}{2\sigma_{g_j}^2}} \right) \frac{(\beta_s)^{\alpha_s}}{\Gamma(\alpha_s)} (s^2)^{\alpha_s-1} e^{-\beta_s s^2} \\
 &\propto (s^2)^{\alpha_s + \frac{m_1 \nu}{2} - 1} \exp \left(-s^2 \left(\frac{\nu}{2} \sum_{j=1}^m I(\sigma_{g_j}^2 > 0) \sigma_{g_j}^{-2} + \beta_s \right) \right)
 \end{aligned}$$

i.e., a Gamma distribution with parameters $\alpha_s + \frac{m_1 \nu}{2}$ and $\frac{\nu}{2} \sum_{j=1}^m I(\sigma_{g_j}^2 > 0) \sigma_{g_j}^{-2} + \beta_s$.

B1.2 Sampling strategy for UNIMH

To sample the degrees of freedom parameter for the random SNP effects, the FCD for sampling ν is as follows:

$$\begin{aligned}
p(\nu | ELSE) &\propto \left(\prod_{j=1}^m I(g_j \neq 0) p(g_j | \nu_g, s_g^2) \right) p(\nu) \\
&= \left[\prod_{j=1}^m I(g_j \neq 0) \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \left(\frac{1}{\pi \nu s^2}\right)^{1/2} \left[1 + \frac{g_j^2}{\nu s^2}\right]^{-\frac{\nu+1}{2}} \right] \frac{1}{(1+\nu)^2}
\end{aligned}$$

As this FCD is not recognizable, we could use a random walk normal MH step on

$\xi = \log(\nu)$. Note that the Jacobian from ν to ξ is $\exp(\xi)$. The corresponding FCD for ξ is as follows:

$$\begin{aligned}
&p(\xi | ELSE) \\
&\propto \left(\frac{\Gamma((\exp(\xi)+1)/2)}{\Gamma(\exp(\xi)/2)} \right)^{m_1} \left(\frac{1}{\pi \exp(\xi) s^2} \right)^{m_1/2} \left(\prod_{j=1}^m I(g_j \neq 0) \left[1 + \frac{g_j^2}{\exp(\xi) s^2} \right]^{-\frac{\exp(\xi)+1}{2}} \right) \\
&\frac{1}{(1 + \exp(\xi))^2} \exp(\xi)
\end{aligned}$$

Where $m_1 = \sum_{j=1}^m I(g_j \neq 0)$, hence

$$\begin{aligned}
&\log p(\xi | ELSE) \\
&= m_1 \left(\log \Gamma\left(\frac{\exp(\xi)+1}{2}\right) - \log \Gamma\left(\frac{\exp(\xi)}{2}\right) + \frac{1}{2} \log\left(\frac{1}{\pi \exp(\xi) s^2}\right) \right) + \\
&\sum_{j=1}^m I(g_j \neq 0) \left(-\left(\frac{\exp(\xi)+1}{2}\right) \log\left(1 + \frac{g_j^2}{\exp(\xi) s^2}\right) - 2 \log(1 + \exp(\xi)) + \xi \right)
\end{aligned}$$

Suppose the value of ξ in the current cycle i is $\xi^{[i]}$, we could propose a random walk

value for $\xi^{[i+1]}$ in the next cycle from a Gaussian distribution:

$$p(\xi^*) = \frac{1}{\sqrt{2\pi}c_v} \exp\left(\frac{-(\xi^* - \xi^{[i]})^2}{2c_v^2}\right)$$

That is equivalent to generate a random variable, say δ from $N(0, c_v^2)$ and add it to $\xi^{[i]}$

to propose $\xi^* = \xi^{[i]} + \delta$. To determine the odds ratio $\alpha = \frac{p(\xi^* | ELSE)}{p(\xi^{[i]} | ELSE)}$, we evaluated

this ratio as:

$$\alpha = \exp\left(\log p(\xi^* | ELSE) - \log p(\xi^{[i]} | ELSE)\right)$$

To implement this Metropolis sampling strategy, we first generated u from a

Uniform(0,1) distribution. Then, 1) If $\alpha > 1$, accept $\xi^{[i+1]} = \xi^*$; 2) If $\alpha > u$, accept

$\xi^{[i+1]} = \xi^*$; 3) If $\alpha < u$, then set $\xi^{[i+1]} = \xi^{[i]}$. The following tuning procedure is to

determine c_v^2 :

1) For the last 10 cycles, the rate of acceptance is greater than 70%, increase c_v^2 by a factor of 1.2.

2) For the last 10 cycles, the rate of acceptance is less than 20%, decrease c_v^2 by a factor of 0.7.

3) After the burn-in, keep c_v^2 constant and monitor subsequent acceptance rates to ensure that they fall within 25 to 75%.

To sample the scale parameter for the random SNP effects, the FCD for sampling

s^2 is as follows:

$$p(s^2 | ELSE) \propto \left(\prod_{j=1}^m I(g_j \neq 0) p(g_j | \nu, s^2) \right) p(s^2 | \alpha_s, \beta_s)$$

$$= \left(\prod_{j=1}^m I(g_j \neq 0) \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \left(\frac{1}{\pi \nu s^2}\right)^{1/2} \left[1 + \frac{g_j^2}{\nu s^2}\right]^{-\frac{\nu+1}{2}} \right) \frac{(\beta_s)^{\alpha_s}}{\Gamma(\alpha_s)} (s^2)^{\alpha_s-1} e^{-\beta_s s^2}$$

Even if this FCD is recognizable, we could use a random walk normal MH step on

$\psi = \log(s^2)$. Note that the Jacobian from s^2 to ψ is $\exp(\psi)$. The corresponding FCD for

ψ is as follows:

$$p(\psi | ELSE)$$

$$\left(\prod_{j=1}^m I(g_j \neq 0) \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \left(\frac{1}{\pi \nu s^2}\right)^{1/2} \left[1 + \frac{g_j^2}{\nu s^2}\right]^{-\frac{\nu+1}{2}} \right) \frac{(\beta_s)^{\alpha_s}}{\Gamma(\alpha_s)} (s^2)^{\alpha_s-1} e^{-\beta_s s^2}$$

$$\propto \left(\frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)} \right)^{m_1} \left(\frac{1}{\pi \nu \exp(\psi)} \right)^{m_1/2} \left(\prod_{j=1}^m I(g_j \neq 0) \left[1 + \frac{g_j^2}{\nu \exp(\psi)}\right]^{-\frac{\nu+1}{2}} \right)$$

$$(\exp(\psi))^{\alpha_s-1} e^{-\beta_s \exp(\psi)} \exp(\psi)$$

Where $m_1 = \sum_{j=1}^m I(g_j \neq 0)$, hence

$$\begin{aligned}
& \log p(\psi \mid ELSE) \\
&= m_1 \left(\log \Gamma\left(\frac{\nu+1}{2}\right) - \log \Gamma\left(\frac{\nu}{2}\right) + \frac{1}{2} \log\left(\frac{1}{\pi \nu \exp(\psi)}\right) \right) + \\
& \sum_{j=1}^m I(g_j \neq 0) \left(-\left(\frac{\nu+1}{2}\right) \log\left(1 + \frac{g_j^2}{\nu \exp(\psi)}\right) \right) + (\alpha_s - 1)\psi - \beta_s \exp(\psi) + \psi
\end{aligned}$$

Suppose the value of ψ in the current cycle i is $\psi^{[i]}$, we could propose a random walk value for $\psi^{[i+1]}$ in the next cycle from a Gaussian distribution:

$$p(\psi^*) = \frac{1}{\sqrt{2\pi}c_s} \exp\left(\frac{-\left(\psi^* - \psi^{[i]}\right)^2}{2c_s^2}\right)$$

That is equivalent to generate a random variable, say δ from $N(0, c_s^2)$ and add it to $\psi^{[i]}$

to propose $\psi^* = \psi^{[i]} + \delta$. To determine the odds ratio $\alpha = \frac{p(\psi^* \mid ELSE)}{p(\psi^{[i]} \mid ELSE)}$, we evaluated

this ratio as:

$$\alpha = \exp\left(\log p(\psi^* \mid ELSE) - \log p(\psi^{[i]} \mid ELSE)\right)$$

To implement this Metropolis sampling strategy, we first generated u from a

Uniform(0,1) distribution. Then, 1) If $\alpha > 1$, accept $\psi^{[i+1]} = \psi^*$; 2) If $\alpha > u$, accept

$\psi^{[i+1]} = \psi^*$; 3) If $\alpha < u$, then set $\psi^{[i+1]} = \psi^{[i]}$. The following tuning procedure is to

determine c_s^2 :

1) For the last 10 cycles, the rate of acceptance is greater than 70%, increase c_s^2 by a factor of 1.2.

2) For the last 10 cycles, the rate of acceptance is less than 20%, decrease c_s^2 by a factor of 0.7.

3) After the burn-in, keep c_s^2 constant and monitor subsequent acceptance rates to ensure that they fall within 25 to 75%.

B1.3 Sampling strategy for BIVMH

To sample the degrees of freedom and scale parameters for the random SNP effects, we divided burn-in into four stages with equal length as follows:

In stage 1, we sample $\log(\nu)$ and $\log(s^2)$ using UNIMH (see sampling strategy 2) with fine-tuning procedure on c_ν^2 and c_s^2 , which are also the variances for the two separate Gaussian proposal densities;

In stage 2, we sample $\log(\nu)$ and $\log(s^2)$ using UNIMH with fixing c_ν^2 and c_s^2 to the values tuned from the last cycle in stage 1 and compute correlation r between samples of $\log(\nu)$ and $\log(s^2)$ within stage 2;

In stage 3, we jointly sample $\log(\nu)$ and $\log(s^2)$ using a bivariate Gaussian proposal density with variances c_ν^2 and c_s^2 based on those tuned at the end of Stage 1 and a covariance based on the correlation computed from Stage 2. Joint density for ν and s^2 is as follows:

$$\begin{aligned}
p(v, s^2 | ELSE) &\propto \left(\prod_{j=1}^m I(g_j \neq 0) p(g_j | v, s^2) \right) p(v) p(s^2 | \alpha_s, \beta_s) \\
&= \left(\prod_{j=1}^m I(g_j \neq 0) \frac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma\left(\frac{v}{2}\right)} \left(\frac{1}{\pi v s^2}\right)^{1/2} \left[1 + \frac{g_j^2}{v s^2}\right]^{-\frac{v+1}{2}} \right) \frac{1}{(1+v)^2} \frac{(\beta_s)^{\alpha_s}}{\Gamma(\alpha_s)} (s^2)^{\alpha_s-1} e^{-\beta_s s^2}
\end{aligned}$$

As this density is not recognizable, we could use a random walk normal MH step on $\xi = \log(v)$ and $\psi = \log(s^2)$. Note that the Jacobian from v to ξ is $\exp(\xi)$. Note that the Jacobian from s^2 to ψ is $\exp(\psi)$. The corresponding joint density for ξ and ψ is as follows:

$$\begin{aligned}
p(\xi, \psi | ELSE) &= \left(\frac{\Gamma((\exp(\xi)+1)/2)}{\Gamma(\exp(\xi)/2)} \right)^{m_1} \left(\frac{1}{\pi \exp(\xi) \exp(\psi)} \right)^{m_1/2} \left(\prod_{j=1}^m I(g_j \neq 0) \left[1 + \frac{g_j^2}{\exp(\xi) \exp(\psi)} \right]^{-\frac{\exp(\xi)+1}{2}} \right) \\
&\quad \frac{1}{(1+\exp(\xi))^2} (\exp(\psi))^{\alpha_s-1} e^{-\beta_s \exp(\psi)} \exp(\psi) \exp(\xi)
\end{aligned}$$

Where $m_1 = \sum_{j=1}^m I(g_j \neq 0)$, hence

$$\begin{aligned}
\log p(\xi, \psi | ELSE) &= m_1 \left(\log \Gamma\left(\frac{\exp(\xi)+1}{2}\right) - \log \Gamma\left(\frac{\exp(\xi)}{2}\right) + \frac{1}{2} \left(\frac{1}{\pi \exp(\xi) \exp(\psi)} \right) \right) + \\
&\quad \sum_{j=1}^m I(g_j \neq 0) \left(- \left(\frac{\exp(\xi)+1}{2} \right) \log \left(1 + \frac{g_j^2}{\exp(\xi) \exp(\psi)} \right) \right) - \\
&\quad 2 \log(1 + \exp(\xi)) + (\alpha_s - 1) \psi - \beta_s \exp(\psi) + \psi + \xi
\end{aligned}$$

Suppose the value of $\boldsymbol{\eta} = [\xi, \psi]'$ in the current cycle i is $\boldsymbol{\eta}^{[i]}$, we could propose a random walk value for $\boldsymbol{\eta}^{[i+1]}$ in the next cycle from a bivariate Gaussian distribution:

$$p(\boldsymbol{\eta}^*) = \frac{1}{2\pi |c_{\boldsymbol{\eta}}^2 \boldsymbol{\Sigma}|^{1/2}} \exp\left(-(\boldsymbol{\eta}^* - \boldsymbol{\eta}^{[i]})' (c_{\boldsymbol{\eta}}^2 \boldsymbol{\Sigma})^{-1} (\boldsymbol{\eta}^* - \boldsymbol{\eta}^{[i]})\right)$$

That is equivalent to generate a random variable, say $\boldsymbol{\delta}$ from $\mathbf{N}(\mathbf{0}, c_{\boldsymbol{\eta}}^2 \boldsymbol{\Sigma})$ and add it to $\boldsymbol{\eta}^{[i]}$

to propose $\boldsymbol{\eta}^* = \boldsymbol{\eta}^{[i]} + \boldsymbol{\delta}$, where $\boldsymbol{\Sigma} = \begin{bmatrix} c_v^2 & r\sqrt{c_v^2 c_s^2} \\ r\sqrt{c_v^2 c_s^2} & c_s^2 \end{bmatrix}$, c_v^2 and c_s^2 were fixed value

from last cycle in stage 1, correlation r between samples of $\log(\nu)$ and $\log(s^2)$

computed from stage 2.

To determine the odds ratio $\alpha = \frac{p(\boldsymbol{\eta}^* | ELSE)}{p(\boldsymbol{\eta}^{[i]} | ELSE)}$, we evaluated this ratio as:

$$\alpha = \exp\left(\log p(\boldsymbol{\eta}^* | ELSE) - \log p(\boldsymbol{\eta}^{[i]} | ELSE)\right)$$

To implement this Metropolis sampling strategy, we first generated u from a

Uniform(0,1) distribution. Then, 1) If $\alpha > 1$, accept $\boldsymbol{\eta}^{[i+1]} = \boldsymbol{\eta}^*$; 2) If $\alpha > u$, accept

$\boldsymbol{\eta}^{[i+1]} = \boldsymbol{\eta}^*$; 3) If $\alpha < u$, then set $\boldsymbol{\eta}^{[i+1]} = \boldsymbol{\eta}^{[i]}$. The following tuning procedure is to

determine $c_{\boldsymbol{\eta}}^2$:

1) For the last 10 cycles, the rate of acceptance is greater than 60%, increase $c_{\boldsymbol{\eta}}^2$ by a factor of 1.2.

2) For the last 10 cycles, the rate of acceptance is less than 10%, decrease c_{η}^2 by a factor of 0.7.

3) After the burn-in, keep c_{η}^2 constant and monitor subsequent acceptance rates to ensure that they fall within 25 to 75%.

In stage 4, we jointly sample $\log(\nu)$ and $\log(s^2)$ using a bivariate Gaussian proposal density with fixing value of c_{η}^2 at the end of stage 3. After burn-in, we started to save all samples on ν and s^2 using MH with the bivariate Gaussian proposal density.

B2 Supplementary tables and figures

Table B2.1: Average posterior means (PMEAN), posterior standard deviations (PSD), posterior medians (PMED), and effective sample size (ESS) for residual variance (σ_e^2), cage variance (σ_c^2), polygenic variance (σ_u^2) and key hyperparameters (ν , s^2 , and π) based on BayesA and BayesB analyses of subsets with 950 SNPs from the heterogeneous stock mice dataset.

| | DFMH | | | UNIMH | | | BIVMH | | |
|--------------|-------|-------|------|-------|-------|------|-------|-------|------|
| Parameter | PMEAN | PSD | ESS | PMEAN | PSD | ESS | PMEAN | PSD | ESS |
| BayesA | | | | | | | | | |
| σ_e^2 | 0.16 | 0.09 | 2964 | 0.16 | 0.08 | 3218 | 0.16 | 0.09 | 3240 |
| σ_c^2 | 2.02 | 0.21 | 2515 | 2.02 | 0.21 | 2797 | 2.02 | 0.21 | 2757 |
| σ_u^2 | 4.02 | 0.36 | 2295 | 3.98 | 0.36 | 2602 | 3.98 | 0.36 | 2690 |
| ν | 18.60 | 47.26 | 291 | 22.28 | 61.07 | 856 | 24.49 | 64.30 | 1092 |
| s^2 | 2e-3 | 6e-4 | 279 | 2e-3 | 6e-4 | 498 | 2e-3 | 6e-4 | 504 |
| BayesB | | | | | | | | | |
| σ_e^2 | 0.16 | 0.06 | 4918 | 0.16 | 0.06 | 6171 | 0.16 | 0.06 | 6283 |
| σ_c^2 | 1.94 | 0.21 | 4301 | 1.91 | 0.21 | 5441 | 1.91 | 0.21 | 5466 |
| σ_u^2 | 3.92 | 0.32 | 4113 | 3.91 | 0.30 | 5280 | 3.91 | 0.30 | 5274 |
| ν | 29.16 | 65.25 | 395 | 30.61 | 72.58 | 2346 | 35.80 | 81.36 | 2957 |
| s^2 | 0.02 | 5e-3 | 314 | 0.02 | 5e-3 | 782 | 0.02 | 5e-3 | 813 |
| π | 0.83 | 0.10 | 391 | 0.83 | 0.10 | 619 | 0.83 | 0.10 | 639 |

Table B2.2: Average posterior means (PMEAN), posterior standard deviations (PSD), posterior medians (PMED), and effective sample size (ESS) for residual variance (σ_e^2), cage variance (σ_c^2), polygenic variance (σ_u^2) and key hyperparameters (ν , s^2 , and π) based on BayesA and BayesB analyses of subsets with 1900 SNPs from the heterogeneous stock mice dataset.

| | DFMH | | | UNIMH | | | BIVMH | | |
|--------------|-------|-------|------|-------|-------|------|-------|-------|------|
| Parameters | PMEAN | PSD | ESS | PMEAN | PSD | ESS | PMEAN | PSD | ESS |
| BayesA | | | | | | | | | |
| σ_e^2 | 0.16 | 0.08 | 1648 | 0.16 | 0.07 | 2228 | 0.16 | 0.07 | 3195 |
| σ_c^2 | 2.03 | 0.21 | 1515 | 2.00 | 0.20 | 2016 | 2.00 | 0.21 | 2786 |
| σ_u^2 | 3.80 | 0.39 | 1314 | 3.80 | 0.38 | 1983 | 3.80 | 0.39 | 2238 |
| ν | 14.82 | 36.03 | 121 | 19.96 | 55.46 | 1075 | 22.75 | 61.18 | 1425 |
| s^2 | 8e-4 | 3e-4 | 109 | 8e-4 | 3e-4 | 303 | 8e-4 | 3e-4 | 315 |
| BayesB | | | | | | | | | |
| σ_e^2 | 0.16 | 0.07 | 4228 | 0.16 | 0.07 | 5210 | 0.16 | 0.07 | 6185 |
| σ_c^2 | 1.98 | 0.19 | 4037 | 1.98 | 0.19 | 4468 | 1.98 | 0.19 | 5891 |
| σ_u^2 | 3.68 | 0.35 | 3971 | 3.68 | 0.36 | 4300 | 3.68 | 0.34 | 5284 |
| ν | 23.57 | 52.94 | 215 | 30.50 | 72.51 | 2474 | 33.75 | 81.49 | 2948 |
| s^2 | 2e-3 | 1e-3 | 194 | 2e-3 | 1e-3 | 475 | 2e-3 | 1e-3 | 542 |
| π | 0.85 | 0.09 | 208 | 0.85 | 0.08 | 424 | 0.85 | 0.09 | 594 |

Table B2.3: Average posterior means (PMEAN), posterior standard deviations (PSD), posterior medians (PMED), and effective sample size (ESS) for residual variance (σ_e^2), cage variance (σ_c^2), polygenic variance (σ_u^2) and key hyperparameters (ν , s^2 , and π) based on BayesA and BayesB analyses of subsets with 3800 SNPs from the heterogeneous stock mice dataset.

| Parameters | DFMH | | | UNIMH | | | BIVMH | | |
|---------------|-------|-------|------|-------|-------|------|-------|-------|------|
| | PMEAN | PSD | ESS | PMEAN | PSD | ESS | PMEAN | PSD | ESS |
| BayesA | | | | | | | | | |
| σ_e^2 | 0.16 | 0.07 | 1030 | 0.16 | 0.07 | 1792 | 0.16 | 0.07 | 2962 |
| σ_c^2 | 2.01 | 0.20 | 1011 | 2.01 | 0.20 | 1565 | 2.01 | 0.20 | 2537 |
| σ_u^2 | 3.43 | 0.33 | 827 | 3.43 | 0.34 | 1529 | 3.43 | 0.34 | 1968 |
| ν | 3.14 | 0.88 | 111 | 3.14 | 2.05 | 1280 | 3.24 | 1.50 | 1339 |
| s^2 | 5e-4 | 2e-4 | 103 | 5e-4 | 2e-4 | 407 | 5e-4 | 2e-4 | 456 |
| BayesB | | | | | | | | | |
| σ_e^2 | 0.16 | 0.07 | 3579 | 0.16 | 0.07 | 4198 | 0.16 | 0.07 | 5230 |
| σ_c^2 | 1.96 | 0.19 | 3134 | 1.96 | 0.19 | 3986 | 1.96 | 0.19 | 4273 |
| σ_u^2 | 3.27 | 0.32 | 2896 | 3.27 | 0.33 | 3761 | 3.27 | 0.31 | 4158 |
| ν | 6.95 | 27.24 | 198 | 6.71 | 20.79 | 2513 | 9.05 | 30.10 | 3127 |
| s^2 | 1e-3 | 9e-4 | 163 | 1e-3 | 9e-4 | 419 | 1e-3 | 9e-4 | 489 |
| π | 0.88 | 0.10 | 194 | 0.88 | 0.09 | 405 | 0.88 | 0.09 | 512 |

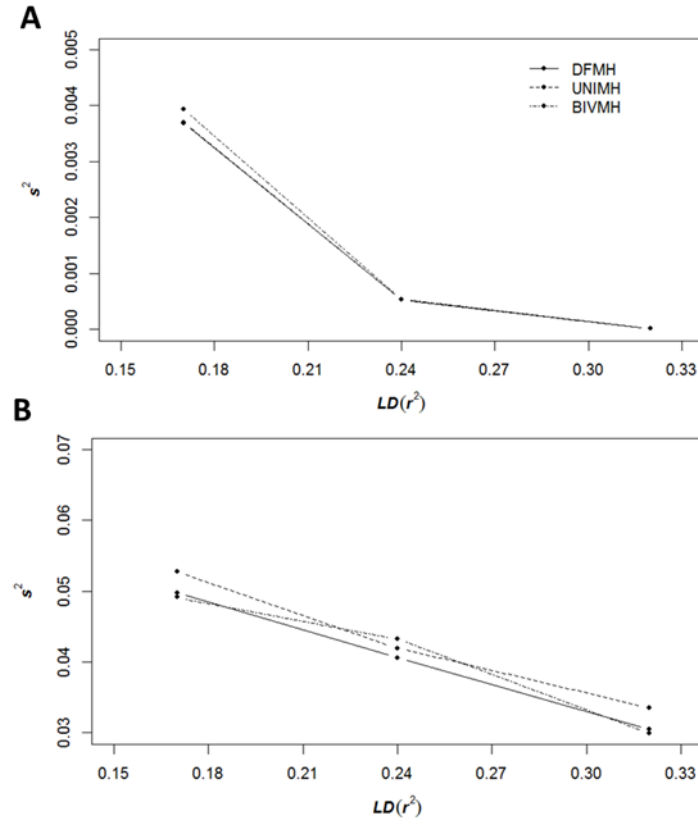


Figure B2.1: Average posterior means of s^2 (BayesA, BayesB) using DFMH, UNIMH and BIVMH across 15 replicates at LD level of 0.17, 0.24 and 0.32 comparing DFMH, UNIMH and BIVMH using BayesA model in (A) and using BayesB model in (B).

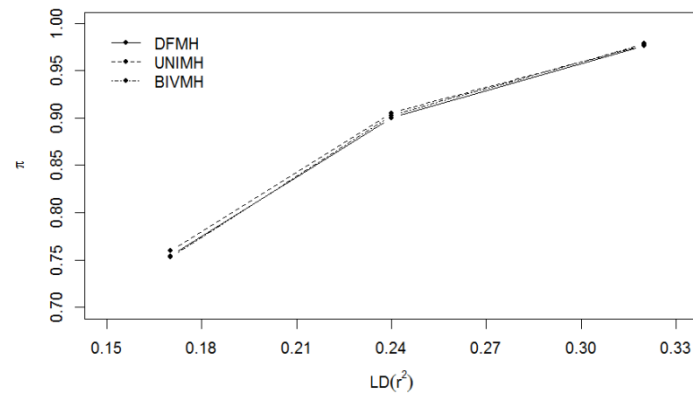


Figure B2.2: Average posterior means of π using BayesB model across 15 replicates as a function of LD levels 0.17, 0.24 and 0.32 comparing DFMH, UNIMH and BIVMH.

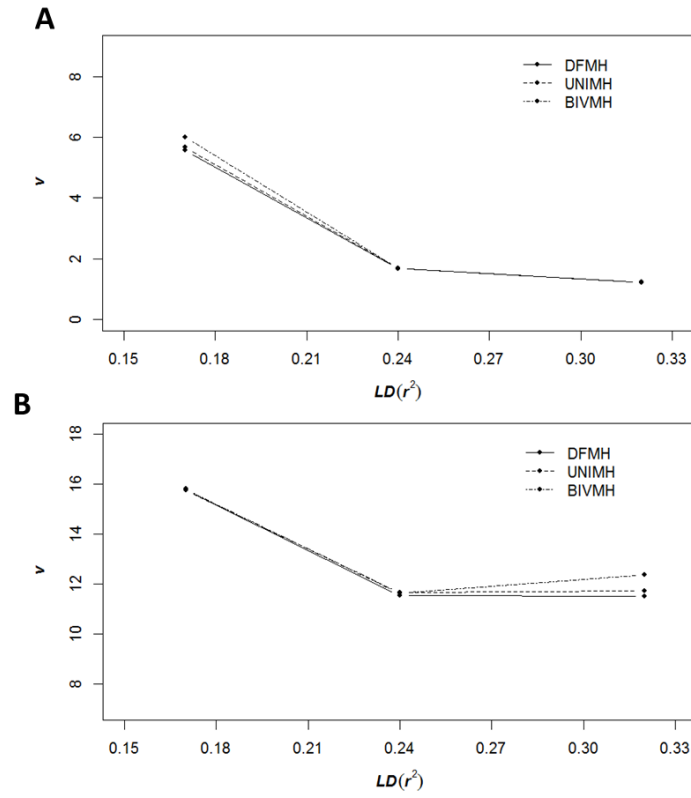


Figure B2.3: Average posterior means of ν (BayesA, BayesB) across 15 replicates for three different levels of LD comparing DFMH, UNIMH and BIVMH in BayesA model (A) and in BayesB model (B).

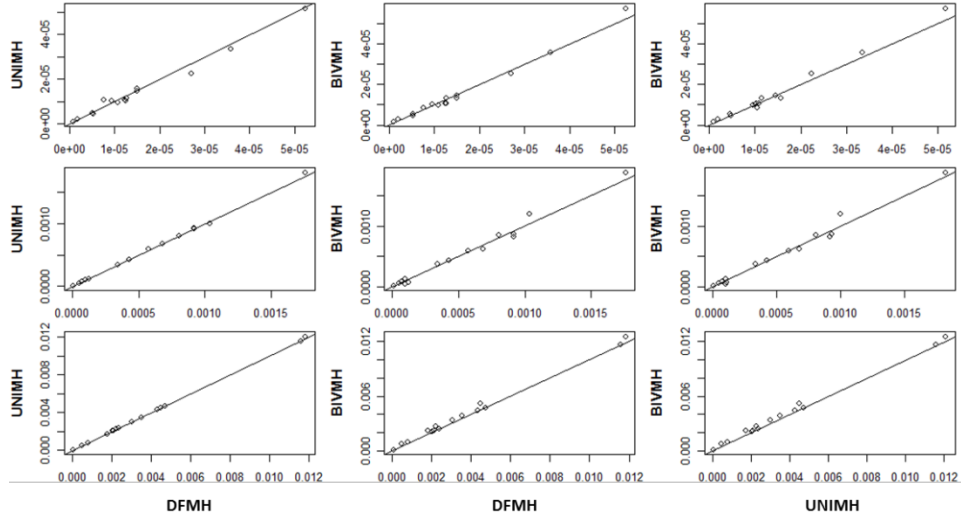


Figure B2.4: Average posterior means of s^2 in BayesA model using DFMH, UNIMH and BIVMH across 15 replicates for three LD levels of 0.17 (bottom), 0.24(middle) and 0.32 (top) comparing DFMH, UNIMH and BIVMH.

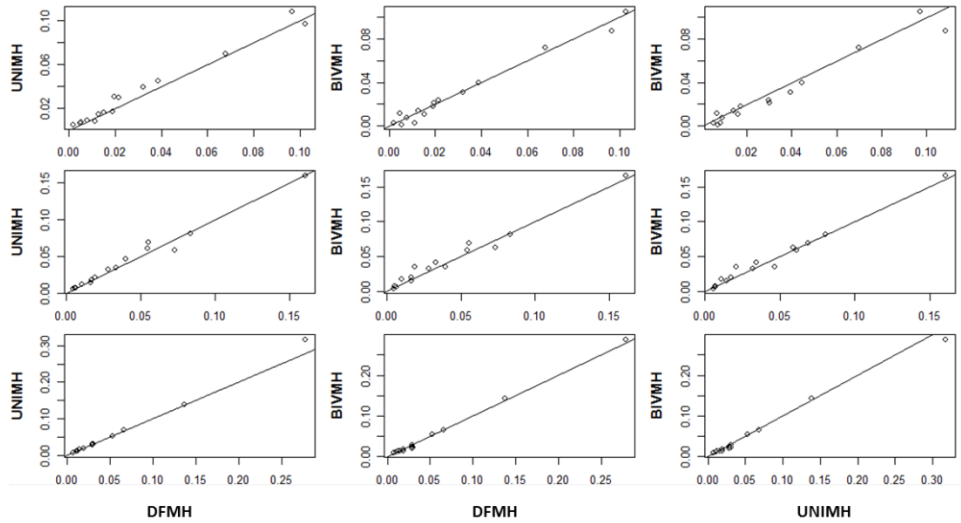


Figure B2.5: Average posterior means of s^2 in BayesB model using DFMH, UNIMH and BIVMH across 15 replicates for three LD levels of 0.17 (bottom), 0.24(middle) and 0.32 (top) comparing DFMH, UNIMH and BIVMH.

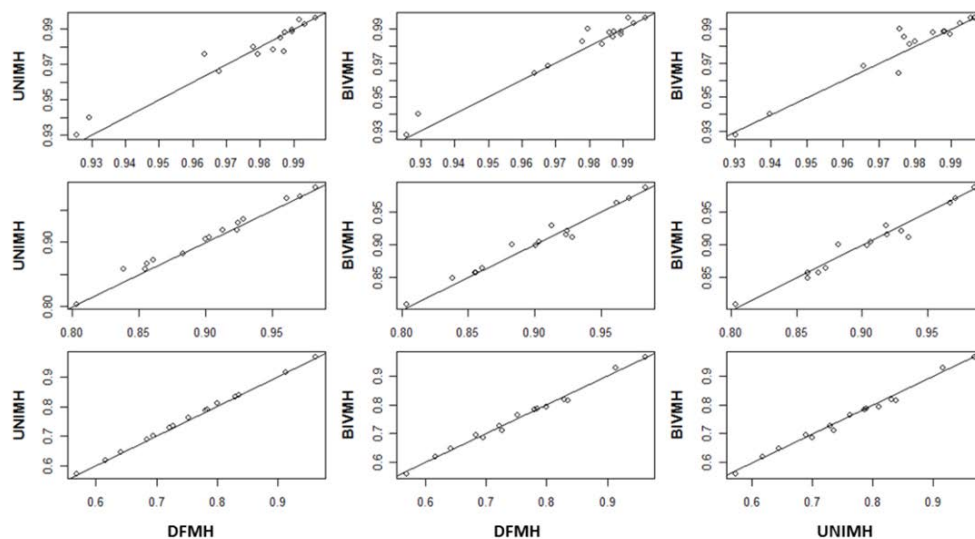


Figure B2.6: Average posterior means of π using BayesB model across 15 replicates at three LD levels of 0.17 (bottom), 0.24(middle) and 0.32 (top) comparing DFMH, UNIMH and BIVMH.

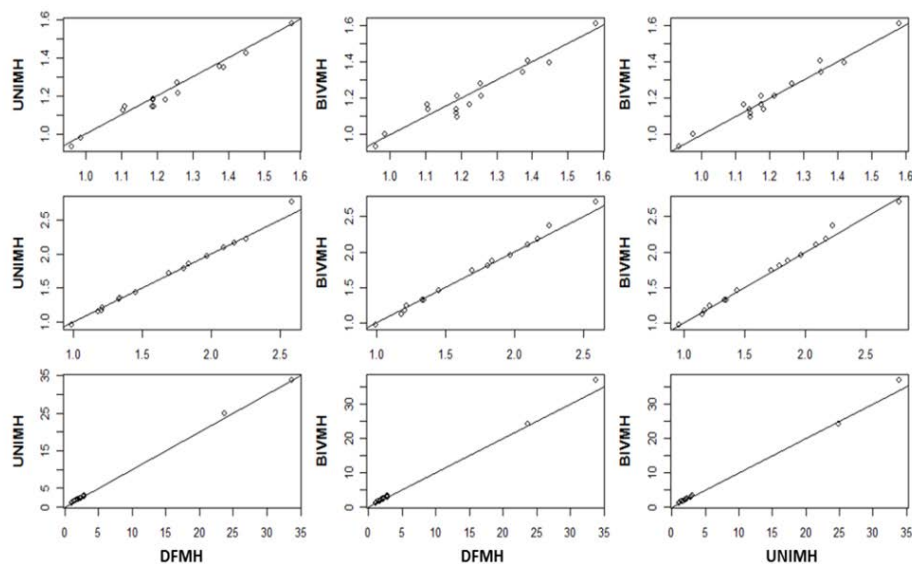


Figure B2.7: Average posterior means of ν across 15 replicates for three LD levels of 0.17 (bottom), 0.24(middle) and 0.32 (top) comparing DFMH, UNIMH and BIVMH in BayesA model.

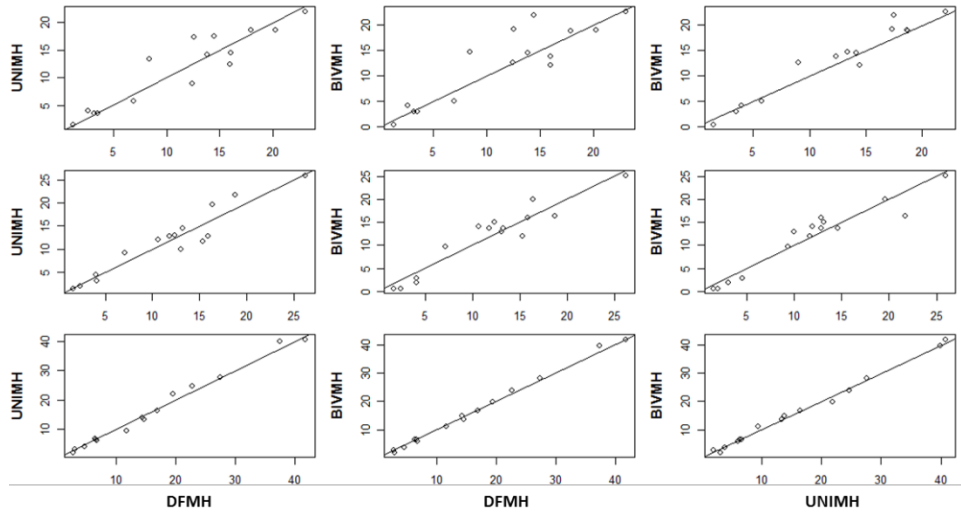


Figure B2.8: Average posterior means of ν across 15 replicates for three LD levels of 0.17 (bottom), 0.24 (middle) and 0.32 (top) comparing DFMH, UNIMH and BIVMH in BayesB model.

APPENDIX C: Chapter 4

C1 Markov Chain Monte Carlo (MCMC) Implementation Strategy for all methods

C1.1 MCMC Implementation Strategy for IW-BayesB/IW-BayesA /IW-BayesC

For the RR/RN WGP model, we know that \mathbf{y} is the $n \times 1$ vector of phenotypes for animals, $\boldsymbol{\beta}$ is the $q \times 1$ vector of fixed effects, \mathbf{g}_1 represents the $m \times 1$ vector of SNP specific random intercept effects, \mathbf{g}_2 represents the $m \times 1$ vector of SNP specific random slope effects, \mathbf{X} is the $n \times q$ design matrix for fixed effect, \mathbf{Z} is a $m \times m$ genotype matrix, \mathbf{D} is the $n \times n$ diagonal matrix with the environmental covariates on the diagonal.

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{g}_1 + \mathbf{DZ}\mathbf{g}_2 + \mathbf{e}$$

To sample location parameters $\boldsymbol{\beta}$, \mathbf{g}_1 , \mathbf{g}_2 computationally efficient, we adopted a Gauss-Seidel updating algorithm in MCMC implementation strategy (LEGARRA and MISZTAL 2008). To sample the random effects more efficiently in order to facilitate MCMC mixing, we block sampled random intercept and slope effects one SNP at a time, such that we define a vector $\mathbf{g} = [g_{11}, g_{21}, \dots, g_{1j}, g_{2j}, \dots, g_{1m}, g_{2m}]'$ where $\mathbf{g}_j = [g_{1j}, g_{2j}]'$ are random intercept and slope effects for SNP j . Based on the description of priors for the three models in materials and methods, we can write the joint posterior density for all unknown parameters as following:

$$\begin{aligned}
& p(\boldsymbol{\beta}, \mathbf{g}, \mathbf{G}, \nu_{\mathbf{g}}, \boldsymbol{\Sigma}_{\mathbf{g}}, \pi, \sigma_e^2 | \mathbf{y}) \\
& \propto p(\mathbf{y} | \boldsymbol{\beta}, \mathbf{g}, \mathbf{G}, \nu_{\mathbf{g}}, \boldsymbol{\Sigma}_{\mathbf{g}}, \pi, \sigma_e^2) p(\boldsymbol{\beta}) \\
& \left(\prod_{j=1}^m p(\mathbf{g}_j | \mathbf{G}_j) \right) \left(\prod_{j=1}^m p(\mathbf{G}_j | \nu_{\mathbf{g}}, \boldsymbol{\Sigma}_{\mathbf{g}}, \pi) \right) p(\boldsymbol{\Sigma}_{\mathbf{g}} | \nu_0, \Sigma_0) p(\nu_{\mathbf{g}}) p(\pi) p(\sigma_e^2)
\end{aligned}$$

where \mathbf{G}_j is a 2 x 2 genetic variance-covariance matrix and $\boldsymbol{\Sigma}_{\mathbf{g}}$ is a 2 x 2 scale matrix for random intercept and slope effects. In IW-BayesB ($\pi < 1$), \mathbf{G} has a mixture prior of point mass at a 2 x 2 matrix of 0's with probability $1 - \pi$ and Inverted Wishart distribution with degrees of freedom $\nu_{\mathbf{g}}$ and scale matrix $\boldsymbol{\Sigma}_{\mathbf{g}}$ with probability π . In IW-BayesA ($\pi = 1$), \mathbf{G} has an Inverted Wishart prior of degree freedom $\nu_{\mathbf{g}}$ and scale matrix $\boldsymbol{\Sigma}_{\mathbf{g}}$. In IW-BayesC ($\pi = 1$ and $\nu_{\mathbf{g}} = \infty$), all SNPs share a common genetic variance-covariance matrix $\boldsymbol{\Sigma}_{\mathbf{g}}$.

To sample fixed effects in IW-BayesA/IW-BayesB/IW-BayesC, FCD for the k^{th} element of $\boldsymbol{\beta}$ is as follows:

$$\beta_k | \mathbf{y}, \boldsymbol{\beta}_{-k}, \text{else} \sim N(\tilde{\beta}_k, \tilde{\nu}_k) \text{ where}$$

$$\begin{aligned}
\tilde{\beta}_k &= \frac{\mathbf{x}_{\cdot k}' ((\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{g}_1 - \mathbf{D}\mathbf{Z}\mathbf{g}_2) + \mathbf{x}_{\cdot k}\beta_k)}{\mathbf{x}_{\cdot k}' \mathbf{x}_{\cdot k}} \\
&= \frac{\mathbf{x}_{\cdot k}' (\mathbf{e} + \mathbf{x}_{\cdot k}\beta_k)}{\mathbf{x}_{\cdot k}' \mathbf{x}_{\cdot k}} = \frac{(\mathbf{x}_{\cdot k}' \mathbf{e} + \mathbf{x}_{\cdot k}' \mathbf{x}_{\cdot k} \beta_k)}{\mathbf{x}_{\cdot k}' \mathbf{x}_{\cdot k}} \\
&= \frac{\left(\sum_{i=1}^n x_{ik} e_i + \left(\sum_{i=1}^n x_{ik}^2 \right) \beta_k \right)}{\mathbf{x}_{\cdot k}' \mathbf{x}_{\cdot k}}
\end{aligned}$$

$$\text{and } \tilde{\nu}_k = \sigma_e^2 (\mathbf{x}_{\cdot k}' \mathbf{x}_{\cdot k})^{-1} = \sigma_e^2 \left(\sum_{i=1}^n x_{ik}^2 \right)^{-1}$$

Immediately after sampling β_k , we update the residual by $\mathbf{e} = \mathbf{e} - \mathbf{x}_{.k} (\beta_k^{[t+1]} - \beta_k^{[t]})$,

where $\beta_k^{[t+1]}$ is the sampled β_k value at cycle [t+1] and $\beta_k^{[t]}$ is the sampled β_k value at cycle [t].

To sample random intercept and slope effects in IW-BayesA, FCD for the j^{th} element of \mathbf{g} is as follows:

$$\mathbf{g}_j \mid \mathbf{y}, \boldsymbol{\beta}, \mathbf{g}_{-j}, \mathbf{G}_j, \boldsymbol{\Sigma}_g, \sigma_e^2 \sim N(\tilde{\mathbf{g}}_j, \tilde{\mathbf{V}}_{g_j})$$

Where

$$\begin{aligned} \tilde{\mathbf{g}}_j &= \left(\begin{bmatrix} \mathbf{z}'_{.j} \\ \mathbf{z}'_{.j} \mathbf{D} \end{bmatrix} \begin{bmatrix} \mathbf{z}_{.j} & \mathbf{Dz}_{.j} \end{bmatrix} + \mathbf{G}_j^{-1} \sigma_e^2 \right)^{-1} \begin{bmatrix} \mathbf{z}'_{.j} \\ \mathbf{z}'_{.j} \mathbf{D} \end{bmatrix} \left((\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Zg}_1 - \mathbf{DZg}_2) + \begin{bmatrix} \mathbf{z}_{.j} & \mathbf{Dz}_{.j} \end{bmatrix} \begin{bmatrix} g_{1j} \\ g_{2j} \end{bmatrix} \right) \\ &= \left(\begin{bmatrix} \mathbf{z}'_{.j} \mathbf{z}_{.j} & \mathbf{z}'_{.j} \mathbf{Dz}_{.j} \\ \mathbf{z}'_{.j} \mathbf{Dz}_{.j} & \mathbf{z}'_{.j} \mathbf{DDz}_{.j} \end{bmatrix} + \mathbf{G}_j^{-1} \sigma_e^2 \right)^{-1} \begin{bmatrix} \mathbf{z}'_{.j} \\ \mathbf{z}'_{.j} \mathbf{D} \end{bmatrix} (\mathbf{e} + \mathbf{z}_{.j} g_{1j} + \mathbf{Dz}_{.j} g_{2j}) \\ &= \left(\begin{bmatrix} \mathbf{z}'_{.j} \mathbf{z}_{.j} & \mathbf{z}'_{.j} \mathbf{Dz}_{.j} \\ \mathbf{z}'_{.j} \mathbf{Dz}_{.j} & \mathbf{z}'_{.j} \mathbf{DDz}_{.j} \end{bmatrix} + \mathbf{G}_j^{-1} \sigma_e^2 \right)^{-1} \begin{bmatrix} \mathbf{z}'_{.j} (\mathbf{e} + \mathbf{z}_{.j} g_{1j} + \mathbf{Dz}_{.j} g_{2j}) \\ \mathbf{z}'_{.j} \mathbf{D} (\mathbf{e} + \mathbf{z}_{.j} g_{1j} + \mathbf{Dz}_{.j} g_{2j}) \end{bmatrix} \end{aligned}$$

$$\text{and } \tilde{\mathbf{V}}_{g_j} = \left(\begin{bmatrix} \mathbf{z}'_{.j} \mathbf{z}_{.j} & \mathbf{z}'_{.j} \mathbf{Dz}_{.j} \\ \mathbf{z}'_{.j} \mathbf{Dz}_{.j} & \mathbf{z}'_{.j} \mathbf{DDz}_{.j} \end{bmatrix} \sigma_e^{-2} + \mathbf{G}_j^{-1} \right)^{-1}$$

After sampling \mathbf{g}_j , we update residual by $\mathbf{e} = \mathbf{e} - \mathbf{z}_{.j} (g_{1j}^{[t+1]} - g_{1j}^{[t]}) - \mathbf{Dz}_{.j} (g_{2j}^{[t+1]} - g_{2j}^{[t]})$,

where $g_{1j}^{[t+1]}, g_{2j}^{[t+1]}$ are the sampled g_{1j}, g_{2j} value at cycle [t+1] and $g_{1j}^{[t]}, g_{2j}^{[t]}$ are the sampled g_{1j}, g_{2j} value at cycle [t].

To sampling genetic variance-covariance matrices for random effects in IW-BayesA, given the specified inverted Wishart prior $p(\mathbf{G}_j | v_g, \Sigma_g) \propto \text{IW}(v_g, \Sigma_g)$, the FCD for \mathbf{G}_j can be written as follows:

$$\begin{aligned} p(\mathbf{G}_j | \mathbf{y}, \text{else}) &\propto p(\mathbf{g}_j | \mathbf{G}_j) p(\mathbf{G}_j | v_g, \Sigma_g) \\ &\propto |\mathbf{G}_j|^{-1/2} \exp(-0.5(\mathbf{g}_j' \mathbf{G}_j^{-1} \mathbf{g}_j)) |\mathbf{G}_j|^{-\frac{1}{2}(v_g+3)} \exp(-0.5 \text{trace}(\mathbf{G}_j^{-1} \Sigma_g (v_g - 3))) \\ &\propto |\mathbf{G}_j|^{-\frac{1}{2}((v_g+1)+3)} \exp\left(-0.5 \text{trace}\left(\mathbf{G}_j^{-1} \left[\begin{array}{cc} g_{1j}^2 & g_{1j}g_{2j} \\ g_{1j}g_{2j} & g_{2j}^2 \end{array}\right] + \Sigma_g (v_g - 3)\right)\right) \end{aligned}$$

$$\text{Hence, } \mathbf{G}_j | \mathbf{y}, \text{else} \sim \text{IW}\left(v_g + 1, \left[\begin{array}{cc} g_{1j}^2 & g_{1j}g_{2j} \\ g_{1j}g_{2j} & g_{2j}^2 \end{array}\right] + \Sigma_g (v_g - 3)\right)$$

To sample random effects and genetic variance-covariance matrices in IW-BayesB, according to the collapsed sampling strategy (Liu 1994), we jointly sampled \mathbf{g}_j and \mathbf{G}_j as adapted in Bayes B (Meuwissen et al. 2001). We first sample from

$$p(\mathbf{G}_j | \mathbf{y}, \text{ELSE except } \mathbf{g}_j) = p(\mathbf{G}_j | v_g, \Sigma_g, \pi) |\mathbf{V}_j|^{-1/2} \exp\left(-\frac{1}{2} \mathbf{y}_j^* \mathbf{V}_j^{-1} \mathbf{y}_j^*\right)$$

$$\text{where } \mathbf{y}_j^* = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}_{-j}\mathbf{g}_{1,-j} - \mathbf{D}\mathbf{Z}_{-j}\mathbf{g}_{2,-j}) = \begin{bmatrix} \mathbf{z}_{\cdot j} & \mathbf{D}\mathbf{z}_{\cdot j} \end{bmatrix} \begin{bmatrix} g_{1j} \\ g_{2j} \end{bmatrix} + \mathbf{e}$$

$$\text{and } \mathbf{V}_j = \text{var}(\mathbf{y}_j^*) = \begin{bmatrix} \mathbf{z}_{\cdot j} & \mathbf{D}\mathbf{z}_{\cdot j} \end{bmatrix} \mathbf{G}_j \begin{bmatrix} \mathbf{z}_{\cdot j}' \\ \mathbf{z}_{\cdot j}' \mathbf{D} \end{bmatrix} + \mathbf{I}\sigma_e^2$$

Since the above FCD is not recognizable, we need to adopt Metropolis-Hastings (MH) algorithm using the mixture prior $p(\mathbf{G}_j | \nu_g, \Sigma_g, \pi)$ as the candidate density. At MCMC cycle [t], we sample a candidate \mathbf{G}_j^* from the candidate density conditional on updated values for hyperparameters. We accept $\mathbf{G}_j^{[t]} = \mathbf{G}_j^*$ with the probability based on the MH acceptance ratio of $\alpha(\mathbf{G}_j^{[t-1]}, \mathbf{G}_j^*)$, where $\mathbf{G}_j^{[t-1]}$ is the value at MCMC cycle [t-1].

Adapting Meuwissen et al. (2001), the MH ratio $\alpha(\mathbf{G}_j^{[t-1]}, \mathbf{G}_j^*)$ looks as follows:

$$\alpha(\mathbf{G}_j^{[t-1]}, \mathbf{G}_j^*) = \begin{cases} \min \left(\frac{p(\mathbf{G}_j^* | \text{ELSE except } \mathbf{g}_j) p(\mathbf{G}_j^{[t-1]} | \nu_{\mathbf{G}}, \Sigma_{\mathbf{G}}, \pi)}{p(\mathbf{G}_j^{[t-1]} | \text{ELSE except } \mathbf{g}_j) p(\mathbf{G}_j^* | \nu_{\mathbf{G}}, \Sigma_{\mathbf{G}}, \pi)}, 1 \right) \\ 1, \text{otherwise} \end{cases}$$

If the candidate \mathbf{G}_j^* is rejected, we can then set $\mathbf{G}_j^{[t]} = \mathbf{G}_j^{[t-1]}$. The MH ratio can be further derived as:

$$\alpha(\mathbf{G}_j^{[t-1]}, \mathbf{G}_j^*) = \begin{cases} \min \left(\frac{|\mathbf{V}_j^*|^{-1/2} \exp\left(-\frac{1}{2} \mathbf{y}^*{}' (\mathbf{V}_j^*)^{-1} \mathbf{y}^*\right)}{|\mathbf{V}_j^{[t-1]}|^{-1/2} \exp\left(-\frac{1}{2} \mathbf{y}^*{}' (\mathbf{V}_j^{[t-1]})^{-1} \mathbf{y}^*\right)}, 1 \right) \\ 1, \text{otherwise} \end{cases}$$

It can be demonstrated that this MH ratio can be written as:

$$\alpha(\mathbf{G}_j^{[t-1]}, \mathbf{G}_j^*) = \begin{cases} \min \left(\frac{\mathbf{R}_j^{*-1/2} \exp\left(-\frac{\mathbf{r}_j^2}{2\mathbf{R}_j^*}\right)}{\mathbf{R}_j^{[t-1]-1/2} \exp\left(-\frac{\mathbf{r}_j^2}{2\mathbf{R}_j^{[t-1]}}\right)}, 1 \right) \\ 1, \text{otherwise} \end{cases}$$

where

$$\mathbf{r}_j = \begin{bmatrix} \mathbf{z}_j' \\ \mathbf{z}_j' \mathbf{D} \end{bmatrix} \mathbf{y}_j^* = \begin{bmatrix} \mathbf{z}_j' \\ \mathbf{z}_j' \mathbf{D} \end{bmatrix} \begin{bmatrix} \mathbf{z}_{.j} & \mathbf{D}\mathbf{z}_{.j} \end{bmatrix} \begin{bmatrix} g_{1j} \\ g_{2j} \end{bmatrix} + \begin{bmatrix} \mathbf{z}_j' \\ \mathbf{z}_j' \mathbf{D} \end{bmatrix} \mathbf{e}$$

and

$$\begin{aligned} \mathbf{R}_j &= \text{var}(\mathbf{r}_j) = \text{var} \left(\begin{bmatrix} \mathbf{z}_j' \\ \mathbf{z}_j' \mathbf{D} \end{bmatrix} \begin{bmatrix} \mathbf{z}_{.j} & \mathbf{D}\mathbf{z}_{.j} \end{bmatrix} \begin{bmatrix} g_{1j} \\ g_{2j} \end{bmatrix} + \begin{bmatrix} \mathbf{z}_j' \\ \mathbf{z}_j' \mathbf{D} \end{bmatrix} \mathbf{e} \right) \\ &= \begin{bmatrix} \mathbf{z}_j' \mathbf{z}_j & \mathbf{z}_j' \mathbf{D}\mathbf{z}_j \\ \mathbf{z}_j' \mathbf{D}\mathbf{z}_j & \mathbf{z}_j' \mathbf{D}\mathbf{D}\mathbf{z}_j \end{bmatrix} \mathbf{G}_j \begin{bmatrix} \mathbf{z}_j' \mathbf{z}_j & \mathbf{z}_j' \mathbf{D}\mathbf{z}_j \\ \mathbf{z}_j' \mathbf{D}\mathbf{z}_j & \mathbf{z}_j' \mathbf{D}\mathbf{D}\mathbf{z}_j \end{bmatrix} + \begin{bmatrix} \mathbf{z}_j' \mathbf{z}_j & \mathbf{z}_j' \mathbf{D}\mathbf{z}_j \\ \mathbf{z}_j' \mathbf{D}\mathbf{z}_j & \mathbf{z}_j' \mathbf{D}\mathbf{D}\mathbf{z}_j \end{bmatrix} \sigma_e^2 \end{aligned}$$

If a non-zero matrix for \mathbf{G}_j is sampled, then one can draw samples of \mathbf{g}_j using the same

full conditionals as with IW-BayesA. If either one of $g_{1j}^{[t+1]}$, $g_{1j}^{[t]}$, $g_{2j}^{[t+1]}$, $g_{2j}^{[t]}$ are non-

zero, we need to update residual $\mathbf{e} = \mathbf{e} - \mathbf{z}_{.j} (g_{1j}^{[t+1]} - g_{1j}^{[t]}) - \mathbf{D}\mathbf{z}_{.j} (g_{2j}^{[t+1]} - g_{2j}^{[t]})$ immediately

after sampling \mathbf{g}_j , where $g_{1j}^{[t+1]}$, $g_{2j}^{[t+1]}$ are the sampled g_{1j} , g_{2j} value at cycle [t+1] and

$g_{1j}^{[t]}$, $g_{2j}^{[t]}$ are the sampled g_{1j} , g_{2j} value at cycle [t].

To sample proportion of SNP markers associated with non-zero effects in IW-BayesB, with specified prior $p(\pi | \alpha_\pi, \beta_\pi) = \text{Beta}(\alpha_\pi, \beta_\pi)$, the FCD of π is based on the following,

$$p(\pi | \mathbf{y}, \text{else}) \propto \left(\prod_{j=1}^m p(\mathbf{G}_j | \mathbf{v}_g, \Sigma_g, \pi) \right) p(\pi | \alpha_\pi, \beta_\pi)$$

Let $m_1 = \sum_{j=1}^m \mathbf{I}(\mathbf{G}_j \neq \mathbf{0}_{2 \times 2})$ denote the number of non-zero vectors sampled in \mathbf{G}_j at a

particular MCMC cycle where $\mathbf{I}(\cdot)$ denotes the indicator function. Then we can write

$$p(\pi | \mathbf{y}, \text{else}) = \text{Beta}(\alpha_\pi + m_1, \beta_\pi + m - m_1).$$

To sample scale matrix for genetic variance-covariance matrices in IW-BayesA/IW-BayesB, given the specified Wishart prior for scale matrix

$p(\Sigma_g) \propto \mathbf{W}(v_0, \Sigma_0 / v_0)$, we can write FCD for Σ_g as the following,

$$\begin{aligned} p(\Sigma_g | \mathbf{y}, \text{else}) &\propto \left(\prod_{j=1}^m \mathbf{I}(\mathbf{G}_j \neq \mathbf{0}_{2 \times 2}) p(\mathbf{G}_j | v_g, \Sigma_g) \right) p(\Sigma_g) \\ &\propto \left(\prod_{j=1}^m \mathbf{I}(\mathbf{G}_j \neq \mathbf{0}_{2 \times 2}) |\Sigma_g|^{v_g/2} \exp(-0.5 \text{trace}(\mathbf{G}_j^{-1} \Sigma_g (v_g - 3))) \right) |\Sigma_g|^{\frac{1}{2}(v_0-3)} \exp(-0.5 \text{trace}(\Sigma_g \Sigma_0^{-1} v_0)) \\ &\propto |\Sigma_g|^{\frac{1}{2}(v_0+v_g m-3)} \exp \left(-0.5 \text{trace} \left(\Sigma_g \left((v_g - 3) \sum_{j=1}^m \mathbf{I}(\mathbf{G}_j \neq \mathbf{0}_{2 \times 2}) \mathbf{G}_j^{-1} + v_0 \Sigma_0^{-1} \right) \right) \right) \\ \text{Hence, } \Sigma_g | \mathbf{y}, \text{else} &\sim \mathbf{W} \left(v_0 + v_g m, \left((v_g - 3) \sum_{j=1}^m \mathbf{I}(\mathbf{G}_j \neq \mathbf{0}_{2 \times 2}) \mathbf{G}_j^{-1} + v_0 \Sigma_0^{-1} \right)^{-1} \right) \end{aligned}$$

To sample the degrees of freedom for genetic variance-covariance matrices in IW-BayesB/IW-BayesA, with a specified non-informative prior for degrees of freedom

$p(v_g) \propto 1 / (1 + v_g)^2$, we can write the FCD for v_g as the following,

$$\begin{aligned} p(v_g | \mathbf{y}, \text{else}) &\propto \left(\prod_{j=1}^m \mathbf{I}(\mathbf{G}_j \neq \mathbf{0}_{2 \times 2}) p(\mathbf{G}_j | v_g, \Sigma_g) \right) p(v_g) \\ &\propto \left(\prod_{j=1}^m \mathbf{I}(\mathbf{G}_j \neq \mathbf{0}_{2 \times 2}) \frac{|\Sigma_g|^{v_g/2}}{2^{-v_g} \Gamma_2(v_g/2)} |\mathbf{G}_j|^{-\frac{1}{2}(v_g+3)} \exp(-0.5 \text{trace}(\mathbf{G}_j^{-1} \Sigma_g (v_g - 3))) \right) \frac{1}{(1 + v_g)^2} \end{aligned}$$

We sampled v_g using a random walk Metropolis Hastings algorithm, which was described in other non-genomic applications involving the sampling of degrees of freedom parameter (Kizilkaya and Tempelman 2005; Bello, Steibel et al. 2010).

To sample random intercept and slope effects in IW-BayesC, FCD for the j^{th} element of \mathbf{g} is as follows:

$$\mathbf{g}_j | \mathbf{y}, \boldsymbol{\beta}, \mathbf{g}_{-j}, \boldsymbol{\Sigma}_g, \sigma_e^2 \sim N(\tilde{\mathbf{g}}_j, \tilde{\mathbf{V}}_{g_j})$$

Where

$$\begin{aligned} \tilde{\mathbf{g}}_j &= \left(\begin{bmatrix} \mathbf{z}'_{.j} \\ \mathbf{z}'_{.j} \mathbf{D} \end{bmatrix} \begin{bmatrix} \mathbf{z}_{.j} & \mathbf{D} \mathbf{z}_{.j} \end{bmatrix} + \boldsymbol{\Sigma}_g^{-1} \sigma_e^2 \right)^{-1} \begin{bmatrix} \mathbf{z}'_{.j} \\ \mathbf{z}'_{.j} \mathbf{D} \end{bmatrix} \left((\mathbf{y} - \mathbf{X} \boldsymbol{\beta} - \mathbf{Z} \mathbf{g}_1 - \mathbf{D} \mathbf{Z} \mathbf{g}_2) + \begin{bmatrix} \mathbf{z}_{.j} & \mathbf{D} \mathbf{z}_{.j} \end{bmatrix} \begin{bmatrix} g_{1j} \\ g_{2j} \end{bmatrix} \right) \\ &= \left(\begin{bmatrix} \mathbf{z}'_{.j} \mathbf{z}_{.j} & \mathbf{z}'_{.j} \mathbf{D} \mathbf{z}_{.j} \\ \mathbf{z}'_{.j} \mathbf{D} \mathbf{z}_{.j} & \mathbf{z}'_{.j} \mathbf{D} \mathbf{D} \mathbf{z}_{.j} \end{bmatrix} + \boldsymbol{\Sigma}_g^{-1} \sigma_e^2 \right)^{-1} \begin{bmatrix} \mathbf{z}'_{.j} \\ \mathbf{z}'_{.j} \mathbf{D} \end{bmatrix} (\mathbf{e} + \mathbf{z}_{.j} g_{1j} + \mathbf{D} \mathbf{z}_{.j} g_{2j}) \\ &= \left(\begin{bmatrix} \mathbf{z}'_{.j} \mathbf{z}_{.j} & \mathbf{z}'_{.j} \mathbf{D} \mathbf{z}_{.j} \\ \mathbf{z}'_{.j} \mathbf{D} \mathbf{z}_{.j} & \mathbf{z}'_{.j} \mathbf{D} \mathbf{D} \mathbf{z}_{.j} \end{bmatrix} + \boldsymbol{\Sigma}_g^{-1} \sigma_e^2 \right)^{-1} \begin{bmatrix} \mathbf{z}'_{.j} (\mathbf{e} + \mathbf{z}_{.j} g_{1j} + \mathbf{D} \mathbf{z}_{.j} g_{2j}) \\ \mathbf{z}'_{.j} \mathbf{D} (\mathbf{e} + \mathbf{z}_{.j} g_{1j} + \mathbf{D} \mathbf{z}_{.j} g_{2j}) \end{bmatrix} \end{aligned}$$

$$\text{and } \tilde{\mathbf{V}}_{g_j} = \left(\begin{bmatrix} \mathbf{z}'_{.j} \mathbf{z}_{.j} & \mathbf{z}'_{.j} \mathbf{D} \mathbf{z}_{.j} \\ \mathbf{z}'_{.j} \mathbf{D} \mathbf{z}_{.j} & \mathbf{z}'_{.j} \mathbf{D} \mathbf{D} \mathbf{z}_{.j} \end{bmatrix} \sigma_e^{-2} + \boldsymbol{\Sigma}_g^{-1} \right)^{-1}$$

After sampling \mathbf{g}_j , we update residual by $\mathbf{e} = \mathbf{e} - \mathbf{z}_{.j} (g_{1j}^{[t+1]} - g_{1j}^{[t]}) - \mathbf{D} \mathbf{z}_{.j} (g_{2j}^{[t+1]} - g_{2j}^{[t]})$,

where $g_{1j}^{[t+1]}, g_{2j}^{[t+1]}$ are the sampled g_{1j}, g_{2j} value at cycle [t+1] and $g_{1j}^{[t]}, g_{2j}^{[t]}$ are the sampled g_{1j}, g_{2j} value at cycle [t].

To sample genetic variance-covariance matrix for random effects in IW-BayesC, from the joint posterior density, we can derive FCD for Σ_g , which has an inverted Wishart prior $p(\Sigma_g | v_0, \Sigma_0) \propto \text{IW}(v_0, \Sigma_0)$.

$$\begin{aligned} p(\Sigma_g | \mathbf{y}, \text{else}) &\propto p(\mathbf{g} | \Sigma_g) p(\Sigma_g | v_0, \Sigma_0) \\ &\propto |\Sigma_g \otimes \mathbf{I}_{m \times m}|^{-1/2} \exp\left(-0.5 \left(\begin{bmatrix} \mathbf{g}'_1 & \mathbf{g}'_2 \end{bmatrix} [\Sigma_g^{-1} \otimes \mathbf{I}_{m \times m}] \begin{bmatrix} \mathbf{g}_1 \\ \mathbf{g}_2 \end{bmatrix} \right)\right) |\Sigma_g|^{\frac{1}{2}(v_0+3)} \exp(-0.5 \text{trace}(\Sigma_g^{-1} \Sigma_0)) \\ &\propto |\Sigma_g|^{\frac{1}{2}(m+v_0+3)} \exp(-0.5 \text{trace}(\Sigma_g^{-1} (\mathbf{S}_g + \Sigma_0))) \end{aligned}$$

$$\text{where } \mathbf{S}_g = \begin{bmatrix} \mathbf{g}'_1 \mathbf{g}_1 & \mathbf{g}'_1 \mathbf{g}_2 \\ \mathbf{g}'_2 \mathbf{g}_1 & \mathbf{g}'_2 \mathbf{g}_2 \end{bmatrix}$$

$$\text{i.e. } p(\Sigma_g | \mathbf{y}, \text{else}) \sim \text{IW}(v_0 + m, \mathbf{S}_g + \Sigma_0)$$

To sample residual variance in IW-BayesA/IW-BayesB/IW-BayesC, given a specified scaled inverted Chi-square prior $p(\sigma_e^2 | \nu_e, S_e) = \chi^{-2}(\nu_e, \nu_e S_e)$, we can write the FCD for σ_e^2 as follows,

$$\begin{aligned} p(\sigma_e^2 | \mathbf{y}, \text{else}) &\propto (2\pi\sigma_e^2)^{-n/2} \exp\left(-\frac{1}{2\sigma_e^2} \mathbf{e}' \mathbf{e}\right) \sigma_e^{2-\left(\frac{\nu_e}{2}+1\right)} e^{-\frac{\nu_e S_e}{2\sigma_e^2}} \\ &\propto \sigma_e^{2-\left(\frac{n+\nu_e}{2}+1\right)} \exp\left(-\frac{1}{2\sigma_e^2} (\mathbf{e}' \mathbf{e} + \nu_e S_e)\right) \end{aligned}$$

$$\text{where } \mathbf{e} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{g}_1 - \mathbf{D}\mathbf{Z}\mathbf{g}_2$$

$$\text{i.e. } p(\sigma_e^2 | \mathbf{y}, \text{else}) \sim \chi^{-2}(df = n + \nu_e, scale = \mathbf{e}' \mathbf{e} + \nu_e S_e)$$

C1.2 MCMC Implementation Strategy for CD-BayesB/CD-BayesA

As presented in the Materials and Methods, a square root free Cholesky decomposition can be applied to the genetic variance-covariance matrix. Let $\mathbf{g}_2 = \mathbf{\Psi}\mathbf{g}_1 + \mathbf{g}_{2||}$ where $\mathbf{g}_{2||}$ is the vector of SNP specific environmental slope effects conditional on intercept effects and $\mathbf{\Psi} = \text{diag}\{\phi_j\}_{j=1}^m$ represents a diagonal matrix of SNP-specific associations between intercept and slope effects. The RR/RN WGP model can be rewritten as:

$$\begin{aligned}\mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{g}_1 + \mathbf{DZ}(\mathbf{\Psi}\mathbf{g}_1 + \mathbf{g}_{2||}) + \mathbf{e} \\ &= \mathbf{X}\boldsymbol{\beta} + (\mathbf{Z} + \mathbf{DZ}\mathbf{\Psi})\mathbf{g}_1 + \mathbf{DZ}\mathbf{g}_{2||} + \mathbf{e}\end{aligned}$$

Based on the description of priors for the two CD models in materials and methods, we can write down the joint posterior density for all unknown parameters:

$$\begin{aligned}& p(\boldsymbol{\beta}, \mathbf{g}_1, \mathbf{g}_{2||}, \boldsymbol{\sigma}_{\mathbf{g}_1}, \boldsymbol{\sigma}_{\mathbf{g}_{2||}}, \boldsymbol{\phi}, \mu_\phi, \sigma_\phi^2, v_1, v_2, s_1^2, s_2^2, \pi_1, \pi_2, \sigma_e^2 | \mathbf{y}) \\ & \propto p(\mathbf{y} | \boldsymbol{\beta}, \mathbf{g}_1, \mathbf{g}_{2||}, \boldsymbol{\sigma}_{\mathbf{g}_1}, \boldsymbol{\sigma}_{\mathbf{g}_{2||}}, \boldsymbol{\phi}, \mu_\phi, \sigma_\phi^2, v_1, v_2, s_1^2, s_2^2, \pi_1, \pi_2, \sigma_e^2) p(\boldsymbol{\beta}) \\ & \left(\prod_{j=1}^m p(g_{1j} | \sigma_{\mathbf{g}_{1j}}^2) \right) \left(\prod_{j=1}^m p(\sigma_{\mathbf{g}_{1j}}^2 | v_1, s_1^2, \pi_1) \right) \left(\prod_{j=1}^m p(g_{2||j} | \sigma_{\mathbf{g}_{2||j}}^2) \right) \left(\prod_{j=1}^m p(\sigma_{\mathbf{g}_{2||j}}^2 | v_2, s_2^2, \pi_2) \right) \\ & \left(\prod_{j=1}^m p(\phi_j | \mu_\phi, \sigma_\phi^2) \right) p(\mu_\phi) p(\sigma_\phi^2) p(v_1) p(v_2) p(s_1^2) p(s_2^2) p(\pi_1) p(\pi_2) p(\sigma_e^2)\end{aligned}$$

Where $\boldsymbol{\sigma}_{\mathbf{g}_1} = [\sigma_{g_{11}}^2 \quad \sigma_{g_{12}}^2 \quad \sigma_{g_{13}}^2 \quad \dots \quad \sigma_{g_{1m}}^2]'$ is a vector of SNP specific variances for

intercept, $\boldsymbol{\sigma}_{\mathbf{g}_{2||}} = [\sigma_{g_{2||,1}}^2 \quad \sigma_{g_{2||,2}}^2 \quad \sigma_{g_{2||,3}}^2 \quad \dots \quad \sigma_{g_{2||,m}}^2]'$ is a vector of SNP specific variances for

slope conditional on intercept, $\boldsymbol{\phi} = [\phi_1 \quad \phi_2 \quad \dots \quad \phi_m]'$ is a vector of SNP specific association

parameters between intercept and slope. In CD-BayesB ($\pi_1 < 1$ and $\pi_2 < 1$), $\sigma_{\mathbf{g}_{1j}}^2$ ($\sigma_{\mathbf{g}_{2||j}}^2$) has a mixture prior of point mass at zero with probability $1 - \pi_1$ ($1 - \pi_2$) and scaled inverted Chi-square distribution of degree freedom ν_1 (ν_2) and scale parameter s_1^2 (s_2^2) with probability π_1 (π_2). In CD-BayesA ($\pi_1 = 1$ and $\pi_2 = 1$), $\sigma_{\mathbf{g}_{1j}}^2$ ($\sigma_{\mathbf{g}_{2||j}}^2$) has a scaled inverted Chi-square distribution of degree freedom ν_1 (ν_2) and scale parameter s_1^2 (s_2^2).

To sample all fixed effects in CD-BayesA/CD-BayesB, FCD for the k^{th} element of $\boldsymbol{\beta}$ is as follows:

$$\beta_k | \mathbf{y}, \boldsymbol{\beta}_{-k}, \text{else} \sim N(\tilde{\beta}_k, \tilde{v}_{\beta k})$$

with

$$\begin{aligned} \tilde{\beta}_k &= \frac{\mathbf{x}_{\cdot k}' \left((\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - (\mathbf{Z} + \mathbf{DZ}\Psi)\mathbf{g}_1 - \mathbf{DZ}\mathbf{g}_{2||}) + \mathbf{x}_{\cdot k}\beta_k \right)}{\mathbf{x}_{\cdot k}'\mathbf{x}_{\cdot k}} \\ &= \frac{\mathbf{x}_{\cdot k}'(\mathbf{e} + \mathbf{x}_{\cdot k}\beta_k)}{\mathbf{x}_{\cdot k}'\mathbf{x}_{\cdot k}} = \frac{(\mathbf{x}_{\cdot k}'\mathbf{e} + \mathbf{x}_{\cdot k}'\mathbf{x}_{\cdot k}\beta_k)}{\mathbf{x}_{\cdot k}'\mathbf{x}_{\cdot k}} \\ &= \frac{\left(\sum_{i=1}^n x_{ik}e_i + \left(\sum_{i=1}^n x_{ik}^2 \right) \beta_k \right)}{\mathbf{x}_{\cdot k}'\mathbf{x}_{\cdot k}} \end{aligned}$$

$$\text{and } \tilde{v}_{\beta k} = \sigma_e^2 (\mathbf{x}_{\cdot k}'\mathbf{x}_{\cdot k})^{-1} = \sigma_e^2 \left(\sum_{i=1}^n x_{ik}^2 \right)^{-1}$$

Immediately after sampling β_k , we update the residual by $\mathbf{e} = \mathbf{e} - \mathbf{x}_{\cdot k}(\beta_k^{[t+1]} - \beta_k^{[t]})$,

where $\beta_k^{[t+1]}$ is the sampled β_k value at cycle $[t+1]$ and $\beta_k^{[t]}$ is the sampled β_k value at cycle $[t]$.

To sample random intercept effects in CD-BayesA, with specified prior on g_{1j} ,

let's define $\mathbf{z}_{\cdot g_{1j}}^* = \begin{bmatrix} z_{1j}(1+d_1\phi_j) & z_{2j}(1+d_2\phi_j) & z_{3j}(1+d_3\phi_j) & \cdots & z_{nj}(1+d_n\phi_j) \end{bmatrix}'$ is

column j of $\mathbf{Z} + \mathbf{DZ}\Psi$. The FCD can be written as follows:

$$g_{1j} | \mathbf{y}, \boldsymbol{\beta}, \mathbf{g}_{-1j}, \mathbf{g}_{2||}, \left\{ \sigma_{g_{1j}}^2 \right\}_{j=1}^m, \left\{ \sigma_{g_{2||j}}^2 \right\}_{j=1}^m, \sigma_e^2, \left\{ \phi_j \right\}_{j=1}^m \sim N\left(\tilde{g}_{1j}, \tilde{v}_{g_{1j}} \right)$$

where

$$\begin{aligned} \tilde{g}_{1j} &= \frac{\mathbf{z}_{\cdot g_{1j}}^* \left((\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) - (\mathbf{Z} + \mathbf{DZ}\Psi)\mathbf{g}_1 - \mathbf{DZ}\mathbf{g}_{2||} \right) + \mathbf{z}_{\cdot g_{1j}}^* g_{1j}}{\mathbf{z}_{\cdot g_{1j}}^* \mathbf{z}_{\cdot g_{1j}}^* + \sigma_e^2 \sigma_{g_{1j}}^{-2}} \\ &= \frac{\mathbf{z}_{\cdot g_{1j}}^* (\mathbf{e} + \mathbf{z}_{\cdot g_{1j}}^* g_{1j})}{\mathbf{z}_{\cdot g_{1j}}^* \mathbf{z}_{\cdot g_{1j}}^* + \sigma_e^2 \sigma_{g_{1j}}^{-2}} = \frac{\mathbf{z}_{\cdot g_{1j}}^* \mathbf{e} + \mathbf{z}_{\cdot g_{1j}}^* \mathbf{z}_{\cdot g_{1j}}^* g_{1j}}{\mathbf{z}_{\cdot g_{1j}}^* \mathbf{z}_{\cdot g_{1j}}^* + \sigma_e^2 \sigma_{g_{1j}}^{-2}} \\ &= \frac{\sum_{i=1}^n z_{ij}(1+d_i\phi_j)e_i + \left(\sum_{i=1}^n z_{ij}^2(1+d_i\phi_j)^2 \right) g_{1j}}{\left(\sum_{i=1}^n z_{ij}^2(1+d_i\phi_j)^2 + \sigma_e^2 \sigma_{g_{1j}}^{-2} \right)} \end{aligned}$$

$$\text{and } \tilde{v}_{g_{1j}} = \left(\sigma_e^{-2} \sum_{i=1}^n \left(z_{ij}(1+d_i\phi_j) \right)^2 + \sigma_{g_{1j}}^{-2} \right)^{-1}$$

Immediately after sampling g_{1j} , we update the residual by $\mathbf{e} = \mathbf{e} - \mathbf{z}_{\cdot g_{1j}}^* \left(g_{1j}^{[t+1]} - g_{1j}^{[t]} \right)$,

where $g_{1j}^{[t+1]}$ is the sampled g_{1j} value at cycle $[t+1]$ and $g_{1j}^{[t]}$ is the sampled g_{1j} value at cycle $[t]$.

To sample random slope effects conditional on intercept in CD-BayesA, with

specified prior on $g_{2||j}$, let's define $\mathbf{z}_{\cdot g_{2||j}}^* = \begin{bmatrix} z_{1j}d_1 & z_{2j}d_2 & z_{3j}d_3 & \cdots & z_{nj}d_n \end{bmatrix}'$ is the j th

column of \mathbf{DZ} . The FCD can be written as follows:

$$g_{2||j} \mid \mathbf{y}, \boldsymbol{\beta}, \mathbf{g}_{-2||j}, \mathbf{g}_1, \left\{ \sigma_{g_{1j}}^2 \right\}_{j=1}^m, \left\{ \sigma_{g_{2||j}}^2 \right\}_{j=1}^m, \sigma_e^2, \left\{ \phi_j \right\}_{j=1}^m \sim N\left(\tilde{g}_{2||j}, \tilde{v}_{g_{2||j}} \right)$$

where

$$\begin{aligned} \tilde{g}_{2||j} &= \frac{\mathbf{z}_{\cdot g_{2||j}}^* \left((\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - (\mathbf{Z} + \mathbf{DZ}\boldsymbol{\Psi})\mathbf{g}_1 - \mathbf{DZ}\mathbf{g}_{2||}) + \mathbf{z}_{\cdot g_{2||j}}^* g_{2||j} \right)}{\mathbf{z}_{\cdot g_{2||j}}^* \mathbf{z}_{\cdot g_{2||j}}^* + \sigma_e^2 \sigma_{g_{2||j}}^{-2}} \\ &= \frac{\mathbf{z}_{\cdot g_{2||j}}^* (\mathbf{e} + \mathbf{z}_{\cdot g_{2||j}}^* g_{2||j})}{\mathbf{z}_{\cdot g_{2||j}}^* \mathbf{z}_{\cdot g_{2||j}}^* + \sigma_e^2 \sigma_{g_{2||j}}^{-2}} = \frac{\mathbf{z}_{\cdot g_{2||j}}^* \mathbf{e} + \mathbf{z}_{\cdot g_{2||j}}^* \mathbf{z}_{\cdot g_{2||j}}^* g_{2||j}}{\mathbf{z}_{\cdot g_{2||j}}^* \mathbf{z}_{\cdot g_{2||j}}^* + \sigma_e^2 \sigma_{g_{2||j}}^{-2}} \\ &= \frac{\sum_{i=1}^n (d_i z_{ij}) e_i + \left(\sum_{i=1}^n (d_i z_{ij})^2 \right) g_{2||j}}{\left(\sum_{i=1}^n (d_i z_{ij})^2 + \sigma_e^2 \sigma_{g_{2||j}}^{-2} \right)} \end{aligned}$$

$$\text{and } \tilde{v}_{g_{2||j}} = \left(\sigma_e^{-2} \sum_{i=1}^n (d_i z_{ij})^2 + \sigma_{g_{2||j}}^{-2} \right)^{-1}$$

Immediately after sampling $g_{2||j}$, we update the residual by

$$\mathbf{e} = \mathbf{e} - \mathbf{z}_{\cdot g_{2||j}}^* \left(g_{2||j}^{[t+1]} - g_{2||j}^{[t]} \right), \text{ where } g_{2||j}^{[t+1]} \text{ is the sampled } g_{2||j} \text{ value at cycle } [t+1]$$

and $g_{2||j}^{[t]}$ is the sampled $g_{2||j}$ value at cycle $[t]$.

To sample variances of SNP intercept effects in CD-BayesA, given the previously specified prior $p(\sigma_{g_{1j}}^2 \mid v_1, s_1^2) = \chi^{-2}(v_1, v_1 s_1^2)$, we can derive FCD as follows:

$$\begin{aligned}
p(\sigma_{g_{1j}}^2 | \mathbf{y}, \text{else}) &\propto p(g_{1j} | \sigma_{g_{1j}}^2) p(\sigma_{g_{1j}}^2 | \nu_1, s_1^2) \\
&\propto (2\pi\sigma_{g_{1j}}^2)^{-1/2} \exp\left(-\frac{g_{1j}^2}{2\sigma_{g_{1j}}^2}\right) \sigma_{g_{1j}}^{2-\left(\frac{\nu_1+1}{2}\right)} \exp\left(-\frac{\nu_1 s_1^2}{2\sigma_{g_{1j}}^2}\right) \\
&\propto \sigma_{g_{1j}}^{2-\left(\frac{\nu_1+1}{2}+1\right)} \exp\left(-\frac{\nu_1 s_1^2 + g_{1j}^2}{2\sigma_{g_{1j}}^2}\right)
\end{aligned}$$

That is, $\sigma_{g_{1j}}^2 | \mathbf{y}, \text{else} \sim \chi^{-2}(df = \nu_1 + 1, scale = g_{1j}^2 + \nu_1 s_1^2)$

To sample variances of SNP slope effects conditional on intercept in CD-BayesA, given the previously specified prior $p(\sigma_{g_{2|l j}}^2 | \nu_2, s_2^2) = \chi^{-2}(\nu_2, \nu_2 s_2^2)$, we can derive FCD as follows:

$$\begin{aligned}
p(\sigma_{g_{2|l j}}^2 | \mathbf{y}, \text{else}) &\propto p(g_{2|l j} | \sigma_{g_{2|l j}}^2) p(\sigma_{g_{2|l j}}^2 | \nu_2, s_2^2) \\
&\propto (2\pi\sigma_{g_{2|l j}}^2)^{-1/2} \exp\left(-\frac{g_{2|l j}^2}{2\sigma_{g_{2|l j}}^2}\right) \sigma_{g_{2|l j}}^{2-\left(\frac{\nu_2+1}{2}\right)} \exp\left(-\frac{\nu_2 s_2^2}{2\sigma_{g_{2|l j}}^2}\right) \\
&\propto \sigma_{g_{2|l j}}^{2-\left(\frac{\nu_2+1}{2}+1\right)} \exp\left(-\frac{\nu_2 s_2^2 + g_{2|l j}^2}{2\sigma_{g_{2|l j}}^2}\right)
\end{aligned}$$

That is, $\sigma_{g_{2|l j}}^2 | \mathbf{y}, \text{else} \sim \chi^{-2}(df = \nu_2 + 1, scale = g_{2|l j}^2 + \nu_2 s_2^2)$

To sample variances and effects of SNP intercepts in CD-BayesB, according to the collapsed sampling strategy (Liu 1994), we jointly sampled \mathbf{g}_{1j} and $\sigma_{g_{1j}}^2$ as adapted in Bayes B (Meuwissen et al. 2001). Let's define

$\mathbf{z}_{\cdot, g_{1j}}^* = \begin{bmatrix} z_{1j}(1 + d_1 \phi_j) & z_{2j}(1 + d_2 \phi_j) & z_{3j}(1 + d_3 \phi_j) & \cdots & z_{nj}(1 + d_n \phi_j) \end{bmatrix}'$ is the j^{th} column of

$\mathbf{Z} + \mathbf{DZ}\Psi$. We first sample from

$$\begin{aligned}
p(\sigma_{g_{1j}}^2 \mid ELSE \text{ except } g_{1j}) &= \int_{g_{1j}} p(\sigma_{g_{1j}}^2, g_{1j} \mid ELSE) dg_{1j} \\
&\propto \int_{g_{1j}} \left(\prod_{i=1}^n p(y_i \mid \boldsymbol{\beta}, \mathbf{g}_{-1j}, g_{1j}, \mathbf{g}_{2||}, \sigma_e^2) \right) p(g_{1j} \mid \sigma_{g_{1j}}^2) p(\sigma_{g_{1j}}^2 \mid \nu_1, s_1^2, \pi_1) dg_{1j} \\
&\propto p(\sigma_{g_{1j}}^2 \mid \nu_1, s_1^2, \pi_1) \left| \mathbf{V}_{g_{1j}} \right|^{-1/2} \exp\left(-\frac{1}{2} \mathbf{y}_{g_{1j}}^* ' \mathbf{V}_{g_{1j}}^{-1} \mathbf{y}_{g_{1j}}^* \right)
\end{aligned}$$

where $\mathbf{y}_{g_{1j}}^* = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - (\mathbf{Z} + \mathbf{DZ}\boldsymbol{\Psi})_{-j} \mathbf{g}_{-1j} - \mathbf{DZ}\mathbf{g}_{2||})$ and $\mathbf{V}_{g_{1j}} = \mathbf{z}_{g_{1j}}^* \mathbf{z}_{g_{1j}}'^* \sigma_{g_{1j}}^2 + \mathbf{I}\sigma_e^2$.

Using that expression, the random walk Metropolis-Hastings acceptance ratio for sampling from $p(\sigma_{g_{1j}}^2 \mid ELSE \text{ except } g_{1j})$ at MCMC cycle [t] based on using

$p(\sigma_{g_{1j}}^2 \mid \nu_1, s_1^2, \pi_1)$ as the candidate density looks as follows:

$$\alpha(\sigma_{g_{1j}}^{2[t-1]}, \sigma_{g_{1j}}^{2*}) = \begin{cases} \min \left(\frac{p(\sigma_{g_{1j}}^{2*} \mid ELSE \text{ except } g_{1j}) p(\sigma_{g_{1j}}^{2[t-1]} \mid \nu_1, s_1^2, \pi_1)}{p(\sigma_{g_{1j}}^{2[t-1]} \mid ELSE \text{ except } g_{1j}) p(\sigma_{g_{1j}}^{2*} \mid \nu_1, s_1^2, \pi_1)}, 1 \right) \\ 1, otherwise; \end{cases}$$

which can be demonstrated (Meuwissen et al. (2001)), to be equal to:

$$\alpha(\sigma_{g_{1j}}^{2[t-1]}, \sigma_{g_{1j}}^{2*}) = \begin{cases} \min \left(\frac{\left| \mathbf{V}_{g_{1j}}^* \right|^{-1/2} \exp\left(-\frac{1}{2} \mathbf{y}_{g_{1j}}^* ' \mathbf{V}_{g_{1j}}^{*-1} \mathbf{y}_{g_{1j}}^* \right)}{\left| \mathbf{V}_{g_{1j}}^{[t-1]} \right|^{-1/2} \exp\left(-\frac{1}{2} \mathbf{y}_{g_{1j}}^* ' (\mathbf{V}_{g_{1j}}^{[t-1]})^{-1} \mathbf{y}_{g_{1j}}^* \right)}, 1 \right) \\ 1, otherwise \end{cases}$$

and further simplified as follows:

$$\alpha\left(\sigma_{g_{1j}}^{2[t-1]}, \sigma_{g_{1j}}^{2*}\right) = \begin{cases} \min\left(\frac{\left(v_{g_{1j}}^*\right)^{-1/2} \exp\left(-\frac{r_{g_{1j}}^2}{2v_{g_{1j}}^*}\right)}{\left(v_{g_{1j}}^{[t-1]}\right)^{-1/2} \exp\left(-\frac{r_{g_{1j}}^2}{2v_{g_{1j}}^{[t-1]}}\right)}, 1\right) \\ 1, \text{otherwise} \end{cases}$$

where

$$r_{g_{1j}} = \mathbf{z}_{\cdot g_{1j}}^* \mathbf{y}_{g_{1j}}^* = \sum_{i=1}^n z_{ij} (1 + d_i \phi_j) e_i + \left(\sum_{i=1}^n z_{ij}^2 (1 + d_i \phi_j)^2 \right) g_{1j}$$

and

$$\begin{aligned} v_{g_{1j}} &= \text{var}(\mathbf{z}_{\cdot g_{1j}}^* \mathbf{y}_{g_{1j}}^*) = \left(\mathbf{z}_{\cdot g_{1j}}^* \mathbf{z}_{\cdot g_{1j}}^* \right) \sigma_{g_{1j}}^2 + \left(\mathbf{z}_{\cdot g_{1j}}^* \mathbf{z}_{\cdot g_{1j}}^* \right) \sigma_e^2 \\ &= \left(\sum_{i=1}^n z_{ij}^2 (1 + d_i \phi_j)^2 \right) \sigma_{g_{1j}}^2 + \left(\sum_{i=1}^n z_{ij}^2 (1 + d_i \phi_j)^2 \right) \sigma_e^2 \end{aligned}$$

If a non-zero value for $\sigma_{g_{1j}}^2$ is sampled, then one can draw samples of g_{1j} using the same

full conditionals as with CD-BayesA. If either $g_{1j}^{[t+1]}$ or $g_{1j}^{[t]}$ are nonzero, the residual

needs to be updated as $\mathbf{e} = \mathbf{e} - \mathbf{z}_{\cdot g_{1j}}^* (g_{1j}^{[t+1]} - g_{1j}^{[t]})$ immediately thereafter.

To sample variances and effects of SNP slopes conditional on intercepts in CD-BayesB, we jointly sampled $g_{2||j}$ and $\sigma_{g_{2||j}}^2$ as adapted in Bayes B (Meuwissen et al.

2001). Let's define $\mathbf{z}_{\cdot g_{2||j}}^* = [z_{1j}d_1 \quad z_{2j}d_2 \quad z_{3j}d_3 \quad \cdots \quad z_{nj}d_n]'$ is the j^{th} column of \mathbf{DZ} . We

first sample from

$$\begin{aligned}
p\left(\sigma_{g_{2||j}}^2 \mid ELSE \text{ except } g_{2||j}\right) &= \int p\left(\sigma_{g_{2||j}}^2, g_{2||j} \mid ELSE\right) dg_{2||j} \\
&\propto \int \left(\prod_{i=1}^n p\left(y_i \mid \boldsymbol{\beta}, \mathbf{g}_{-2||j}, g_{2||j}, \mathbf{g}_1, \sigma_e^2\right) \right) p\left(g_{2||j} \mid \sigma_{g_{2||j}}^2\right) p\left(\sigma_{g_{2||j}}^2 \mid \nu_2, s_2^2, \pi_2\right) dg_{2||j} \\
&\propto p\left(\sigma_{g_{2||j}}^2 \mid \nu_2, s_2^2, \pi_2\right) \left| \mathbf{V}_{g_{2||j}} \right|^{-1/2} \exp\left(-\frac{1}{2} \mathbf{y}_{g_{2||j}}^* \mathbf{V}_{g_{2||j}}^{-1} \mathbf{y}_{g_{2||j}}^*\right)
\end{aligned}$$

where $\mathbf{y}_{g_{2||j}}^* = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - (\mathbf{Z} + \mathbf{DZ}\Psi)\mathbf{g}_1 - (\mathbf{DZ})_{-j} \mathbf{g}_{-2||j})$ and $\mathbf{V}_{g_{2||j}} = \mathbf{z}_{g_{2||j}}^* \mathbf{z}_{g_{2||j}}'^* \sigma_{g_{2||j}}^2 + \mathbf{I} \sigma_e^2$.

Using that expression, the random walk Metropolis-Hastings acceptance ratio for

sampling from $p\left(\sigma_{g_{2||j}}^2 \mid ELSE \text{ except } g_{2||j}\right)$ at MCMC cycle [t] based on using

$p\left(\sigma_{g_{2||j}}^2 \mid \nu_2, s_2^2, \pi_2\right)$ as the candidate density looks as follows:

$$\alpha\left(\sigma_{g_{2||j}}^{2[t-1]}, \sigma_{g_{2||j}}^{2*}\right) = \begin{cases} \min\left(\frac{p\left(\sigma_{g_{2||j}}^{2*} \mid ELSE \text{ except } g_{2||j}\right) p\left(\sigma_{g_{2||j}}^{2[t-1]} \mid \nu_2, s_2^2, \pi_2\right)}{p\left(\sigma_{g_{2||j}}^{2[t-1]} \mid ELSE \text{ except } g_{2||j}\right) p\left(\sigma_{g_{2||j}}^{2*} \mid \nu_2, s_2^2, \pi_2\right)}, 1\right) \\ 1, otherwise; \end{cases}$$

According to Meuwissen et al. (2001), this ratio is further equal to:

$$\alpha\left(\sigma_{g_{2||j}}^{2[t-1]}, \sigma_{g_{2||j}}^{2*}\right) = \begin{cases} \min\left(\frac{\left| \mathbf{V}_{g_{2||j}}^* \right|^{-1/2} \exp\left(-\frac{1}{2} \mathbf{y}_{g_{2||j}}^* \mathbf{V}_{g_{2||j}}^{*-1} \mathbf{y}_{g_{2||j}}^*\right)}{\left| \mathbf{V}_{g_{2||j}}^{[t-1]} \right|^{-1/2} \exp\left(-\frac{1}{2} \mathbf{y}_{g_{2||j}}^* \left(\mathbf{V}_{g_{2||j}}^{[t-1]}\right)^{-1} \mathbf{y}_{g_{2||j}}^*\right)}, 1\right) \\ 1, otherwise \end{cases}$$

and can be further simplified as:

$$\alpha\left(\sigma_{g_{2||j}}^{2[t-1]}, \sigma_{g_{2||j}}^{2*}\right) = \begin{cases} \min\left(\frac{\left(v_{g_{2||j}}^*\right)^{-1/2} \exp\left(-\frac{r_{g_{2||j}}^2}{2v_{g_{2||j}}^*}\right)}{\left(v_{g_{2||j}}^{[t-1]}\right)^{-1/2} \exp\left(-\frac{r_{g_{2||j}}^2}{2v_{g_{2||j}}^{[t-1]}}\right)}, 1\right) \\ 1, \text{otherwise} \end{cases}$$

where

$$r_{g_{2||j}} = \mathbf{z}_{\cdot g_{2||j}}^* \mathbf{y}_{g_{2||j}}^* = \sum_{i=1}^n z_{ij} d_i e_i + \left(\sum_{i=1}^n (z_{ij} d_i)^2 \right) g_{2||j}$$

and

$$\begin{aligned} v_{g_{2||j}} &= \text{var}(\mathbf{z}_{\cdot g_{2||j}}^* \mathbf{y}_{g_{2||j}}^*) = \left(\mathbf{z}_{\cdot g_{2||j}}^* \mathbf{z}_{\cdot g_{2||j}}^* \right)^2 \sigma_{g_{2||j}}^2 + \left(\mathbf{z}_{\cdot g_{2||j}}^* \mathbf{z}_{\cdot g_{2||j}}^* \right) \sigma_e^2 \\ &= \left(\sum_{i=1}^n (z_{ij} d_i)^2 \right)^2 \sigma_{g_{2||j}}^2 + \left(\sum_{i=1}^n (z_{ij} d_i)^2 \right) \sigma_e^2 \end{aligned}$$

If a non-zero value for $\sigma_{g_{2||j}}^2$ is sampled, then one can draw samples of $g_{2||j}$ using the

same full conditionals as with CD-BayesA. If either $g_{2||j}^{[t+1]}$ or $g_{2||j}^{[t]}$ are nonzero, the

residual needs to be updated as $\mathbf{e} = \mathbf{e} - \mathbf{z}_{\cdot g_{2||j}}^* \left(g_{2||j}^{[t+1]} - g_{2||j}^{[t]} \right)$ immediately thereafter.

To sample proportion of SNP markers associated with non-zero intercept effects in CD-BayesB, with specified prior $p(\pi_1 | \alpha_{\pi_1}, \beta_{\pi_1}) = \text{Beta}(\alpha_{\pi_1}, \beta_{\pi_1})$, the FCD of π_1 is based on the following,

$$p(\pi_1 | \mathbf{y}, \text{else}) \propto \left(\prod_{j=1}^m p\left(\sigma_{g_{1j}}^2 | v_1, s_1^2, \pi_1\right) \right) p(\pi_1 | \alpha_{\pi_1}, \beta_{\pi_1})$$

Let $m_1 = \sum_{j=1}^m I(\sigma_{g_{1j}}^2 > 0)$ denote the number of non-zero values sampled in $\sigma_{g_{1j}}^2$ at a

particular MCMC cycle where $I(\cdot)$ denotes the indicator function. Then we can write

$$p(\pi_1 | \mathbf{y}, \text{else}) = \text{Beta}(\alpha_{\pi_1} + m_1, \beta_{\pi_1} + m - m_1) .$$

To sample proportion of SNP markers associated with non-zero slope effects conditional on intercept in CD-BayesB, with specified prior

$$p(\pi_2 | \alpha_{\pi_2}, \beta_{\pi_2}) = \text{Beta}(\alpha_{\pi_2}, \beta_{\pi_2}) , \text{ the FCD of } \pi_2 \text{ is based on the following,}$$

$$p(\pi_2 | \mathbf{y}, \text{else}) \propto \left(\prod_{j=1}^m p(\sigma_{g_{2|j}}^2 | v_2, s_2^2, \pi_2) \right) p(\pi_2 | \alpha_{\pi_2}, \beta_{\pi_2})$$

Let $m_2 = \sum_{j=1}^m I(\sigma_{g_{2|j}}^2 > 0)$ denote the number of non-zero values sampled in $\sigma_{g_{2|j}}^2$ at a

particular MCMC cycle where $I(\cdot)$ denotes the indicator function. Then we can write

$$p(\pi_2 | \mathbf{y}, \text{else}) = \text{Beta}(\alpha_{\pi_2} + m_2, \beta_{\pi_2} + m - m_2) .$$

To sample association parameters between intercept and slope effects in CD-BayesA/CD-BayesB, with specified prior $\phi_j \sim N(\mu_\phi, \sigma_\phi^2)$, let's define $\mathbf{G} = \text{diag}\{g_{1j}\}_{j=1}^m$

and $\mathbf{z}_{\cdot\phi j}^* = [z_{1j}d_1g_{1j} \quad z_{2j}d_2g_{1j} \quad z_{3j}d_3g_{1j} \quad \cdots \quad z_{nj}d_ng_{1j}]'$ is column j of \mathbf{DZG} . The FCD for

ϕ_j can be written as follows:

$$\phi_j | \mathbf{y}, \boldsymbol{\beta}, \phi_{-j}, \mathbf{g}_1, \mathbf{g}_{2|j}, \{\sigma_{g_{1j}}^2\}_{j=1}^m, \{\sigma_{g_{2|j}}^2\}_{j=1}^m, \sigma_e^2, \sim N(\tilde{\phi}_j, \tilde{v}_{\phi_j})$$

where

$$\begin{aligned}
\tilde{\phi}_j &= \frac{\sigma_e^2 \sigma_\phi^{-2} \mu_\phi + \mathbf{z}_{\cdot\phi_j}^* \left((\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{g}_1 - \mathbf{DZG}\boldsymbol{\phi} - \mathbf{DZg}_{2||}) + \mathbf{z}_{\cdot\phi_j}^* \phi_j \right)}{\mathbf{z}_{\cdot\phi_j}^* \mathbf{z}_{\cdot\phi_j}^* + \sigma_\phi^{-2} \sigma_e^2} \\
&= \frac{\sigma_e^2 \sigma_\phi^{-2} \mu_\phi + \mathbf{z}_{\cdot\phi_j}^* (\mathbf{e} + \mathbf{z}_{\cdot\phi_j}^* \phi_j)}{\mathbf{z}_{\cdot\phi_j}^* \mathbf{z}_{\cdot\phi_j}^* + \sigma_\phi^{-2} \sigma_e^2} = \frac{\sigma_e^2 \sigma_\phi^{-2} \mu_\phi + \mathbf{z}_{\cdot\phi_j}^* \mathbf{e} + \mathbf{z}_{\cdot\phi_j}^* \mathbf{z}_{\cdot\phi_j}^* \phi_j}{\mathbf{z}_{\cdot\phi_j}^* \mathbf{z}_{\cdot\phi_j}^* + \sigma_\phi^{-2} \sigma_e^2} \\
&= \frac{\sigma_e^2 \sigma_\phi^{-2} \mu_\phi + \sum_{i=1}^n z_{ij} d_i g_{1j} e_i + \left(\sum_{i=1}^n (z_{ij} d_i g_{1j})^2 \right) \phi_j}{\left(\sum_{i=1}^n (z_{ij} d_i g_{1j})^2 + \sigma_\phi^{-2} \sigma_e^2 \right)}
\end{aligned}$$

and

$$\tilde{v}_{\phi_j} = \left(\sigma_e^{-2} \sum_{i=1}^n (z_{ij} d_i g_{1j})^2 + \sigma_\phi^{-2} \right)^{-1}$$

Immediately after sampling ϕ_j , we update the residual by $\mathbf{e} = \mathbf{e} - \mathbf{z}_{\cdot\phi_j}^* (\phi_j^{[t+1]} - \phi_j^{[t]})$, where

$\phi_j^{[t+1]}$ is the sampled ϕ_j value at cycle [t+1] and $\phi_j^{[t]}$ is the sampled ϕ_j value at cycle [t].

To sampling mean for association parameters in CD-BayesA/CD-BayesB, given a specified prior $p(\mu_\phi) = N(\tau, \zeta^2)$, FCD for μ_ϕ can be written as follows,

$$\begin{aligned}
p(\mu_\phi | \mathbf{y}, ELSE) &\propto \left(\prod_{j=1}^m p(\phi_j | \mu_\phi, \sigma_\phi^2) \right) p(\mu_\phi | \tau, \zeta^2) \\
&\propto \exp \left(-\frac{1}{2\sigma_\phi^2} \sum_{j=1}^m (\phi_j - \mu_\phi)^2 \right) \exp \left(-\frac{(\mu_\phi - \tau)^2}{2\zeta^2} \right) \\
&\propto \exp \left(-\frac{1}{2\tilde{\sigma}_{\mu_\phi}^2} \sum_{i=1}^m (\mu_\phi - \tilde{\mu}_\phi)^2 \right)
\end{aligned}$$

$$\text{Where } \tilde{\mu}_\phi = \frac{\left(\frac{\sigma_\phi^2}{m}\right)^{-1} \bar{\phi} + (\zeta^2)^{-1} \tau}{(\zeta^2)^{-1} + \left(\frac{\sigma_\phi^2}{m}\right)^{-1}} \text{ for } \bar{\phi} = \frac{\sum_{j=1}^m \phi_j}{m} \text{ and } \tilde{\sigma}_{\mu_\phi}^2 = \left((\zeta^2)^{-1} + \left(\frac{\sigma_\phi^2}{m}\right)^{-1} \right)^{-1}$$

$$\text{i.e. } \mu_\phi | \mathbf{y}, ELSE \sim N\left(\tilde{\mu}_\phi, \tilde{\sigma}_{\mu_\phi}^2\right)$$

To sampling variance of association parameters in CD-BayesA/CD-BayesB, with a specified prior $p(\sigma_\phi^2) = \text{Gamma}(\alpha_\phi, \beta_\phi)$, FCD for σ_ϕ^2 can be written as follows,

$$\begin{aligned} p(\sigma_\phi^2 | \mathbf{y}, ELSE) &\propto \left(\prod_{j=1}^m p(\phi_j | \mu_\phi, \sigma_\phi^2) \right) p(\sigma_\phi^2) \\ &\left((2\pi\sigma_\phi^2)^{-m/2} \exp\left(-\frac{1}{2\sigma_\phi^2} \sum_{j=1}^m (\phi_j - \mu_\phi)^2\right) \right) (\sigma_\phi^2)^{\alpha_\phi-1} e^{-\beta_\phi\sigma_\phi^2} \\ &\propto \sigma_\phi^{2-\left(\frac{m-2\alpha_\phi}{2}+1\right)} \exp\left(-\frac{(\phi - \mu_\phi)'(\phi - \mu_\phi) + \beta_\phi}{2\sigma_\phi^2}\right) \end{aligned}$$

That is, $\sigma_\phi^2 | \mathbf{y}, ELSE \sim \chi^2(df = (m - 2\alpha_\phi), scale = (\phi - \mu_\phi)'(\phi - \mu_\phi) + \beta_\phi)$.

To sample scale parameter for SNP intercepts in CD-BayesA/CD-BayesB, with a specified prior for scale parameter $p(s_1^2 | \alpha_1, \beta_1) = \text{Gamma}(\alpha_1, \beta_1)$, we can write FCD as follows:

$$\begin{aligned}
p(s_1^2 | \mathbf{y}, ELSE) &\propto \left(\prod_{j=1}^m I(\sigma_{g_{1j}}^2 > 0) p(\sigma_{g_{1j}}^2 | \nu_1, s_1^2, \pi_1) \right) p(s_1^2 | \alpha_1, \beta_1) \\
&= \left(\prod_{j=1}^m I(\sigma_{g_{1j}}^2 > 0) \frac{\left(\frac{s_1^2 \nu_1}{2} \right)^{\frac{\nu_1}{2}}}{\Gamma\left(\frac{\nu_1}{2}\right)} \sigma_{g_{1j}}^{-2} e^{-\left(\frac{\nu_1+1}{2}\right) \frac{s_1^2 \nu_1}{2\sigma_{g_{1j}}^2}} \right) (s_1^2)^{\alpha_1-1} e^{-\beta_1 s_1^2} \\
&\propto (s_1^2)^{\alpha_1 + \frac{m_1 \nu_1}{2} - 1} \exp \left(-s_1^2 \left(\frac{\nu_1}{2} \sum_{j=1}^m I(\sigma_{g_{1j}}^2 > 0) \sigma_{g_{1j}}^{-2} + \beta_1 \right) \right)
\end{aligned}$$

That is, a gamma distribution with parameters $\alpha_1 + \frac{m_1 \nu_1}{2}$ and $\frac{\nu_1}{2} \sum_{j=1}^m I(\sigma_{g_{1j}}^2 > 0) \sigma_{g_{1j}}^{-2} + \beta_1$.

To sample scale parameter for SNP slopes conditional on intercept in CD-BayesA/CD-BayesB, with a specified prior for scale parameter

$p(s_2^2 | \alpha_2, \beta_2) = \text{Gamma}(\alpha_2, \beta_2)$, we can write FCD as follows,

$$\begin{aligned}
p(s_2^2 | \mathbf{y}, ELSE) &\propto \left(\prod_{j=1}^m I(\sigma_{g_{2|j}}^2 > 0) p(\sigma_{g_{2|j}}^2 | \nu_2, s_2^2, \pi_2) \right) p(s_2^2 | \alpha_2, \beta_2) \\
&= \left(\prod_{j=1}^m I(\sigma_{g_{2|j}}^2 > 0) \frac{\left(\frac{s_2^2 \nu_2}{2} \right)^{\frac{\nu_2}{2}}}{\Gamma\left(\frac{\nu_2}{2}\right)} \sigma_{g_{2|j}}^{-2} e^{-\left(\frac{\nu_2+1}{2}\right) \frac{s_2^2 \nu_2}{2\sigma_{g_{2|j}}^2}} \right) (s_2^2)^{\alpha_2-1} e^{-\beta_2 s_2^2} \\
&\propto (s_2^2)^{\alpha_2 + \frac{m_2 \nu_2}{2} - 1} \exp \left(-s_2^2 \left(\frac{\nu_2}{2} \sum_{j=1}^m I(\sigma_{g_{2|j}}^2 > 0) \sigma_{g_{2|j}}^{-2} + \beta_2 \right) \right)
\end{aligned}$$

That is, a gamma distribution with parameters $\alpha_2 + \frac{m_2 \nu_2}{2}$ and $\frac{\nu_2}{2} \sum_{j=1}^m I(\sigma_{g_{2|j}}^2 > 0) \sigma_{g_{2|j}}^{-2} + \beta_2$.

To sampling degrees of freedom for SNP intercept effects in CD-BayesA/CD-BayesB, with a specified non-informative prior for degrees of freedom $p(v_1) = 1 / (1 + v_1)^2$, we can write FCD for v_1 as follows:

$$p(v_1 | \mathbf{y}, ELSE) \propto \left(\prod_{j=1}^m I(\sigma_{g_{1j}}^2 > 0) p(\sigma_{g_{1j}}^2 | v_1, s_1^2, \pi_1) \right) p(v_1)$$

$$= \left(\prod_{j=1}^m I(\sigma_{g_{1j}}^2 > 0) \frac{\left(\frac{s_1^2}{2}\right)^{\frac{v_1}{2}}}{\Gamma\left(\frac{v_1}{2}\right)} \sigma_{g_{1j}}^2^{-\left(\frac{v_1}{2}+1\right)} e^{-\frac{s_1^2}{2\sigma_{g_{1j}}^2}} \right) p(v_1)$$

We sampled v_1 using a random walk Metropolis Hastings algorithm, which was described in other non-genomic applications involving the sampling of degrees of freedom parameter (Kizilkaya & Tempelman 2005; Bello 2010).

To sample degrees of freedom for SNP slopes conditional on intercepts in CD-BayesA/CD-BayesB, with a specified non-informative prior for degrees of freedom

$p(v_2) = 1 / (1 + v_2)^2$, we can write FCD for v_2 as follows:

$$p(v_2 | \mathbf{y}, ELSE) \propto \left(\prod_{j=1}^m I(\sigma_{g_{2|j}}^2 > 0) p(\sigma_{g_{2|j}}^2 | v_2, s_2^2, \pi_2) \right) p(v_2)$$

$$= \left(\prod_{j=1}^m I(\sigma_{g_{2|j}}^2 > 0) \frac{\left(\frac{s_2^2}{2}\right)^{\frac{v_2}{2}}}{\Gamma\left(\frac{v_2}{2}\right)} \sigma_{g_{2|j}}^2^{-\left(\frac{v_2}{2}+1\right)} e^{-\frac{s_2^2}{2\sigma_{g_{2|j}}^2}} \right) p(v_2)$$

We sampled ν_2 using a random walk Metropolis Hastings algorithm, which was described in other non-genomic applications involving the sampling of degrees of freedom parameter (Kizilkaya & Tempelman 2005; Bello 2010).

To sample residual variance in CD-BayesA/CD-BayesB, given a specified scaled inverted Chi-square prior $p(\sigma_e^2 | \nu_e, S_e) = \chi^{-2}(\nu_e, \nu_e S_e)$, we can write the FCD for σ_e^2 as follows,

$$\begin{aligned} p(\sigma_e^2 | \mathbf{y}, else) \\ \propto (2\pi\sigma_e^2)^{-n/2} \exp\left(-\frac{1}{2\sigma_e^2} \mathbf{e}'\mathbf{e}\right) \sigma_e^{2-\left(\frac{\nu_e}{2}+1\right)} e^{-\frac{\nu_e S_e}{2\sigma_e^2}} \\ \propto \sigma_e^{2-\left(\frac{n+\nu_e}{2}+1\right)} \exp\left(-\frac{1}{2\sigma_e^2} (\mathbf{e}'\mathbf{e} + \nu_e S_e)\right) \end{aligned}$$

where $\mathbf{e} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta} - (\mathbf{Z} + \mathbf{DZ}\Psi)\mathbf{g}_1 - \mathbf{DZ}\mathbf{g}_{2||}$

i.e. $p(\sigma_e^2 | \mathbf{y}, else) \sim \chi^{-2}(df = n + \nu_e, scale = \mathbf{e}'\mathbf{e} + \nu_e S_e)$

C1.3 Derivation for overall genetic correlation between intercept and slope

The RR/RN WGP model can be written as follows,

$$y_i = \mathbf{x}'_i \boldsymbol{\beta} + \sum_{j=1}^m z_{ij} (g_{1j} + d_i g_{2j}) + e_i; i = 1, 2, \dots, n$$

Then, the genetic variance at environmental covariate d_i for animal i with genotype z_{ij} at locus j can be defined as,

$$\begin{aligned} \text{var} \left(\sum_{j=1}^m z_{ij} (g_{1j} + d_i g_{2j}) \right) &= \sum_{j=1}^m z_{ij}^2 (\sigma_{g_{1j}}^2 + d_i^2 \sigma_{g_{2j}}^2 + 2d_i \sigma_{g_{1j}g_{2j}}) \\ &= \sum_{j=1}^m z_{ij}^2 \sigma_{g_{1j}}^2 + \sum_{j=1}^m z_{ij}^2 d_i^2 \sigma_{g_{2j}}^2 + \sum_{j=1}^m z_{ij}^2 2d_i \sigma_{g_{1j}g_{2j}} \end{aligned}$$

Thus, for any environment d , we can define the genetic variance as follows,

$$\begin{aligned} \text{var} \left(\sum_{j=1}^m z_{ij} (g_{1j} + dg_{2j}) \right) \\ = \sum_{j=1}^m z_{ij}^2 \sigma_{g_{1j}}^2 + d^2 \sum_{j=1}^m z_{ij}^2 \sigma_{g_{2j}}^2 + 2d \sum_{j=1}^m z_{ij}^2 \sigma_{g_{1j}g_{2j}} \end{aligned}$$

Using de los Campos et al. (2012), we can compute the overall genetic variance across all animals as follows:

$$\begin{aligned} V_g &= n^{-1} \sum_{i=1}^n \left(\sum_{j=1}^m z_{ij}^2 \sigma_{g_{1j}}^2 + d^2 \sum_{j=1}^m z_{ij}^2 \sigma_{g_{2j}}^2 + 2d \sum_{j=1}^m z_{ij}^2 \sigma_{g_{1j}g_{2j}} \right) \\ &= n^{-1} \sum_{i=1}^n \sum_{j=1}^m z_{ij}^2 \sigma_{g_{1j}}^2 + d^2 n^{-1} \sum_{i=1}^n \sum_{j=1}^m z_{ij}^2 \sigma_{g_{2j}}^2 + 2dn^{-1} \sum_{i=1}^n \sum_{j=1}^m z_{ij}^2 \sigma_{g_{1j}g_{2j}} \end{aligned}$$

where $n^{-1} \sum_{i=1}^n \sum_{j=1}^m z_{ij}^2 \sigma_{g_{1j}}^2$ is the overall genetic variance for intercept across animals,

$n^{-1} \sum_{i=1}^n \sum_{j=1}^m z_{ij}^2 \sigma_{g_{2j}}^2$ is the overall genetic variance for slope across animals and

$n^{-1} \sum_{i=1}^n \sum_{j=1}^m z_{ij}^2 \sigma_{g_{1g_{2j}}}$ is the overall genetic covariance between intercept and slope across

animals. Therefore, the overall genetic correlation $\bar{\rho}_{g_1 g_2}$ between intercept and slope can be written as,

$$\bar{\rho}_{g_1 g_2} = \frac{n^{-1} \sum_{i=1}^n \sum_{j=1}^m z_{ij}^2 \sigma_{g_{1g_{2j}}}}{\sqrt{\left(n^{-1} \sum_{i=1}^n \sum_{j=1}^m z_{ij}^2 \sigma_{g_{1j}}^2 \right) \left(n^{-1} \sum_{i=1}^n \sum_{j=1}^m z_{ij}^2 \sigma_{g_{2j}}^2 \right)}}$$

C2 Supplementary tables and figures

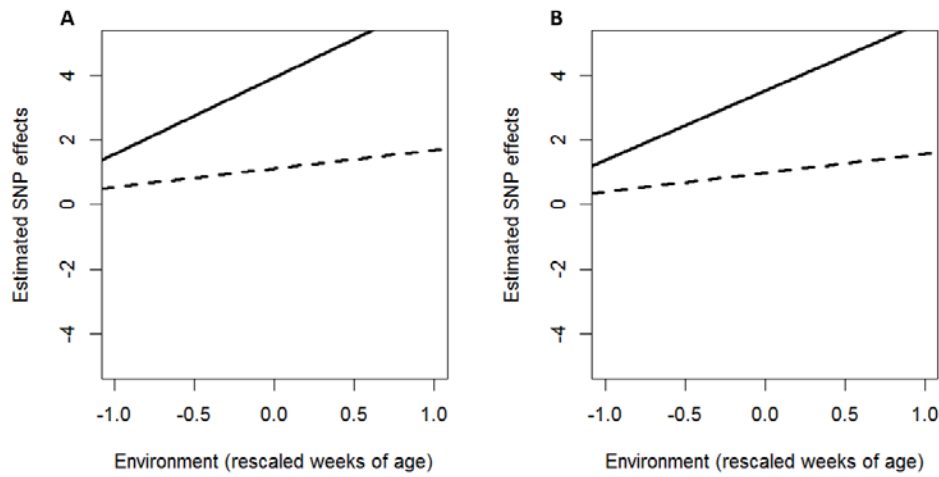


Figure C2.1: Estimated SNP effects with Environmental (rescaled weeks of age 10, 13, 16, 19 and 22) dependence on back fat thickness for two SNP markers (solid line and dash line) using the complete final analyses data in MSU Pig Resource Population under model A) IW-BayesA, B) CD-BayesA.

APPENDIX D: Chapter 5

D1 Markov Chain Monte Carlo (MCMC) Implementation Strategy for Bayesian hierarchical methods

D1.1 MCMC Implementation Strategy for IWBayesA

To sample location parameters computationally efficiently, we adopted Gauss-Seidel updating algorithm in MCMC implementation strategy (Legarra and Misztal, 2008). In the bivariate trait WGP model, \mathbf{y}_1 is the $n \times 1$ vector of phenotypes for animals on trait 1, \mathbf{y}_2 is the $n \times 1$ vector of phenotypes for animals on trait 2, $\boldsymbol{\beta}_1$ ($\boldsymbol{\beta}_2$) is the $q \times 1$ vector of fixed effects on trait 1 (trait 2), \mathbf{g}_1 (\mathbf{g}_2) represents the $m \times 1$ vector of SNP substitution effects on trait 1 (trait 2), \mathbf{X}_1 (\mathbf{X}_2) is the $n \times q$ design matrix for fixed effect $\boldsymbol{\beta}_1$ ($\boldsymbol{\beta}_2$), \mathbf{Z} is the $m \times m$ genotype matrix.

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{Z} & \mathbf{0} \\ \mathbf{0} & \mathbf{Z} \end{bmatrix} \begin{bmatrix} \mathbf{g}_1 \\ \mathbf{g}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix} \quad [1]$$

To estimate the random SNP effects more efficiently, we block sampling trait specific

effects by SNP, such that we have a vector $\mathbf{g} = [g_{11}, g_{21}, \dots, g_{1k}, g_{2k}, \dots, g_{1m}, g_{2m}]'$ where

$\mathbf{g}_k = [g_{1k}, g_{2k}]'$ is random SNP effects on trait 1 and trait 2 for k^{th} SNP. Let $\boldsymbol{\beta} = [\boldsymbol{\beta}_1 \ \boldsymbol{\beta}_2]'$

and $\mathbf{y} = [\mathbf{y}_1 \ \mathbf{y}_2]'$, based on the description of priors for IWBayesA in methods, we can

write the joint posterior density for all unknown parameters as follows:

$$\begin{aligned}
& p(\boldsymbol{\beta}, \mathbf{g}, \mathbf{G}, v_g, \boldsymbol{\Sigma}_g, \boldsymbol{\Sigma}_e | \mathbf{y}) \\
& \propto p(\mathbf{y} | \boldsymbol{\beta}, \mathbf{g}, \mathbf{G}, v_g, \boldsymbol{\Sigma}_g, \boldsymbol{\Sigma}_e) p(\boldsymbol{\beta}) \\
& \left(\prod_{k=1}^m p(\mathbf{g}_k | \mathbf{G}_k) \right) \left(\prod_{k=1}^m p(\mathbf{G}_k | v_g, \boldsymbol{\Sigma}_g) \right) p(\boldsymbol{\Sigma}_g | v_0, \Sigma_0) p(v_g) p(\boldsymbol{\Sigma}_e)
\end{aligned} \tag{2}$$

where \mathbf{G} is a $2m \times 2m$ genetic variance-covariance matrix and $\boldsymbol{\Sigma}_g$ is a 2×2 scale matrix

for trait specific random SNP effects. We assume that \mathbf{G} has an Inverted Wishart (IW)

prior of degree freedom v_g and scale matrix $\boldsymbol{\Sigma}_g$. Residuals for the two traits $\mathbf{e} = [\mathbf{e}_1 \ \mathbf{e}_2]'$

were assumed to have a bivariate normal distribution with null mean and variance-

covariance matrix $\mathbf{R} = \begin{bmatrix} \mathbf{I}\sigma_{e_1}^2 & \mathbf{I}\sigma_{e_1e_2} \\ \mathbf{I}\sigma_{e_1e_2} & \mathbf{I}\sigma_{e_2}^2 \end{bmatrix}$. Thus, we know the inverse of the residual

variance-covariance matrix is $\mathbf{R}^{-1} = \begin{bmatrix} \mathbf{I}r^{11} & \mathbf{I}r^{12} \\ \mathbf{I}r^{12} & \mathbf{I}r^{22} \end{bmatrix}$ where $\begin{bmatrix} r^{11} & r^{12} \\ r^{12} & r^{22} \end{bmatrix} = \begin{bmatrix} \sigma_{e_1}^2 & \sigma_{e_1e_2} \\ \sigma_{e_1e_2} & \sigma_{e_2}^2 \end{bmatrix}^{-1}$.

To sample fixed effects on trait 1, FCD for l^{th} element of $\boldsymbol{\beta}_1$ can be written as follows:

$$\beta_{1l} | \mathbf{y}, \boldsymbol{\beta}_{-1l}, \mathbf{g}_1, \mathbf{g}_2, \mathbf{G}, \boldsymbol{\Sigma}_e \sim N(\tilde{\beta}_{1l}, \tilde{v}_{\beta_{1l}}) \quad \text{with}$$

$$\begin{aligned}
\tilde{\beta}_{1l} &= \frac{\mathbf{x}_{.l}^{(1)'} \left((r^{11}\mathbf{y}_1 + r^{12}\mathbf{y}_2 - r^{11}\mathbf{X}_1\boldsymbol{\beta}_1 - r^{12}\mathbf{X}_2\boldsymbol{\beta}_2 - r^{11}\mathbf{Z}\mathbf{g}_1 - r^{12}\mathbf{Z}\mathbf{g}_2) + r^{11}\mathbf{x}_{.l}^{(1)}\beta_{1l} \right)}{\mathbf{x}_{.l}^{(1)'} \mathbf{x}_{.l}^{(1)} r^{11}} \\
&= \frac{\mathbf{x}_{.l}^{(1)'} \left((r^{11}(\mathbf{y}_1 - \mathbf{X}_1\boldsymbol{\beta}_1 - \mathbf{Z}\mathbf{g}_1) + r^{12}(\mathbf{y}_2 - \mathbf{X}_2\boldsymbol{\beta}_2 - \mathbf{Z}\mathbf{g}_2)) + r^{11}\mathbf{x}_{.l}^{(1)}\beta_{1l} \right)}{\mathbf{x}_{.l}^{(1)'} \mathbf{x}_{.l}^{(1)} r^{11}} \\
&= \frac{\mathbf{x}_{.l}^{(1)'} (r^{11}\mathbf{e}_1 + r^{12}\mathbf{e}_2 + r^{11}\mathbf{x}_{.l}^{(1)}\beta_{1l})}{\mathbf{x}_{.l}^{(1)'} \mathbf{x}_{.l}^{(1)} r^{11}} \\
&= \frac{r^{11} \sum_{i=1}^n x_{1il} e_{1i} + r^{12} \sum_{i=1}^n x_{1il} e_{2i} + r^{11} \left(\sum_{i=1}^n x_{1il}^2 \right) \beta_{1l}}{r^{11} \left(\sum_{i=1}^n x_{1il}^2 \right)}
\end{aligned}$$

$$\text{and } \tilde{v}_{\beta_{1l}} = \left(\mathbf{x}_{.l}^{(1)'} \mathbf{x}_{.l}^{(1)} r^{11} \right)^{-1} = \frac{\left(\sum_{i=1}^n x_{1il}^2 \right)^{-1}}{r^{11}}$$

Immediately after sampling β_{1l} , we update the first trait residual using

$\mathbf{e}_1 = \mathbf{e}_1 - \mathbf{x}_{.l}^{(1)} \left(\beta_{1l}^{[t+1]} - \beta_{1l}^{[t]} \right)$, where $\mathbf{x}_{.l}^{(1)}$ is the covariate/dummy variable values for fixed effect variable l on trait 1, $\beta_{1l}^{[t+1]}$ is the sampled β_{1l} value at cycle $[t+1]$ and $\beta_{1l}^{[t]}$ is the sampled β_{1l} value at cycle $[t]$.

To sample fixed effects on trait 2, FCD for l^{th} element of $\boldsymbol{\beta}_2$ can be written as follows:

$$\beta_{2l} \mid \mathbf{y}, \boldsymbol{\beta}_{-2l}, \mathbf{g}_1, \mathbf{g}_2, \mathbf{G}, \boldsymbol{\Sigma}_e \sim N\left(\tilde{\beta}_{2l}, \tilde{v}_{\beta_{2l}}\right) \text{ with}$$

$$\begin{aligned}
\tilde{\beta}_{2l} &= \frac{\mathbf{x}_{.l}^{(2)'} \left((r^{21} \mathbf{y}_1 + r^{22} \mathbf{y}_2 - r^{21} \mathbf{X}_1 \boldsymbol{\beta}_1 - r^{22} \mathbf{X}_2 \boldsymbol{\beta}_2 - r^{21} \mathbf{Z} \mathbf{g}_1 - r^{22} \mathbf{Z} \mathbf{g}_2) + \mathbf{x}_{.l}^{(2)} \beta_{2l} \right)}{\mathbf{x}_{.l}^{(2)'} \mathbf{x}_{.l}^{(2)} r^{22}} \\
&= \frac{\mathbf{x}_{.l}^{(2)'} \left((r^{12} (\mathbf{y}_1 - \mathbf{X}_1 \boldsymbol{\beta}_1 - \mathbf{Z} \mathbf{g}_1) + r^{22} (\mathbf{y}_2 - \mathbf{X}_2 \boldsymbol{\beta}_2 - \mathbf{Z} \mathbf{g}_2)) + r^{22} \mathbf{x}_{.l}^{(2)} \beta_{2l} \right)}{\mathbf{x}_{.l}^{(2)'} \mathbf{x}_{.l}^{(2)} r^{22}} \\
&= \frac{\mathbf{x}_{.l}^{(2)'} (r^{12} \mathbf{e}_1 + r^{22} \mathbf{e}_2 + r^{22} \mathbf{x}_{.l}^{(2)} \beta_{2l})}{\mathbf{x}_{.l}^{(2)'} \mathbf{x}_{.l}^{(2)} r^{22}} = \frac{r^{12} \mathbf{x}_{.l}^{(2)} \mathbf{e}_1 + r^{22} \mathbf{x}_{.l}^{(2)} \mathbf{e}_1 + r^{22} \mathbf{x}_{.l}^{(2)'} \mathbf{x}_{.l}^{(2)} \beta_{2l}}{\mathbf{x}_{.l}^{(2)'} \mathbf{x}_{.l}^{(2)} r^{22}} \\
&= \frac{r^{12} \sum_{i=1}^n x_{2il} e_{1i} + r^{22} \sum_{i=1}^n x_{1il} e_{2i} + r^{22} \left(\sum_{i=1}^n x_{2il}^2 \right) \beta_{2l}}{r^{22} \left(\sum_{i=1}^n x_{2il}^2 \right)}
\end{aligned}$$

$$\text{and } \tilde{v}_{\beta_{2l}} = \left(\mathbf{x}_{.l}^{(2)'} \mathbf{x}_{.l}^{(2)} r^{22} \right)^{-1} = \frac{\left(\sum_{i=1}^n x_{2il}^2 \right)^{-1}}{r^{22}}$$

Immediately after sampling β_{2l} , we update the second trait residual using

$\mathbf{e}_2 = \mathbf{e}_2 - \mathbf{x}_{.l}^{(2)} \left(\beta_{2l}^{[t+1]} - \beta_{2l}^{[t]} \right)$, where $\mathbf{x}_{.l}^{(2)}$ is the covariate/dummy variable values for fixed effect variable l on trait 2, $\beta_{2l}^{[t+1]}$ is the sampled β_{2l} value at cycle $[t+1]$ and $\beta_{2l}^{[t]}$ is the sampled β_{2l} value at cycle $[t]$.

To sample trait specific random SNP effects, FCD for k^{th} element of \mathbf{g} can be written as follows:

$$\mathbf{g}_k = \begin{bmatrix} g_{1k} \\ g_{2k} \end{bmatrix} | \mathbf{y}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \mathbf{g}_{-k}, \{\mathbf{G}_k\}_{k=1}^m, \boldsymbol{\Sigma}_e \sim N(\tilde{\mathbf{g}}_k, \tilde{\mathbf{V}}_{g_k})$$

Where

$$\tilde{\mathbf{g}}_k = \left(\begin{bmatrix} r^{11} \mathbf{z}'_{.k} \mathbf{z}_{.k} & r^{12} \mathbf{z}'_{.k} \mathbf{z}_{.k} \\ r^{21} \mathbf{z}'_{.k} \mathbf{z}_{.k} & r^{22} \mathbf{z}'_{.k} \mathbf{z}_{.k} \end{bmatrix} + \mathbf{G}_k^{-1} \right)^{-1} \begin{bmatrix} \mathbf{z}'_{.k} (r^{11} \mathbf{e}_1 + r^{12} \mathbf{e}_2 + r^{11} \mathbf{z}_{.k} g_{1k} + r^{12} \mathbf{z}_{.k} g_{2k}) \\ \mathbf{z}'_{.k} (r^{21} \mathbf{e}_1 + r^{22} \mathbf{e}_2 + r^{12} \mathbf{z}_{.k} g_{1k} + r^{22} \mathbf{z}_{.k} g_{2k}) \end{bmatrix}$$

and

$$\tilde{\mathbf{V}}_{g_k} = \left(\begin{bmatrix} r^{11} \mathbf{z}'_{.k} \mathbf{z}_{.k} & r^{12} \mathbf{z}'_{.k} \mathbf{z}_{.k} \\ r^{21} \mathbf{z}'_{.k} \mathbf{z}_{.k} & r^{22} \mathbf{z}'_{.k} \mathbf{z}_{.k} \end{bmatrix} + \mathbf{G}_k^{-1} \right)^{-1}$$

Immediately after sampling \mathbf{g}_k , we update the first trait residual using

$$\mathbf{e}_1 = \mathbf{e}_1 - \mathbf{z}_{.k} (g_{1k}^{[t+1]} - g_{1k}^{[t]}) \text{ and update the second trait residual using}$$

$$\mathbf{e}_2 = \mathbf{e}_2 - \mathbf{z}_{.k} (g_{2k}^{[t+1]} - g_{2k}^{[t]}), \text{ where } g_{1k}^{[t+1]} (g_{2k}^{[t+1]}) \text{ is the sampled } g_{1k} (g_{2k}) \text{ value at cycle [t+1] and } g_{1k}^{[t]} (g_{2k}^{[t]}) \text{ is the sampled } g_{1k} (g_{2k}) \text{ value at cycle [t].}$$

To sample genetic variance-covariance matrix for random SNP effects, given the specified conjugate prior on genetic variance-covariance matrix

$$p(\mathbf{G}_k | v_g, \Sigma_g) \propto \text{IW}(v_g, \Sigma_g) \text{ for } k^{th} \text{ SNP, FCD of } \mathbf{G}_k \text{ can be written as follows:}$$

$$\begin{aligned} p(\mathbf{G}_k | else) &\propto p(\mathbf{g}_k | \mathbf{G}_k) p(\mathbf{G}_k | v_g, \Sigma_g) \\ &\propto |\mathbf{G}_k|^{-1/2} \exp(-0.5(\mathbf{g}'_k \mathbf{G}_k^{-1} \mathbf{g}_k)) |\mathbf{G}_k|^{-\frac{1}{2}(v_g+3)} \exp(-0.5 \text{trace}(\mathbf{G}_k^{-1} \Sigma_g)) \\ &\propto |\mathbf{G}_k|^{-\frac{1}{2}((v_g+1)+3)} \exp\left(-0.5 \text{trace}\left(\mathbf{G}_k^{-1} \left(\begin{bmatrix} g_{1k}^2 & g_{1k} g_{2k} \\ g_{1k} g_{2k} & g_{2k}^2 \end{bmatrix} + \Sigma_g \right)\right)\right) \end{aligned}$$

$$\text{Hence, } \mathbf{G}_k | else \sim \text{IW}\left(v_g + 1, \left(\begin{bmatrix} g_{1k}^2 & g_{1k} g_{2k} \\ g_{1k} g_{2k} & g_{2k}^2 \end{bmatrix} + \Sigma_g \right)\right)$$

To sample scale matrix for genetic variance-covariance matrices, given a conjugate Wishart prior for scale matrix $p(\Sigma_g) \propto W(v_0, \Sigma_0 / v_0)$, FCD of Σ_g can be written as follows:

$$\begin{aligned}
p(\Sigma_g | else) &\propto \prod_{k=1}^m p(\mathbf{G}_k | v_g, \Sigma_g) W(v_0, \Sigma_0 / v_0) \\
&\propto \left(\prod_{k=1}^m |\Sigma_g|^{v_g/2} \exp(-0.5 \text{trace}(\mathbf{G}_k^{-1} \Sigma_g)) \right) |\Sigma_g|^{\frac{1}{2}(v_0-3)} \exp(-0.5 \text{trace}(\Sigma_g \Sigma_0^{-1} v_0)) \\
&\propto |\Sigma_g|^{v_g m/2} \exp\left(-0.5 \text{trace}\left(\Sigma_g \sum_{j=1}^m \mathbf{G}_j^{-1}\right)\right) |\Sigma_g|^{\frac{1}{2}(v_0-3)} \exp(-0.5 \text{trace}(\Sigma_g \Sigma_0^{-1} v_0)) \\
&\propto |\Sigma_g|^{\frac{1}{2}(v_0+v_g m-3)} \exp\left(-0.5 \text{trace}\left(\Sigma_g \left(\sum_{k=1}^m \mathbf{G}_k^{-1} + v_0 \Sigma_0^{-1}\right)\right)\right)
\end{aligned}$$

$$\text{Hence, } \Sigma_g | else \sim W\left(v_0 + v_g m, \left(\sum_{k=1}^m \mathbf{G}_k^{-1} + v_0 \Sigma_0^{-1}\right)^{-1}\right)$$

To sample degrees of freedom for genetic variance-covariance matrices, given a “non-informative” prior on degrees of freedom $p(v_g) \propto 1/(1+v_g)^2$, we can write FCD for v_g as follows:

$$p(v_g | else) \propto 2^{-v_g m} \Gamma_2(v_g / 2)^{-m} |\Sigma_g|^{v_g m/2} \prod_{j=1}^m |\mathbf{G}_j|^{-\frac{1}{2}(v_g+3)} \exp(-0.5 \text{trace}(\mathbf{G}_j^{-1} \Sigma_g)) \frac{1}{(1+v_g)^2}$$

$$\text{where } \Gamma_2(v_g / 2) = \pi^{2(2-1)/4} \prod_{p=1}^2 \Gamma[v_g / 2 + (1-p) / 2] \propto \Gamma[v_g / 2] \Gamma[v_g / 2 - 1 / 2]$$

The FCD for v_g is not recognizable and so a sampling strategy for nonstandard distributions is required. In order to use proposal densities, especially in a random walk Metropolis implementation, it may be more appropriate to change the variable such that

its parameter space is defined on the real line. Using $\zeta = \log(v_g)$, the relevant FCD in this case is:

$$p(\zeta | else) \propto \left(0.5|\Sigma_g|\right)^{m \exp(\zeta)} \left(\Gamma[\exp(\zeta)/2] \Gamma[\exp(\zeta)/2 - 1/2]\right)^{-m} \\ \prod_{j=1}^m |\mathbf{G}_j|^{-\frac{1}{2}(\exp(\zeta)+3)} \exp\left(-0.5 \text{trace}(\mathbf{G}_j^{-1} \Sigma_g)\right) \frac{1}{(1 + \exp(\zeta))^2} \exp(\zeta)$$

Hence,

$$\log p(\zeta | else) = \\ m \left(\exp(\zeta) (\log(0.5) + 0.5 \log |\Sigma_g|) \right. \\ \left. - \log \Gamma[\exp(\zeta)/2] - \log \Gamma[\exp(\zeta)/2 - 0.5] \right) \\ - 0.5(\exp(\zeta) + 3) \sum_{j=1}^m \log |\mathbf{G}_j| - 0.5 \sum_{j=1}^m \text{trace}(\Sigma_g \mathbf{G}_j^{-1}) - 2 \log(1 + \exp(\zeta)) + \zeta$$

We sampled v_g using a random walk Metropolis Hastings algorithm, which was described in other non-genomic applications involving the sampling of degrees of freedom parameter (Kizilkaya & Tempelman 2005; Bello 2010). Suppose the value of ζ in the current cycle i is $\zeta^{[i]}$. Then generate a random variable, say δ from $N(0, c^2)$ and add it to $\zeta^{[i]}$ to propose $\zeta^* = \zeta^{[i]} + \delta$. Determine the following odds ratio:

$$\alpha = p(\zeta^* | else) / p(\zeta^{[i]} | else). \text{ For numerical stability, it is perhaps much wiser to}$$

evaluate this ratio as: $\alpha = \exp(\log p(\zeta^{[i+1]} | else) - \log p(\zeta^{[i]} | else))$. To implement this

Metropolis (within Gibbs) scheme, one would first generate U from a Uniform(0, 1)

distribution. If 1) $\alpha > 1$, accept $\zeta^{[i+1]} = \zeta^*$; 2) If $\alpha > U$, accept $\zeta^{[i+1]} = \zeta^*$; 3) If $\alpha < U$,

then set $\zeta^{[i+1]} = \zeta^{[i]}$.

To sample residual variance-covariance matrix, given a specified Inverted

Wishart prior $p(\Sigma_e | \nu_0, \Sigma_0) = \text{IW}(\nu_0, \Sigma_0)$, we can write the FCD for Σ_e as follows,

$$\begin{aligned} p(\Sigma_e | ELSE, \mathbf{y}) &\propto p(\mathbf{y} | \boldsymbol{\beta}, \mathbf{g}, \Sigma_e) p(\Sigma_e | \nu_0, \Sigma_0) \\ &\propto |\Sigma_e|^{-\frac{n}{2}} \exp\left(-0.5 \left[\text{trace}(\Sigma_e^{-1} \mathbf{S}_e) \right]\right) |\Sigma_e|^{-\frac{1}{2}(\nu_0+3)} \exp\left(-0.5 \text{trace}(\Sigma_e^{-1} \Sigma_0^{-1})\right) \\ &\propto |\Sigma_e|^{-\frac{1}{2}(n+\nu_0+3)} \exp\left(-0.5 \left[\text{trace}(\Sigma_e^{-1} (\mathbf{S}_e + \Sigma_0^{-1})) \right]\right) \end{aligned}$$

$$\text{where } \mathbf{S}_e = \begin{bmatrix} \mathbf{e}_1' \mathbf{e}_1 & \mathbf{e}_1' \mathbf{e}_2 \\ \mathbf{e}_2' \mathbf{e}_1 & \mathbf{e}_2' \mathbf{e}_2 \end{bmatrix}$$

$$\text{Hence, } \Sigma_e | else \sim \text{IW}\left(n + \nu_0, (\mathbf{S}_e + \Sigma_0^{-1})\right)$$

D1.2 MCMC Implementation Strategy for CD-BayesA/CD-BayesB

As defined in the methods, a square root free Cholesky decomposition (CD) can be applied to the genetic variance-covariance matrices. Thereby, if let $\mathbf{g}_2 = \mathbf{\Psi}\mathbf{g}_1 + \mathbf{g}_{2||}$ where $\mathbf{g}_{2||}$ is the vector of SNP substitution effects for trait 2 conditional on trait 1 and $\mathbf{\Psi} = \text{diag}\{\phi_j\}_{j=1}^m$ represents a diagonal matrix of SNP-specific associations between two traits. Based on the description of priors for the two CD models in methods, we can write down the joint posterior density for all unknown parameters:

$$\begin{aligned} & p(\boldsymbol{\beta}, \mathbf{g}_1, \mathbf{g}_{2||}, \boldsymbol{\sigma}_{\mathbf{g}_1}, \boldsymbol{\sigma}_{\mathbf{g}_{2||}}, \boldsymbol{\phi}, \mu_\phi, \sigma_\phi^2, v_1, v_2, s_1^2, s_2^2, \pi_1, \pi_2, \pi_\phi, \boldsymbol{\Sigma}_e | \mathbf{y}) \\ & \propto p(\mathbf{y} | \boldsymbol{\beta}, \mathbf{g}_1, \mathbf{g}_{2||}, \boldsymbol{\sigma}_{\mathbf{g}_1}, \boldsymbol{\sigma}_{\mathbf{g}_{2||}}, \boldsymbol{\phi}, \mu_\phi, \sigma_\phi^2, v_1, v_2, s_1^2, s_2^2, \pi_1, \pi_2, \pi_\phi, \boldsymbol{\Sigma}_e) p(\boldsymbol{\beta}) \\ & \left(\prod_{k=1}^m p(g_{1k} | \sigma_{\mathbf{g}_{1k}}^2) \right) \left(\prod_{k=1}^m p(\sigma_{\mathbf{g}_{1k}}^2 | v_1, s_1^2, \pi_1) \right) \left(\prod_{k=1}^m p(g_{2||k} | \sigma_{\mathbf{g}_{2||k}}^2) \right) \left(\prod_{j=1}^m p(\sigma_{\mathbf{g}_{2||j}}^2 | v_2, s_2^2, \pi_2) \right) \\ & \left(\prod_{k=1}^m p(\phi_k | \mu_\phi, \sigma_\phi^2, \pi_\phi) \right) p(\mu_\phi) p(\sigma_\phi^2) p(v_1) p(v_2) p(s_1^2) p(s_2^2) p(\pi_1) p(\pi_2) p(\pi_\phi) p(\boldsymbol{\Sigma}_e) \end{aligned}$$

Where $\boldsymbol{\sigma}_{\mathbf{g}_1} = [\sigma_{g_{11}}^2 \quad \sigma_{g_{12}}^2 \quad \sigma_{g_{13}}^2 \quad \dots \quad \sigma_{g_{1m}}^2]'$ is a vector of SNP specific variances for trait 1,

$\boldsymbol{\sigma}_{\mathbf{g}_{2||}} = [\sigma_{g_{2||,1}}^2 \quad \sigma_{g_{2||,2}}^2 \quad \sigma_{g_{2||,3}}^2 \quad \dots \quad \sigma_{g_{2||,m}}^2]'$ is a vector of SNP specific variances for trait 2

conditional on trait 1, $\boldsymbol{\phi} = [\phi_1 \quad \phi_2 \quad \dots \quad \phi_m]'$ is a vector of SNP specific association

parameters between two traits. In CDBayesB ($\pi_1 < 1$, $\pi_2 < 1$ and $\pi_\phi < 1$), $\sigma_{\mathbf{g}_{1k}}^2$ ($\sigma_{\mathbf{g}_{2||k}}^2$)

has a mixture prior of point mass at zero with probability $1 - \pi_1$ ($1 - \pi_2$) and scaled

inverted Chi-square distribution of degree freedom v_1 (v_2) and scale parameter s_1^2 (s_2^2)

with probability π_1 (π_2). In CDBayesA ($\pi_1 = 1$, $\pi_2 = 1$ and $\pi_\phi = 1$), $\sigma_{\mathbf{g}_{1j}}^2$ ($\sigma_{\mathbf{g}_{2||j}}^2$) has a

scaled inverted Chi-square distribution of degree freedom $v_1(v_2)$ and scale parameter s_1^2 (s_2^2). To sample location parameters computationally efficiently, we again adopted Gauss-Seidel updating algorithm in MCMC implementation strategy (Legarra and Misztal, 2008).

To sample fixed effects on trait 1 in CDBayesA\CDBayesB, FCD for l^{th} element of β_1 can be written as follows:

$$\beta_{1l} | \mathbf{y}, \beta_{-1l}, \mathbf{g}_1, \mathbf{g}_{2l}, \{\sigma_{g1k}^2\}_{k=1}^m, \{\sigma_{g2lk}^2\}_{k=1}^m, \Sigma_e, \{\phi_k\}_{k=1}^m \sim N(\tilde{\beta}_{1l}, \tilde{v}_{\beta_{1l}})$$

with

$$\begin{aligned} \tilde{\beta}_{1l} &= \frac{\mathbf{x}_{.l}^{(1)'} \left((r^{11} \mathbf{y}_1 + r^{12} \mathbf{y}_2 - r^{11} \mathbf{X}_1 \beta_1 - r^{12} \mathbf{X}_2 \beta_2 - r^{11} \mathbf{Z} \mathbf{g}_1 - r^{12} \mathbf{Z} \Psi \mathbf{g}_1 - r^{12} \mathbf{Z} \mathbf{g}_{2l}) + r^{11} \mathbf{x}_{.l}^{(1)} \beta_{1l} \right)}{\mathbf{x}_{.l}^{(1)'} \mathbf{x}_{.l}^{(1)} r^{11}} \\ &= \frac{\mathbf{x}_{.l}^{(1)'} \left((r^{11} (\mathbf{y}_1 - \mathbf{X}_1 \beta_1 - \mathbf{Z} \mathbf{g}_1) + r^{12} (\mathbf{y}_2 - \mathbf{X}_2 \beta_2 - \mathbf{Z} \Psi \mathbf{g}_1 - \mathbf{Z} \mathbf{g}_{2l})) + r^{11} \mathbf{x}_{.l}^{(1)} \beta_{1l} \right)}{\mathbf{x}_{.l}^{(1)'} \mathbf{x}_{.l}^{(1)} r^{11}} \\ &= \frac{\mathbf{x}_{.l}^{(1)'} (r^{11} \mathbf{e}_1 + r^{12} \mathbf{e}_2 + r^{11} \mathbf{x}_{.l}^{(1)} \beta_{1l})}{\mathbf{x}_{.l}^{(1)'} \mathbf{x}_{.l}^{(1)} r^{11}} \\ &= \frac{r^{11} \sum_{i=1}^n x_{1il} e_{1i} + r^{12} \sum_{i=1}^n x_{1il} e_{2i} + r^{11} \left(\sum_{i=1}^n x_{1il}^2 \right) \beta_{1l}}{r^{11} \left(\sum_{i=1}^n x_{1il}^2 \right)} \end{aligned}$$

$$\text{and } \tilde{v}_{\beta_{1l}} = \left(\mathbf{x}_{.l}^{(1)'} \mathbf{x}_{.l}^{(1)} r^{11} \right)^{-1} = \frac{\left(\sum_{i=1}^n x_{1il}^2 \right)^{-1}}{r^{11}}$$

Immediately after sampling β_{1l} , we update the first trait residual using

$$\mathbf{e}_1 = \mathbf{e}_1 - \mathbf{x}_{.l}^{(1)} \left(\beta_{1l}^{[t+1]} - \beta_{1l}^{[t]} \right), \text{ where } \mathbf{x}_{.l}^{(1)} \text{ is the covariate/dummy variable values for fixed}$$

effect variable l on trait 1, $\beta_{1l}^{[t+1]}$ is the sampled β_{1l} value at cycle $[t+1]$ and $\beta_{1l}^{[t]}$ is the sampled β_{1l} value at cycle $[t]$.

To sample fixed effects on trait 2 in CDBayesA\CDBayesB, FCD for l^{th} element of β_2 can be written as follows:

$$\beta_{2l} | \mathbf{y}, \beta_{-2l}, \mathbf{g}_1, \mathbf{g}_{2l}, \{\sigma_{g1k}^2\}_{k=1}^m, \{\sigma_{g2lk}^2\}_{k=1}^m, \sigma_e^2, \{\phi_k\}_{k=1}^m \sim N(\tilde{\beta}_{2l}, \tilde{v}_{\beta 2l})$$

Where

$$\begin{aligned} \tilde{\beta}_{2l} &= \frac{\mathbf{x}_{.l}^{(2)'} \left((r^{12} \mathbf{y}_1 + r^{22} \mathbf{y}_2 - r^{12} \mathbf{X}_1 \beta_1 - r^{22} \mathbf{X}_2 \beta_2 - r^{12} \mathbf{Z} \mathbf{g}_1 - r^{22} \mathbf{Z} \Psi \mathbf{g}_1 - r^{22} \mathbf{Z} \mathbf{g}_{2l}) + r^{22} \mathbf{x}_{.l}^{(2)} \beta_{2l} \right)}{\mathbf{x}_{.l}^{(2)'} \mathbf{x}_{.l}^{(2)} r^{22}} \\ &= \frac{\mathbf{x}_{.l}^{(2)'} \left((r^{12} (\mathbf{y}_1 - \mathbf{X}_1 \beta_1 - \mathbf{Z} \mathbf{g}_1) + r^{22} (\mathbf{y}_2 - \mathbf{X}_2 \beta_2 - \mathbf{Z} \Psi \mathbf{g}_1 - \mathbf{Z} \mathbf{g}_{2l})) + r^{22} \mathbf{x}_{.l}^{(2)} \beta_{2l} \right)}{\mathbf{x}_{.l}^{(2)'} \mathbf{x}_{.l}^{(2)} r^{22}} \\ &= \frac{\mathbf{x}_{.l}^{(2)'} (r^{12} \mathbf{e}_1 + r^{22} \mathbf{e}_2 + r^{22} \mathbf{x}_{.l}^{(2)} \beta_{2l})}{\mathbf{x}_{.l}^{(2)'} \mathbf{x}_{.l}^{(2)} r^{22}} \\ &= \frac{r^{12} \sum_{i=1}^n x_{2il} e_{1i} + r^{22} \sum_{i=1}^n x_{1il} e_{2i} + r^{22} \left(\sum_{i=1}^n x_{2il}^2 \right) \beta_{2l}}{r^{22} \left(\sum_{i=1}^n x_{2il}^2 \right)} \end{aligned}$$

$$\text{and } \tilde{v}_{\beta 2l} = \left(\mathbf{x}_{.l}^{(2)'} \mathbf{x}_{.l}^{(2)} r^{22} \right)^{-1} = \frac{\left(\sum_{i=1}^n x_{2il}^2 \right)^{-1}}{r^{22}}$$

Immediately after sampling β_{2l} , we update the second trait residual using

$$\mathbf{e}_2 = \mathbf{e}_2 - \mathbf{x}_{.l}^{(2)} \left(\beta_{2l}^{[t+1]} - \beta_{2l}^{[t]} \right), \text{ where } \mathbf{x}_{.l}^{(2)} \text{ is the covariate/dummy variable values for fixed}$$

effect variable l on trait 2, $\beta_{2l}^{[t+1]}$ is the sampled β_{2l} value at cycle $[t+1]$ and $\beta_{2l}^{[t]}$ is the

sampled β_{2l} value at cycle $[t]$.

To sample random SNP effects for trait 1 in CDBayesA, FCD for k^{th} element of \mathbf{g}_1 can be written as follows:

$$g_{1k} \mid \mathbf{y}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \mathbf{g}_{-1k}, \mathbf{g}_{2||}, \{\sigma_{g1k}^2\}_{k=1}^m, \{\sigma_{g2||k}^2\}_{k=1}^m, \boldsymbol{\Sigma}_e, \{\phi_k\}_{k=1}^m \sim N(\tilde{g}_{1k}, \tilde{v}_{1k})$$

Where

$$\begin{aligned} \tilde{g}_{1k} &= \frac{\left(\mathbf{z}_{\cdot,k}' \left((r^{11} \mathbf{y}_1 + r^{12} \mathbf{y}_2 - r^{11} \mathbf{X}_1 \boldsymbol{\beta}_1 - r^{12} \mathbf{X}_2 \boldsymbol{\beta}_2 - r^{11} \mathbf{Z} \mathbf{g}_1 - r^{12} \mathbf{Z} \Psi \mathbf{g}_1 - r^{12} \mathbf{Z} \mathbf{g}_{2||} + r^{11} \mathbf{z}_{\cdot,k} g_{1k} + r^{12} \phi_k \mathbf{z}_{\cdot,k} g_{1k}) \right) \right. \\ &\quad \left. + \phi_k \mathbf{z}_{\cdot,k}' (r^{12} \mathbf{y}_1 + r^{22} \mathbf{y}_2 - r^{12} \mathbf{X}_1 \boldsymbol{\beta}_1 - r^{22} \mathbf{X}_2 \boldsymbol{\beta}_2 - r^{12} \mathbf{Z} \mathbf{g}_1 - r^{22} \mathbf{Z} \Psi \mathbf{g}_1 - r^{22} \mathbf{Z} \mathbf{g}_{2||} + r^{12} \mathbf{z}_{\cdot,k} g_{1k} + r^{22} \phi_k \mathbf{z}_{\cdot,k} g_{1k}) \right) \\ &\quad \left. r^{11} \mathbf{z}_{\cdot,k}' \mathbf{z}_{\cdot,k} + r^{12} \phi_k \mathbf{z}_{\cdot,k}' \mathbf{z}_{\cdot,k} + r^{12} \phi_k \mathbf{z}_{\cdot,k}' \mathbf{z}_{\cdot,k} + r^{22} \phi_k \mathbf{z}_{\cdot,k}' \phi_k \mathbf{z}_{\cdot,k} + \sigma_{g1k}^{-2} \right) \\ &= \frac{\left(\mathbf{z}_{\cdot,k}' (r^{11} \mathbf{e}_1 + r^{12} \mathbf{e}_2 + r^{11} \mathbf{z}_{\cdot,k} g_{1k} + r^{12} \phi_k \mathbf{z}_{\cdot,k} g_{1k}) \right. \\ &\quad \left. + \phi_k \mathbf{z}_{\cdot,k}' (r^{12} \mathbf{e}_1 + r^{22} \mathbf{e}_2 + r^{12} \mathbf{z}_{\cdot,k} g_{1k} + r^{22} \phi_k \mathbf{z}_{\cdot,k} g_{1k}) \right) \\ &\quad \left. r^{11} \mathbf{z}_{\cdot,k}' \mathbf{z}_{\cdot,k} + r^{12} \phi_k \mathbf{z}_{\cdot,k}' \mathbf{z}_{\cdot,k} + r^{12} \phi_k \mathbf{z}_{\cdot,k}' \mathbf{z}_{\cdot,k} + r^{22} \phi_k \mathbf{z}_{\cdot,k}' \phi_k \mathbf{z}_{\cdot,k} + \sigma_{g1k}^{-2} \right) \\ &= \frac{\left(r^{11} \sum_{i=1}^n z_{ik} e_{1i} + r^{12} \sum_{i=1}^n z_{ik} e_{2i} + r^{11} \left(\sum_{i=1}^n z_{ik}^2 \right) g_{1k} + r^{12} \left(\sum_{i=1}^n z_{ik}^2 \right) \phi_k g_{1k} \right. \\ &\quad \left. + r^{12} \phi_k \sum_{i=1}^n z_{ik} e_{1i} + r^{22} \phi_k \sum_{i=1}^n z_{ik} e_{2i} + r^{12} \phi_k \left(\sum_{i=1}^n z_{ik}^2 \right) g_{1k} + r^{22} \phi_k^2 \left(\sum_{i=1}^n z_{ik}^2 \right) g_{1k} \right) \\ &\quad \left. (r^{11} + 2r^{12} \phi_k + r^{22} \phi_k^2) \left(\sum_{i=1}^n z_{ik}^2 \right) + \sigma_{g1k}^{-2} \right) \\ &= \frac{\left((r^{11} + r^{12} \phi_k) \left(\sum_{i=1}^n z_{ik} e_{1i} \right) + (r^{12} + r^{22} \phi_k) \sum_{i=1}^n z_{ik} e_{2i} + (r^{11} + 2r^{12} \phi_k + r^{22} \phi_k^2) \left(\sum_{i=1}^n z_{ik}^2 \right) g_{1k} \right) \\ &\quad \left. (r^{11} + 2r^{12} \phi_k + r^{22} \phi_k^2) \left(\sum_{i=1}^n z_{ik}^2 \right) + \sigma_{g1k}^{-2} \right) \end{aligned}$$

and

$$\tilde{v}_{1k} = \left((r^{11} + 2r^{12} \phi_k + r^{22} \phi_k^2) \left(\sum_{i=1}^n z_{ik}^2 \right) + \sigma_{g1k}^{-2} \right)^{-1}$$

Immediately after sampling g_{1k} , we update the first trait residual using

$\mathbf{e}_1 = \mathbf{e}_1 - \mathbf{z}_{.k} \left(g_{1k}^{[t+1]} - g_{1k}^{[t]} \right)$ and update the second trait residual using

$\mathbf{e}_2 = \mathbf{e}_2 - \phi_k \mathbf{z}_{.k} \left(g_{1k}^{[t+1]} - g_{1k}^{[t]} \right)$, where $g_{1k}^{[t+1]}$ is the sampled g_{1k} value at cycle [t+1] and

$g_{1k}^{[t]}$ is the sampled g_{1k} value at cycle [t].

To sample random SNP effects for trait 2 conditional on trait 1 in CDBayesA,

FCD for k^{th} element of $\mathbf{g}_{2||}$ can be written as follows:

$$g_{2||k} \mid \mathbf{y}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \mathbf{g}_1, \mathbf{g}_{-2||k}, \left\{ \sigma_{g1k}^2 \right\}_{k=1}^m, \left\{ \sigma_{g2||k}^2 \right\}_{k=1}^m, \sigma_e^2, \left\{ \phi_k \right\}_{k=1}^m \sim N \left(\tilde{g}_{2||k}, \tilde{v}_{2||k} \right)$$

where

$$\begin{aligned} \tilde{g}_{2||k} &= \frac{\mathbf{z}_{.k}' \left(\left(r^{12} \mathbf{y}_1 + r^{22} \mathbf{y}_2 - r^{12} \mathbf{X}_1 \boldsymbol{\beta}_1 - r^{22} \mathbf{X}_2 \boldsymbol{\beta}_2 - r^{12} \mathbf{Z} \mathbf{g}_1 - r^{22} \mathbf{Z} \Psi \mathbf{g}_1 - r^{22} \mathbf{Z} \mathbf{g}_{2||} + r^{22} \mathbf{z}_{.k} g_{2||k} \right) \right)}{r^{22} \mathbf{z}_{.k}' \mathbf{z}_{.k} + \sigma_{g2||k}^{-2}} \\ &= \frac{\mathbf{z}_{.k}' \left(r^{12} (\mathbf{y}_1 - \mathbf{X}_1 \boldsymbol{\beta}_1 - \mathbf{Z} \mathbf{g}_1) + r^{22} (\mathbf{y}_2 - \mathbf{X}_2 \boldsymbol{\beta}_2 - \mathbf{Z} \Psi \mathbf{g}_1 - \mathbf{Z} \mathbf{g}_{2||}) + r^{22} \mathbf{z}_{.k} g_{2||k} \right)}{r^{22} \mathbf{z}_{.k}' \mathbf{z}_{.k} + \sigma_{g2||k}^{-2}} \\ &= \frac{\mathbf{z}_{.k}' \left(r^{12} \mathbf{e}_1 + r^{22} \mathbf{e}_2 + r^{22} \mathbf{z}_{.k} g_{2||k} \right)}{r^{22} \mathbf{z}_{.k}' \mathbf{z}_{.k} + \sigma_{g2||k}^{-2}} \\ &= \frac{r^{12} \mathbf{z}_{.k}' \mathbf{e}_1 + r^{22} \mathbf{z}_{.k}' \mathbf{e}_2 + r^{22} \mathbf{z}_{.k}' \mathbf{z}_{.k} g_{2||k}}{r^{22} \mathbf{z}_{.k}' \mathbf{z}_{.k} + \sigma_{g2||k}^{-2}} \\ &= \frac{r^{12} \sum_{i=1}^n z_{ik} e_{1i} + r^{22} \sum_{i=1}^n z_{ik} e_{2i} + r^{22} \left(\sum_{i=1}^n z_{ik}^2 \right) g_{2||k}}{r^{22} \left(\sum_{i=1}^n z_{ik}^2 \right) + \sigma_{g2||k}^{-2}} \end{aligned}$$

and

$$\tilde{v}_{2||k} = \left(r^{22} \left(\sum_{i=1}^n z_{ik}^2 \right) + \sigma_{g2||k}^{-2} \right)^{-1}$$

Immediately after sampling g_{2k} , we update the second trait residual using

$\mathbf{e}_2 = \mathbf{e}_2 - \mathbf{z}_{.k} \left(g_{2|k}^{[t+1]} - g_{2|k}^{[t]} \right)$, where $g_{2|k}^{[t+1]}$ is the sampled $g_{2|k}$ value at cycle $[t+1]$ and

$g_{2|k}^{[t]}$ is the sampled $g_{2|k}$ value at cycle $[t]$.

To sample random association parameters in CDBayesA, FCD for ϕ_k

($k=1,2,\dots,m$) can be written as follows:

$$\phi_k | \mathbf{y}, \boldsymbol{\beta}, \mathbf{g}_1, \mathbf{g}_{2|1}, \left\{ \sigma_{g_{1k}}^2 \right\}_{k=1}^m, \left\{ \sigma_{g_{2|k}}^2 \right\}_{k=1}^m, \Sigma_e, \boldsymbol{\phi}_{-k} \sim N(\tilde{\phi}_k, \tilde{v}_{\phi k})$$

with

$$\begin{aligned} \tilde{\phi}_k &= \frac{\sigma_{e_2}^2 \sigma_{\phi}^{-2} \mu_{\phi} + \mathbf{z}_{.k}^g ' \left((\mathbf{y}_2 - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{g}_{2|1}) + \mathbf{z}_{.k}^g \phi_k \right)}{\mathbf{z}_{.k}^g ' \mathbf{z}_{.k}^g + \sigma_{e_2}^2 \sigma_{\phi}^{-2}} \\ &= \frac{\sigma_{e_2}^2 \sigma_{\phi}^{-2} \mu_{\phi} + \mathbf{z}_{.k}^g ' (\mathbf{e}_2 + \mathbf{z}_{.k}^g \phi_j)}{\mathbf{z}_{.k}^g ' \mathbf{z}_{.k}^g + \sigma_{e_2}^2 \sigma_{\phi}^{-2}} = \frac{\sigma_{e_2}^2 \sigma_{\phi}^{-2} \mu_{\phi} + \mathbf{z}_{.k}^g ' \mathbf{e}_2 + \mathbf{z}_{.k}^g ' \mathbf{z}_{.k}^g \phi_k}{\mathbf{z}_{.k}^g ' \mathbf{z}_{.k}^g + \sigma_{e_2}^2 \sigma_{\phi}^{-2}} \\ &= \frac{\sigma_{e_2}^2 \sigma_{\phi}^{-2} \mu_{\phi} + g_{1k} \sum_{i=1}^n z_{ik} e_{2j} + \left(g_{1k}^2 \sum_{i=1}^n (z_{ik})^2 \right) \phi_k}{\left(g_{1k}^2 \sum_{i=1}^n (z_{ik})^2 \right) + \sigma_{e_2}^2 \sigma_{\phi}^{-2}} \end{aligned}$$

and

$$\tilde{v}_{\phi k} = \left(\frac{\left(g_{1k}^2 \sum_{j=1}^n (z_{jk})^2 \right)}{\sigma_{e_2}^2} + \sigma_{\phi}^{-2} \right)^{-1}$$

Immediately after sampling ϕ_k , we update the second trait residual using

$\mathbf{e}_2 = \mathbf{e}_2 - g_{1k} \mathbf{z}_{.k} (\phi_k^{[t+1]} - \phi_k^{[t]})$, where $\mathbf{z}_{.k}^g = g_{1k} \mathbf{z}_{.k}$, $\phi_k^{[t+1]}$ is the sampled ϕ_k value at cycle [t+1] and $\phi_k^{[t]}$ is the sampled ϕ_k value at cycle [t].

To sampling SNP specific variances on trait 1 in CDBayesA,

given that the prior density is scaled inverted chi-square: $\sigma_{g_{1k}}^2 | v_1, s_1^2 \sim \chi^{-2}(v_1, v_1 s_1^2)$, FCD

for $\sigma_{g_{1k}}^2$ ($k=1,2,\dots,m$) can be derived as follows:

$$\begin{aligned} p(\sigma_{g_{1k}}^2 | \text{else}) &\propto p(g_{1k} | \sigma_{g_{1k}}^2) p(\sigma_{g_{1k}}^2 | v_1, s_1^2) \\ &\propto (2\pi\sigma_{g_{1k}}^2)^{-1/2} \exp\left(-\frac{g_{1k}^2}{2\sigma_{g_{1k}}^2}\right) \frac{\left(\frac{v_1 s_1^2}{2}\right)^{\frac{v_1}{2}}}{\Gamma\left(\frac{v_1}{2}\right)} \sigma_{g_{1k}}^{2 - \left(\frac{v_1}{2} + 1\right)} e^{-\frac{v_1 s_1^2}{2\sigma_{g_{1k}}^2}} \\ &\propto \sigma_{g_{1k}}^{2 - \left(\frac{(v_1+1)}{2} + 1\right)} e^{-\frac{v_1 s_1^2 + g_{1k}^2}{2\sigma_{g_{1k}}^2}} \end{aligned}$$

Thus, $\sigma_{g_{1k}}^2 | \text{else} \sim \chi^{-2}(df = v_1 + 1, scale = g_{1k}^2 + v_1 s_1^2)$

To sample SNP specific variances on trait 2 conditional on trait 1 in CDBayesA, given

that the prior density is scaled inverted chi-square: $\sigma_{g_{2||k}}^2 | v_{2||}, s_{2||}^2 \sim \chi^{-2}(v_{2||}, v_{2||} s_{2||}^2)$, FCD

for $\sigma_{g_{2||k}}^2$ ($k=1,2,\dots,m$) can be derived as follows:

$$\begin{aligned}
p\left(\sigma_{g_{2|l}k}^2 \mid \text{else}\right) &\propto p\left(g_{2|l}k \mid \sigma_{g_{2|l}k}^2\right) p\left(\sigma_{g_{2|l}k}^2 \mid v_{2|l}, s_{2|l}^2\right) \\
&\propto \left(2\pi\sigma_{g_{2|l}k}^2\right)^{-1/2} \exp\left(-\frac{g_{2|l}k^2}{2\sigma_{g_{2|l}k}^2}\right) \sigma_{g_{2|l}k}^2^{-\left(\frac{v_{2|l}}{2}+1\right)} e^{-\frac{v_{2|l}s_{2|l}^2}{2\sigma_{g_{2|l}k}^2}} \\
&\propto \sigma_{g_{2|l}k}^2^{-\left(\frac{(v_{2|l}+1)}{2}+1\right)} e^{-\frac{v_{2|l}s_{2|l}^2 + g_{2|l}k^2}{2\sigma_{g_{2|l}k}^2}}
\end{aligned}$$

Thus, $\sigma_{g_{2|l}k}^2 \mid \text{else} \sim \chi^{-2}\left(df = v_{2|l} + 1, scale = g_{2|l}k^2 + v_{2|l}s_{2|l}^2\right)$

To sample random SNP effects and variances for trait 1 in CDBayesB, according to the collapsed sampling strategy (Liu 1994), we jointly sampled g_{1k} and $\sigma_{g_{1k}}^2$ as adapted in Bayes B (Meuwissen et al. 2001). Let's first sample from

$$\begin{aligned}
p\left(\sigma_{g_{1k}}^2 \mid ELSE \text{ except } g_{1k}\right) &= \int_{g_{1k}} p\left(\sigma_{g_{1k}}^2, g_{1k} \mid ELSE\right) dg_{1k} \\
&\propto \int_{g_{1k}} \left(\prod_{i=1}^n p\left(y_{i1}, y_{i2} \mid \beta, \mathbf{g}_{-1k}, g_{1k}, \mathbf{g}_{2|l}, \mathbf{R}\right)\right) p\left(g_{1k} \mid \sigma_{g_{1k}}^2\right) p\left(\sigma_{g_{1k}}^2 \mid v_1, s_1^2, \pi_1\right) dg_{1k} \\
&\propto p\left(\sigma_{g_{1k}}^2 \mid v_1, s_1^2, \pi_1\right) |\mathbf{V}_{1k}|^{-1/2} \exp\left(-\frac{1}{2} \mathbf{y}_{-1k}^* ' \mathbf{V}_{1k}^{-1} \mathbf{y}_{-1k}^*\right)
\end{aligned}$$

where $\mathbf{y}_{-1k}^* = \left[\mathbf{y}_{1,-1k}^* \quad \mathbf{y}_{2,-1k}^*\right]'$, $\mathbf{y}_{1,-1k}^* = \mathbf{y}_1 - \mathbf{X}_1\beta_1 - \mathbf{Z}_{-k}\mathbf{g}_{-1k}$,

$$\mathbf{y}_{2,-1k}^* = \mathbf{y}_2 - \mathbf{X}_2\beta_2 - (\mathbf{Z}\Psi)_{-k}\mathbf{g}_{-1k} - \mathbf{Z}\mathbf{g}_{2|l} \quad \text{and} \quad \mathbf{V}_{1k} = \begin{bmatrix} \mathbf{z}_{.k} \\ \phi_k \mathbf{z}_{.k} \end{bmatrix} \begin{bmatrix} \mathbf{z}_{.k} & \phi_k \mathbf{z}_{.k} \end{bmatrix} \sigma_{g_{1k}}^2 + \mathbf{R} \otimes \mathbf{I}$$

Using that expression, the random walk Metropolis-Hastings acceptance ratio for sampling from $p\left(\sigma_{g_{1k}}^2 \mid ELSE \text{ except } g_{1k}\right)$ at MCMC cycle [t] based on using

$p\left(\sigma_{g_{1k}}^2 \mid v_1, s_1^2, \pi_1\right)$ as the candidate density looks as follows:

$$\alpha\left(\sigma_{g_{1k}}^{2[t-1]}, \sigma_{g_{1k}}^{2*}\right) = \begin{cases} \min\left(\frac{p\left(\sigma_{g_{1k}}^{2*} \mid ELSE \text{ except } g_{1k}\right) p\left(\sigma_{g_{1k}}^{2[t-1]} \mid \nu_1, s_1^2, \pi_1\right)}{p\left(\sigma_{g_{1k}}^{2[t-1]} \mid ELSE \text{ except } g_{1k}\right) p\left(\sigma_{g_{1k}}^{2*} \mid \nu_1, s_1^2, \pi_1\right)}, 1\right) \\ 1, otherwise; \end{cases}$$

According to Meuwissen et al. (2001), this ratio is further equal to:

$$\alpha\left(\sigma_{g_{1k}}^{2[t-1]}, \sigma_{g_{1k}}^{2*}\right) = \begin{cases} \min\left(\frac{\left|\mathbf{V}_{1k}^*\right|^{-1/2} \exp\left(-\frac{1}{2} \mathbf{y}_{-1k}^* \mathbf{V}_{1k}^{*-1} \mathbf{y}_{-1k}^*\right)}{\left|\mathbf{V}_{1k}^{[t-1]}\right|^{-1/2} \exp\left(-\frac{1}{2} \mathbf{y}_{-1k}^* \left(\mathbf{V}_{1k}^{[t-1]}\right)^{-1} \mathbf{y}_{-1k}^*\right)}, 1\right) \\ 1, otherwise \end{cases}$$

Using results from Rohan Fernando, we could simplify the Metropolis acceptance ratio for sampling SNP specific variance for trait 1 further as follows:

$$\alpha\left(\sigma_{g_{1k}}^{2[t-1]}, \sigma_{g_{1k}}^{2*}\right) = \begin{cases} \min\left(\frac{\left(\nu_{1k}^*\right)^{-1/2} \exp\left(-\frac{w_{1k}^2}{2\nu_{1k}^*}\right)}{\left(\nu_{1k}^{[t-1]}\right)^{-1/2} \exp\left(-\frac{w_{1k}^2}{2\nu_{1k}^{[t-1]}}\right)}, 1\right) \\ 1, otherwise \end{cases}$$

Where

$$\begin{aligned} w_{1k} &= \begin{bmatrix} \mathbf{z}_{.k}' & \phi_k \mathbf{z}_{.k}' \end{bmatrix} \begin{bmatrix} \mathbf{I}r^{11} & \mathbf{I}r^{12} \\ \mathbf{I}r^{12} & \mathbf{I}r^{22} \end{bmatrix} \begin{bmatrix} \mathbf{z}_{.k} g_{1k} + \mathbf{e}_1 \\ \phi_k \mathbf{z}_{.k} g_{1k} + \mathbf{e}_2 \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{z}_{.k}' & \phi_k \mathbf{z}_{.k}' \end{bmatrix} \begin{bmatrix} r^{11} \mathbf{z}_{.k} g_{1k} + r^{11} \mathbf{e}_1 + r^{12} \phi_k \mathbf{z}_{.k} g_{1k} + r^{12} \mathbf{e}_2 \\ r^{12} \mathbf{z}_{.k} g_{1k} + r^{12} \mathbf{e}_1 + r^{22} \phi_k \mathbf{z}_{.k} g_{1k} + r^{22} \mathbf{e}_2 \end{bmatrix} \\ &= \left(r^{11} + 2r^{12} \phi_k + r^{22} \phi_k^2\right) \mathbf{z}_{.k}' \mathbf{z}_{.k} g_{1k} + \left(r^{11} + \phi_k r^{12}\right) \mathbf{z}_{.k}' \mathbf{e}_1 + \left(r^{12} + \phi_k r^{22}\right) \mathbf{z}_{.k}' \mathbf{e}_2 \end{aligned}$$

and

$$\begin{aligned}
v_{1k} &= \text{var}(w_{1k}) = (r^{11} + 2r^{12}\phi_k + r^{22}\phi_k^2)^2 (\mathbf{z}'_{.k} \mathbf{z}_{.k})^2 \sigma_{g_{1k}}^2 + (r^{11} + \phi_k r^{12})^2 (\mathbf{z}'_{.k} \mathbf{z}_{.k}) \sigma_{e_1}^2 \\
&+ (r^{12} + \phi_k r^{22})^2 (\mathbf{z}'_{.k} \mathbf{z}_{.k}) \sigma_{e_2}^2 + (r^{11} + \phi_k r^{12})(r^{12} + \phi_k r^{22})(\mathbf{z}'_{.k} \mathbf{z}_{.k}) \sigma_{e_{12}} \\
&= (r^{11} + 2r^{12}\phi_k + r^{22}\phi_k^2)^2 \left(\sum_{i=1}^n z_{ik}^2 \right) \sigma_{g_{1k}}^2 + (r^{11} + \phi_k r^{12})^2 \left(\sum_{i=1}^n z_{ik}^2 \right) \sigma_{e_1}^2 \\
&+ (r^{12} + \phi_k r^{22})^2 \left(\sum_{i=1}^n z_{ik}^2 \right) \sigma_{e_2}^2 + (r^{11} + \phi_k r^{12})(r^{12} + \phi_k r^{22}) \left(\sum_{i=1}^n z_{ik}^2 \right) \sigma_{e_{12}}
\end{aligned}$$

If a non-zero value for $\sigma_{g_{1k}}^2$ is sampled, then one can draw samples of g_{1k} using the same full conditionals as with CD-BayesA. If either $g_{1k}^{[t+1]}$ or $g_{1k}^{[t]}$ are nonzero, residual for trait 1 needs to be updated as $\mathbf{e}_1 = \mathbf{e}_1 - \mathbf{z}_{.k} (g_{1k}^{[t+1]} - g_{1k}^{[t]})$ and residual for trait 2 also needs to be updated as $\mathbf{e}_2 = \mathbf{e}_2 - \phi_k \mathbf{z}_{.k} (g_{1k}^{[t+1]} - g_{1k}^{[t]})$ immediately thereafter.

To sample random SNP effects and variances for trait 2 conditional on trait 1 in CDBayesB, according to the collapsed sampling strategy (Liu 1994), we jointly sampled $g_{2||k}$ and $\sigma_{g_{2||k}}^2$ as adapted in Bayes B (Meuwissen et al. 2001). Let's first sample from

$$\begin{aligned}
p(\sigma_{g_{2||k}}^2 | ELSE \text{ except } g_{2k}) &= \int_{g_{2||k}} p(\sigma_{g_{2||k}}^2, g_{2||k} | ELSE) dg_{2||k} \\
&\propto \int_{g_{2||k}} \left(\prod_{i=1}^n p(y_{i1}, y_{i2} | \boldsymbol{\beta}, \mathbf{g}_{1k}, g_{2||k}, \mathbf{g}_{-2||k}, \mathbf{R}) \right) p(g_{2||k} | \sigma_{g_{2||k}}^2) p(\sigma_{g_{2||k}}^2 | v_2, s_2^2, \pi_2) dg_{2||k} \\
&\propto p(\sigma_{g_{2||k}}^2 | v_2, s_2^2, \pi_2) |\mathbf{V}_{2||k}|^{-1/2} \exp\left(-\frac{1}{2} \mathbf{y}_{-2||k}^* \mathbf{V}_{2||k}^{-1} \mathbf{y}_{-2||k}^*\right)
\end{aligned}$$

where $\mathbf{y}_{-2||k}^* = [\mathbf{y}_1^* \mathbf{y}_{2,-2||k}^*]'$, $\mathbf{y}_1^* = \mathbf{y}_1 - \mathbf{X}_1 \boldsymbol{\beta}_1 - \mathbf{Z}_1 \mathbf{g}_1$, $\mathbf{y}_{2,-2||k}^* = \mathbf{y}_2 - \mathbf{X}_2 \boldsymbol{\beta}_2 - \mathbf{Z} \Psi \mathbf{g}_1 - \mathbf{Z} \mathbf{g}_{-2||k}$ and

$$\mathbf{V}_{2||k} = \begin{bmatrix} 0 \\ \mathbf{z}_{.k} \end{bmatrix} [0 \quad \mathbf{z}_{.k}] \sigma_{g_{2||k}}^2 + \mathbf{R} \otimes \mathbf{I}$$

Using that expression, the random walk Metropolis-Hastings acceptance ratio for

sampling from $p(\sigma_{g_{2|k}}^2 | ELSE \text{ except } g_{2|k})$ at MCMC cycle [t] based on using

$p(\sigma_{g_{2|k}}^2 | \nu_2, s_2^2, \pi_2)$ as the candidate density looks as follows:

$$\alpha(\sigma_{g_{2|k}}^{2[t-1]}, \sigma_{g_{2|k}}^{2*}) = \begin{cases} \min \left(\frac{p(\sigma_{g_{2|k}}^{2*} | ELSE \text{ except } g_{2|k}) p(\sigma_{g_{2|k}}^{2[t-1]} | \nu_2, s_2^2, \pi_2)}{p(\sigma_{g_{2|k}}^{2[t-1]} | ELSE \text{ except } g_{2|k}) p(\sigma_{g_{2|k}}^{2*} | \nu_2, s_2^2, \pi_2)}, 1 \right) \\ 1, otherwise; \end{cases}$$

According to Meuwissen et al. (2001), this ratio is further equal to:

$$\alpha(\sigma_{g_{2|k}}^{2[t-1]}, \sigma_{g_{2|k}}^{2*}) = \begin{cases} \min \left(\frac{|\mathbf{V}_{2|k}^*|^{-1/2} \exp\left(-\frac{1}{2} \mathbf{y}_{-2|k}^* ' \mathbf{V}_{2|k}^{*-1} \mathbf{y}_{-2|k}^*\right)}{|\mathbf{V}_{2|k}^{[t-1]}|^{-1/2} \exp\left(-\frac{1}{2} \mathbf{y}_{-2|k}^* ' (\mathbf{V}_{2|k}^{[t-1]})^{-1} \mathbf{y}_{-2|k}^*\right)}, 1 \right) \\ 1, otherwise \end{cases}$$

Using results from Rohan Fernando, we could simplify the Metropolis acceptance ratio

for sampling SNP specific variance for trait 1 further as follows:

$$\alpha(\sigma_{g_{2|k}}^{2[t-1]}, \sigma_{g_{2|k}}^{2*}) = \begin{cases} \min \left(\frac{(v_{2|k}^*)^{-1/2} \exp\left(-\frac{w_{2|k}^2}{2v_{2|k}^*}\right)}{(v_{2|k}^{[t-1]})^{-1/2} \exp\left(-\frac{w_{2|k}^2}{2v_{2|k}^{[t-1]}}\right)}, 1 \right) \\ 1, otherwise \end{cases}$$

Where

$$\begin{aligned}
w_{2||k} &= \begin{bmatrix} 0 & \mathbf{z}_{\cdot k}' \end{bmatrix} \begin{bmatrix} \mathbf{I}r^{11} & \mathbf{I}r^{12} \\ \mathbf{I}r^{12} & \mathbf{I}r^{22} \end{bmatrix} \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{z}_{\cdot k} g_{2||k} + \mathbf{e}_2 \end{bmatrix} \\
&= \begin{bmatrix} 0 & \mathbf{z}_{\cdot k}' \end{bmatrix} \begin{bmatrix} r^{11}\mathbf{e}_1 + r^{12}\mathbf{z}_{\cdot k} g_{2||k} + r^{12}\mathbf{e}_2 \\ r^{12}\mathbf{e}_1 + r^{22}\mathbf{z}_{\cdot k} g_{2||k} + r^{22}\mathbf{e}_2 \end{bmatrix} \\
&= r^{12}\mathbf{z}_{\cdot k}'\mathbf{e}_1 + r^{22}\mathbf{z}_{\cdot k}'\mathbf{e}_2 + r^{22}\mathbf{z}_{\cdot k}'\mathbf{z}_{\cdot k} g_{2||k}
\end{aligned}$$

and

$$\begin{aligned}
v_{2||k} &= \text{var}(w_{2||k}) \\
&= (r^{12})^2 (\mathbf{z}_{\cdot k}'\mathbf{z}_{\cdot k}) \sigma_{e_1}^2 + (r^{22})^2 (\mathbf{z}_{\cdot k}'\mathbf{z}_{\cdot k}) \sigma_{e_2}^2 + r^{12}r^{22} (\mathbf{z}_{\cdot k}'\mathbf{z}_{\cdot k}) \sigma_{e_{12}} + (r^{22})^2 (\mathbf{z}_{\cdot k}'\mathbf{z}_{\cdot k})^2 \sigma_{g_{1k}}^2 \\
&= (r^{12})^2 \left(\sum_{i=1}^n z_{ik}^2 \right) \sigma_{e_1}^2 + (r^{22})^2 \left(\sum_{i=1}^n z_{ik}^2 \right) \sigma_{e_2}^2 + r^{12}r^{22} \left(\sum_{i=1}^n z_{ik}^2 \right) \sigma_{e_{12}} + (r^{22})^2 \left(\sum_{i=1}^n z_{ik}^2 \right)^2 \sigma_{g_{1k}}^2
\end{aligned}$$

If a non-zero value for $\sigma_{g_{2||k}}^2$ is sampled, then one can draw samples of $g_{2||k}$ using the

same full conditionals as with CDBayesA. If either $g_{2||k}^{[t+1]}$ or $g_{2||k}^{[t]}$ are nonzero, residual

for trait 2 also needs to be updated as $\mathbf{e}_2 = \mathbf{e}_2 - \mathbf{z}_{\cdot k} (g_{2||k}^{[t+1]} - g_{2||k}^{[t]})$ immediately

thereafter.

To sample random association parameters ϕ_k ($k=1,2,\dots,m$) in CDBayesB, given

the specified prior distribution, i.e. $\phi_k \sim \begin{cases} N(\mu_\phi, \sigma_\phi^2) & \text{with prob } \pi_\phi \\ 0 & \text{with prob } 1 - \pi_\phi \end{cases}$, FCD on ϕ_k can be

derived following Geweke (1994):

If we define $\mathbf{y}_{2,-\phi_k}^* = \mathbf{y}_2 - \mathbf{X}_2 \boldsymbol{\beta}_2 - (\mathbf{Z}\boldsymbol{\Gamma})_{-k} \boldsymbol{\phi}_k - \mathbf{Z}\mathbf{g}_{2||} = \mathbf{z}_{\cdot k}^g \phi_k + \mathbf{e}_2$ and $\mathbf{z}_{\cdot k}^g = g_{1k} \mathbf{z}_{\cdot k}$, then the

likelihood function kernel is : $\exp \left(-\frac{(\mathbf{y}_{2,-\phi_k}^* - \mathbf{z}_{\cdot k}^g \phi_k)' (\mathbf{y}_{2,-\phi_k}^* - \mathbf{z}_{\cdot k}^g \phi_k)}{2\sigma_{e_2}^2} \right)$. Conditional on

$\phi_k = 0$, the value of the kernel is: $\exp\left(-\frac{\mathbf{y}_{2,-\phi_k}^* \mathbf{y}_{2,-\phi_k}^*}{2\sigma_{e_2}^2}\right)$. Conditional on $\phi_k \neq 0$, the

corresponding kernel density is:

$$\begin{aligned} & \sigma_\phi^{-1} \exp\left(-\frac{(\mathbf{y}_{2,-\phi_k}^* - \mathbf{z}_{\cdot,k}^g \phi_k)' (\mathbf{y}_{2,-\phi_k}^* - \mathbf{z}_{\cdot,k}^g \phi_k)}{2\sigma_{e_2}^2}\right) \exp\left(-\frac{(\phi_k - \mu_\phi)^2}{2\sigma_\phi^2}\right) \\ &= \sigma_\phi^{-1} \exp\left(-\frac{(\mathbf{y}_{2,-\phi_k}^* - \mathbf{z}_{\cdot,k}^g \tilde{\phi}_k)' (\mathbf{y}_{2,-\phi_k}^* - \mathbf{z}_{\cdot,k}^g \tilde{\phi}_k)}{2\sigma_{e_2}^2}\right) \exp\left(-\frac{(\phi_k - \tilde{\phi}_k)^2}{2\omega_k^2} - \frac{(\phi_k - \mu_\phi)^2}{2\sigma_\phi^2}\right) \\ &= \sigma_\phi^{-1} \exp\left(-\frac{(\mathbf{y}_{2,-\phi_k}^* - \mathbf{z}_{\cdot,k}^g \tilde{\phi}_k)' (\mathbf{y}_{2,-\phi_k}^* - \mathbf{z}_{\cdot,k}^g \tilde{\phi}_k)}{2\sigma_{e_2}^2}\right) \exp\left(-\frac{(\phi_k - \hat{\phi}_k)^2}{2v_{\phi k}^2}\right) \exp\left(-\left(\frac{\tilde{\phi}_k^2}{2\omega_k^2} + \frac{\mu_\phi^2}{2\sigma_\phi^2} - \frac{\hat{\phi}_k^2}{2v_{\phi k}^2}\right)\right) \end{aligned}$$

where

$$\hat{\phi}_k = \frac{\sigma_{e_2}^2 \sigma_\phi^{-2} \mu_\phi + g_{1k} \sum_{i=1}^n z_{ik} e_{2j} + \left(g_{1k}^2 \sum_{i=1}^n (z_{ik})^2\right) \tilde{\phi}_k}{\left(g_{1k}^2 \sum_{i=1}^n (z_{ik})^2\right) + \sigma_{e_2}^2 \sigma_\phi^{-2}} \quad \text{and} \quad v_{\phi k}^2 = \left(\frac{g_{1k}^2 \sum_{j=1}^n (z_{jk})^2}{\sigma_{e_2}^2} + \sigma_\phi^{-2}\right)^{-1}$$

To remove the conditioning on $\phi_k = 0$ or on $\phi_k \neq 0$, it is necessary to further integrate

this expression over ϕ_k . This integration yields:

$$= \frac{v_{\phi k}}{\sigma_\phi} \exp\left(-\frac{(\mathbf{y}_{2,-\phi_k}^* - \mathbf{z}_{\cdot,k}^g \tilde{\phi}_k)' (\mathbf{y}_{2,-\phi_k}^* - \mathbf{z}_{\cdot,k}^g \tilde{\phi}_k)}{2\sigma_{e_2}^2}\right) \exp\left(-\left(\frac{\tilde{\phi}_k^2}{2\omega_k^2} + \frac{\mu_\phi^2}{2\sigma_\phi^2} - \frac{\hat{\phi}_k^2}{2v_{\phi k}^2}\right)\right)$$

Thus, the conditional Bayes Factor in favor of $\phi_k \neq 0$ over $\phi_k = 0$ is,

$$BF_k = \frac{v_{\phi k}}{\sigma_\phi} \exp \left(\frac{\mathbf{y}_{2,-\phi_k}^* \mathbf{y}_{2,-\phi_k}^* - (\mathbf{y}_{2,-\phi_k}^* - \mathbf{z}_{.k}^g \tilde{\phi}_k)' (\mathbf{y}_{2,-\phi_k}^* - \mathbf{z}_{.k}^g \tilde{\phi}_k)}{2\sigma_{e_2}^2} \right) \exp \left(- \left(\frac{\tilde{\phi}_k^2}{2\omega_k^2} + \frac{\mu_\phi^2}{2\sigma_\phi^2} - \frac{\hat{\phi}_k^2}{2v_{\phi k}^2} \right) \right)$$

$$= \frac{v_{\phi k}}{\sigma_\phi} \exp \left(\frac{\hat{\phi}_k^2}{2v_{\phi k}^2} - \frac{\mu_\phi^2}{2\sigma_\phi^2} \right)$$

To draw ϕ_k from its conditional distribution, the conditional posterior probability that

$$\phi_k = 0 \text{ is computed from the conditional Bayes factor (BF) is } \hat{p}_k = \frac{1 - \pi_\phi}{(1 - \pi_\phi) + \pi_\phi BF_k}.$$

Based on a comparison of this probability with a drawing from a Uniform(0,1), the

choice $\phi_k = 0$ or $\phi_k \neq 0$ is made. If $\phi_k \neq 0$, then a draw is made from a $N(\hat{\phi}_k, v_{\phi k}^2)$

distribution. Immediately after sampling ϕ_k , if either $\phi_k^{[t+1]}$ or $\phi_k^{[t]}$ are nonzero, we

update residual for the second trait using $\mathbf{e}_2 = \mathbf{e}_2 - g_{1k} \mathbf{z}_{.k} (\phi_k^{[t+1]} - \phi_k^{[t]})$.

To sample proportion of SNP markers associated with non-zero SNP effects for trait 1 in CDBayesB, given a specified prior $p(\pi_1 | \alpha_{\pi_1}, \beta_{\pi_1}) = \text{Beta}(\alpha_{\pi_1}, \beta_{\pi_1})$, the FCD of π_1 is based on the following,

$$p(\pi_1 | ELSE) \propto \left(\prod_{j=1}^m p(\sigma_{g_{1j}}^2 | v_1, s_1^2, \pi_1) \right) p(\pi_1 | \alpha_{\pi_1}, \beta_{\pi_1})$$

Let $m_1 = \sum_{j=1}^m I(\sigma_{g_{1j}}^2 > 0)$ denote the number of non-zero values sampled in $\sigma_{g_{1j}}^2$ at a

particular MCMC cycle where $I(\cdot)$ denotes the indicator function. Then we can write

$$p(\pi_1 | ELSE) = \text{beta}(\alpha_{\pi_1} + m_1, \beta_{\pi_1} + m - m_1).$$

To sample proportion of SNP markers associated with non-zero SNP effects for trait 2 conditional on trait 1 in CDBayesB, given a specified prior

$p(\pi_{2||} | \alpha_{\pi_{2||}}, \beta_{\pi_{2||}}) = \text{Beta}(\alpha_{\pi_{2||}}, \beta_{\pi_{2||}})$, the FCD of $\pi_{2||}$ is based on the following,

$$p(\pi_{2||} | ELSE) \propto \left(\prod_{j=1}^m p(\sigma_{g_{2||j}}^2 | \nu_{2||}, s_{2||}^2, \pi_{2||}) \right) p(\pi_{2||} | \alpha_{\pi_{2||}}, \beta_{\pi_{2||}})$$

Let $m_2 = \sum_{j=1}^m I(\sigma_{g_{2||j}}^2 > 0)$ denote the number of non-zero values sampled in $\sigma_{g_{2||j}}^2$ at a

particular MCMC cycle where $I(\cdot)$ denotes the indicator function. Then we can write

$$p(\pi_{2||} | ELSE) = \text{beta}(\alpha_{\pi_{2||}} + m_2, \beta_{\pi_{2||}} + m - m_2).$$

To sample proportion of non-zero association parameters between two traits in CDBayesB, given a specified prior $p(\pi_{\phi} | \alpha_{\phi}, \beta_{\phi}) = \text{Beta}(\alpha_{\phi}, \beta_{\phi})$, the FCD of π_{ϕ} is based on the following,

$$p(\pi_{\phi} | ELSE) \propto \left(\prod_{k=1}^m p(\phi_k | \mu_{\phi}, \sigma_{\phi}^2, \pi_{\phi}) \right) p(\pi_{\phi} | \alpha_{\phi}, \beta_{\phi})$$

Let $m_{\phi} = \sum_{k=1}^m I(\phi_k \neq 0)$ denote the number of non-zero values sampled in ϕ_k at a particular

MCMC cycle where $I(\cdot)$ denotes the indicator function. Then we can write

$$p(\pi_{\phi} | ELSE) = \text{beta}(\alpha_{\phi} + m_{\phi}, \beta_{\phi} + m - m_{\phi}).$$

To sample scale parameter s_1^2 for random SNP effects for trait 1 in CDBayesA\CDBayesB, given a specified prior for scale parameter

$p(s_1^2 | \alpha_1, \beta_1) = \text{Gamma}(\alpha_1, \beta_1)$, we can write FCD as follows:

$$\begin{aligned}
 p(s_1^2 | ELSE) &\propto \left(\prod_{j=1}^m I(\sigma_{g_{1j}}^2 > 0) p(\sigma_{g_{1j}}^2 | \nu_1, s_1^2, \pi_1) \right) p(s_1^2 | \alpha_1, \beta_1) \\
 &= \left(\prod_{j=1}^m I(\sigma_{g_{1j}}^2 > 0) \frac{\left(\frac{s_1^2 \nu_1}{2} \right)^{\frac{\nu_1}{2}}}{\Gamma\left(\frac{\nu_1}{2}\right)} \sigma_{g_{1j}}^2^{-\left(\frac{\nu_1}{2}+1\right)} e^{-\frac{s_1^2 \nu_1}{2\sigma_{g_{1j}}^2}} \right) (s_1^2)^{\alpha_1-1} e^{-\beta_1 s_1^2} \\
 &\propto (s_1^2)^{\alpha_1 + \frac{m_1 \nu_1}{2} - 1} \exp\left(-s_1^2 \left(\frac{\nu_1}{2} \sum_{j=1}^m I(\sigma_{g_{1j}}^2 > 0) \sigma_{g_{1j}}^{-2} + \beta_1 \right)\right)
 \end{aligned}$$

That is, a gamma distribution with parameters $\alpha_1 + \frac{m_1 \nu_1}{2}$ and $\frac{\nu_1}{2} \sum_{j=1}^m I(\sigma_{g_{1j}}^2 > 0) \sigma_{g_{1j}}^{-2} + \beta_1$.

To sample scale parameter $s_{2||}^2$ for random SNP effects for trait 2 conditional on trait 1 in CDBayesA\CDBayesB, given a specified prior for scale parameter

$p(s_{2||}^2 | \alpha_2, \beta_2) = \text{Gamma}(\alpha_2, \beta_2)$, we can write FCD as follows:

$$\begin{aligned}
p(s_{2||}^2 | ELSE) &\propto \left(\prod_{j=1}^m I(\sigma_{g_{2||j}}^2 > 0) p(\sigma_{g_{2||j}}^2 | \nu_{2||}, s_{2||}^2, \pi_{2||}) \right) p(s_{2||}^2 | \alpha_2, \beta_2) \\
&= \left(\prod_{j=1}^m I(\sigma_{g_{2||j}}^2 > 0) \frac{\left(\frac{s_{2||}^2 \nu_{2||}}{2} \right)^{\frac{\nu_{2||}}{2}}}{\Gamma\left(\frac{\nu_{2||}}{2}\right)} \sigma_{g_{2||j}}^2^{-\left(\frac{\nu_{2||}}{2}+1\right)} e^{-\frac{s_{2||}^2 \nu_{2||}}{2\sigma_{g_{2||j}}^2}} \right) (s_{2||}^2)^{\alpha_2-1} e^{-\beta_2 s_{2||}^2} \\
&\propto (s_{2||}^2)^{\alpha_2 + \frac{m_2 \nu_{2||}}{2} - 1} \exp\left(-s_{2||}^2 \left(\frac{\nu_{2||}}{2} \sum_{j=1}^m I(\sigma_{g_{2||j}}^2 > 0) \sigma_{g_{2||j}}^{-2} + \beta_2 \right)\right)
\end{aligned}$$

That is, a gamma distribution with parameters $\alpha_2 + \frac{m_2 \nu_{2||}}{2}$ and

$$\frac{\nu_{2||}}{2} \sum_{j=1}^m I(\sigma_{g_{2||j}}^2 > 0) \sigma_{g_{2||j}}^{-2} + \beta_2.$$

To sample degrees of freedom for SNP effects on trait 1 in CDBayesA/CDBayesB, with a specified non-informative prior for degrees of freedom

$p(\nu_1) = 1 / (1 + \nu_1)^2$, we can write FCD for ν_1 as follows,

$$\begin{aligned}
p(\nu_1 | ELSE) &\propto \left(\prod_{j=1}^m I(\sigma_{g_{1j}}^2 > 0) p(\sigma_{g_{1j}}^2 | \nu_1, s_1^2, \pi_1) \right) p(\nu_1) \\
&= \left(\prod_{j=1}^m I(\sigma_{g_{1j}}^2 > 0) \frac{\left(\frac{s_1^2}{2} \right)^{\frac{\nu_1}{2}}}{\Gamma\left(\frac{\nu_1}{2}\right)} \sigma_{g_{1j}}^2^{-\left(\frac{\nu_1}{2}+1\right)} e^{-\frac{s_1^2}{2\sigma_{g_{1j}}^2}} \right) p(\nu_1)
\end{aligned}$$

We sampled ν_1 using a random walk Metropolis Hastings algorithm, which was described in other non-genomic applications involving the sampling of degrees of freedom parameter (Kizilkaya & Tempelman 2005; Bello 2010).

To sample degrees of freedom for SNP effects on trait 2 conditional on trait 1 in CDBayesA/CDBayesB, with a specified non-informative prior for degrees of freedom

$p(v_{2||}) = 1 / (1 + v_{2||})^2$, we can write FCD for $v_{2||}$ as the following,

$$p(v_{2||} | ELSE) \propto \left(\prod_{j=1}^m I(\sigma_{g_{2||j}}^2 > 0) p(\sigma_{g_{2||j}}^2 | v_{2||}, s_{2||}^2, \pi_{2||}) \right) p(v_{2||})$$

$$= \left(\prod_{j=1}^m I(\sigma_{g_{2||j}}^2 > 0) \frac{\left(\frac{s_{2||}^2}{2}\right)^{\frac{v_{2||}}{2}}}{\Gamma\left(\frac{v_{2||}}{2}\right)} \sigma_{g_{2||j}}^2^{-\left(\frac{v_{2||}}{2}+1\right)} e^{-\frac{s_{2||}^2}{2\sigma_{g_{2||j}}^2}} \right) p(v_{2||})$$

We sampled $v_{2||}$ using a random walk Metropolis Hastings algorithm, which was described in other non-genomic applications involving the sampling of degrees of freedom parameter (Kizilkaya & Tempelman 2005; Bello 2010).

To sample mean for association parameters between two traits in CDBayesA/CD-BayesB, given a specified prior $\mu_\phi \sim N(\tau, \zeta^2)$, FCD can be written as the following,

$$p(\mu_\phi | ELSE) \propto \left(\prod_{j=1}^m I(\phi_k \neq 0) p(\phi_k | \mu_\phi, \sigma_\phi^2) \right) p(\mu_\phi | \tau, \zeta^2)$$

$$\propto \exp\left(-\frac{1}{2\sigma_\phi^2} \sum_{k=1}^m I(\phi_k \neq 0) (\phi_k - \mu_\phi)^2\right) \exp\left(-\frac{(\mu_\phi - \tau)^2}{2\zeta^2}\right)$$

$$\propto \exp\left(-\frac{1}{2\tilde{\sigma}_{\mu_\phi}^2} (\mu_\phi - \tilde{\mu}_\phi)^2\right)$$

Where

$$\tilde{\mu}_\phi = \frac{\bar{\phi}\zeta^2 + \tau \frac{\sigma_\phi^2}{m_\phi}}{\zeta^2 + \frac{\sigma_\phi^2}{m_\phi}} \quad \text{for} \quad \bar{\phi} = \frac{\sum_{k=1}^m \phi_k I(\phi_k \neq 0)}{m_\phi} \quad \text{and} \quad \tilde{\sigma}_{\mu_\phi}^2 = \frac{\zeta^2 \frac{\sigma_\phi^2}{m_\phi}}{\zeta^2 + \frac{\sigma_\phi^2}{m_\phi}}$$

Hence, $\mu_\phi | ELSE \sim N(\tilde{\mu}_\phi, \tilde{\sigma}_{\mu_\phi}^2)$

To sample variance for association parameters between two traits in

CDBayesA/CDBayesB, given a specified prior $p(\sigma_\phi^2) = (\sigma_\phi^2)^{-1/2}$, FCD for σ_ϕ^2 can be

written as follows,

$$\begin{aligned} p(\sigma_\phi^2 | ELSE) &\propto \left(\prod_{k=1}^m I(\phi_k \neq 0) p(\phi_k | \mu_\phi, \sigma_\phi^2) \right) p(\sigma_\phi^2) \\ &\left((2\pi\sigma_\phi^2)^{-m/2} \exp\left(-\frac{1}{2\sigma_\phi^2} \sum_{k=1}^m I(\phi_k \neq 0) (\phi_k - \mu_\phi)^2\right) \right) (\sigma_\phi^2)^{-1/2} \\ &\propto \sigma_\phi^2^{-\left(\frac{(m_\phi-1)}{2}+1\right)} e^{-\frac{\sum_{k=1}^m I(\phi_k \neq 0) (\phi_k - \mu_\phi)^2}{2\sigma_\phi^2}} \end{aligned}$$

$$\text{Hence, } \sigma_\phi^2 | else \sim \chi^2 \left(df = (m_\phi - 1), scale = \sum_{k=1}^m 1(\phi_k \neq 0) (\phi_k - \mu_\phi)^2 \right)$$

To sample residual variance-covariance matrix in CDBayesA\CDBayesB, given a

specified Inverted Wishart prior $p(\Sigma_e | v_0, \Sigma_0) = IW(v_0, \Sigma_0)$, we can write the FCD for

Σ_e as follows,

$$\begin{aligned} p(\Sigma_e | ELSE, \mathbf{y}) &\propto p(\mathbf{y} | \boldsymbol{\beta}, \mathbf{g}, \Sigma_e) p(\Sigma_e | v_0, \Sigma_0) \\ &\propto |\Sigma_e|^{\frac{n}{2}} \exp\left(-0.5 \left[\text{trace}(\Sigma_e^{-1} \mathbf{S}_e) \right]\right) |\Sigma_e|^{\frac{1}{2}(v_0+3)} \exp\left(-0.5 \text{trace}(\Sigma_e^{-1} \Sigma_0^{-1})\right) \\ &\propto |\Sigma_e|^{\frac{1}{2}(n+v_0+3)} \exp\left(-0.5 \left[\text{trace}(\Sigma_e^{-1} (\mathbf{S}_e + \Sigma_0^{-1})) \right]\right) \end{aligned}$$

$$\text{where } \mathbf{S}_e = \begin{bmatrix} \mathbf{e}_1' \mathbf{e}_1 & \mathbf{e}_1' \mathbf{e}_2 \\ \mathbf{e}_2' \mathbf{e}_1 & \mathbf{e}_2' \mathbf{e}_2 \end{bmatrix}$$

$$\text{Hence, } \Sigma_e | else \sim \text{IW} \left(n + \nu_0, \left(\mathbf{S}_e + \Sigma_0^{-1} \right) \right)$$

D2 Supplementary tables and figures

Table D2.1: Summary of Responsible surface designs (RSD) in LE simulation.

| Factors | RSD |
|--|----------------|
| Controlling | |
| Heritability on Trait 1 | 0.8 |
| Heritability on Trait 2 | 0.1 |
| Residual covariance between two traits | 0 |
| Number of SNPs | 2000 |
| Mean on association parameters | 0.8 |
| Investigating | |
| Number of animals | 2000,4000,6000 |
| Variance on association parameters | 2e-3,1.001,2 |
| Number of QTLs on Trait 1 | 20,310,600 |
| Number of QTLs on Trait 2 | 20,310,600 |
| Number of QTLs on both traits | 20,310,600 |

Table D2.2: P values for the fixed effects by fitting accuracy on the lower heritability trait as response variable in LE simulation under RSD.

| Fixed effects | P-value |
|--|---------|
| Number of animals (n) | <0.0001 |
| Number of QTL on Trait 2 ($M2$) | <0.0001 |
| Number of QTL on both traits ($M12$) | <0.0001 |
| Variance on association parameters (σ_{ϕ}^2) | 0.0005 |
| $M2*M12$ | <0.0001 |
| $M12*M12$ | <0.0001 |
| $M12*Method$ | <0.0001 |
| $M12*M12*Method$ | 0.0038 |

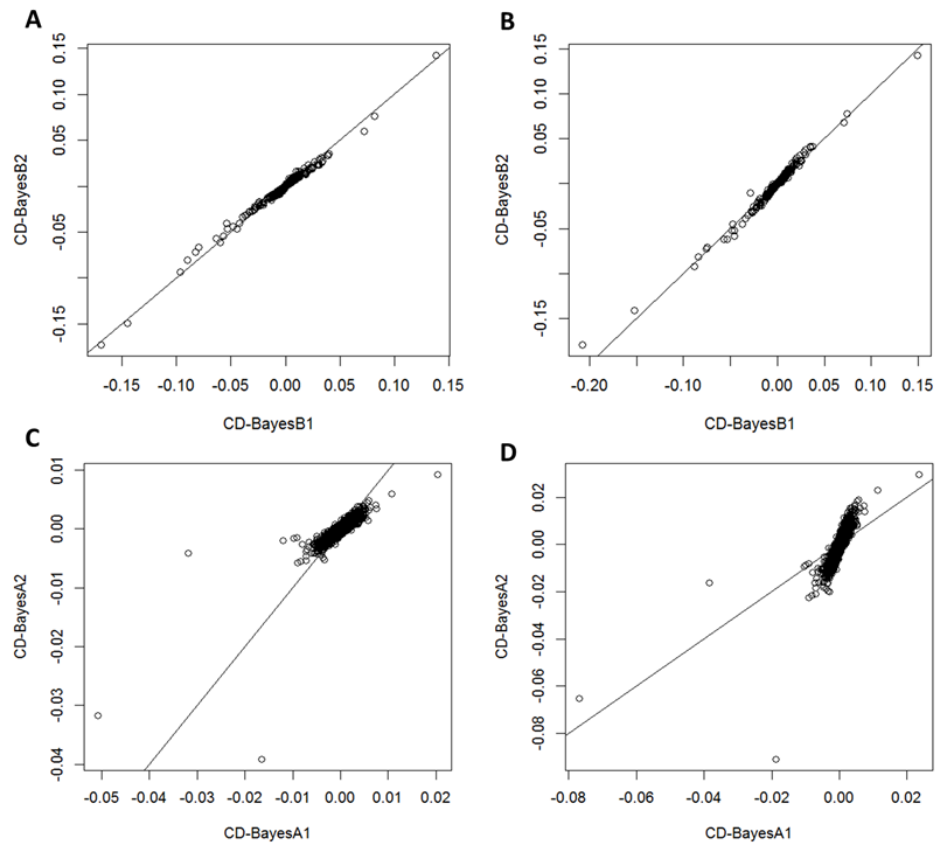


Figure D2.1: Estimated SNP effects for Rust_gall_vol and Rust_bin against SNP index using whole Pine data set comparing CDBayesA1 and CDBayesA2, CDBayesB1 and CDBayesB2. A) and C) were for Rust_gall_vol, B) and D) were for Rust_bin.

BIBLIOGRAPHY

BIBLIOGRAPHY

Abasht, B., E. Sandford, et al. (2009). "Extent and consistency of linkage disequilibrium and identification of DNA markers for production and egg quality traits in commercial layer chicken populations." Bmc Genomics **10**(Suppl. 2): S2.

Badke, Y., R. Bates, et al. (2013). "Methods of tagSNP selection and other variables affecting imputation accuracy in swine." Bmc Genetics **14**(1): 8.

Banerjee, S., B. S. Yandell, et al. (2008). "Bayesian quantitative trait loci mapping for multiple traits." Genetics **179**(4): 2275-2289.

Beerda, B., W. Ouweltjes, et al. (2007). "Effects of genotype by environment interactions on milk yield, energy balance, and protein balance." Journal of Dairy Science **90**(1): 219-228.

Bello, N. M., J. P. Steibel, et al. (2010). "Hierarchical Bayesian modeling of random and residual variance-covariance matrices in bivariate mixed effects models." Biometrical Journal **52**(3): 297-313.

Berry, D. P., F. Buckley, et al. (2003). "Estimation of genotype X environment interactions, in a grassbased system, for milk yield, body condition score, and body weight using random regression models." Livestock Production Science **83**(2-3): 191-203.

Bohmanova, J., I. Misztal, et al. (2008). "Short communication: Genotype by environment interaction due to heat stress." Journal of Dairy Science **91**(2): 840-846.

Burgueno, J., G. de los Campos, et al. (2012). "Genomic Prediction of Breeding Values when Modeling Genotype x Environment Interaction using Pedigree and Dense Molecular Markers." Crop Science **52**(2): 707-719.

Calus, M. P. L., A. F. Groen, et al. (2002). "Genotype x environment interaction for protein yield in Dutch dairy cattle as quantified by different models." Journal of Dairy Science **85**(11): 3115-3123.

Calus, M. P. L., T. H. E. Meuwissen, et al. (2008). "Accuracy of genomic selection using different methods to define haplotypes." Genetics **178**(1): 553-561.

Calus, M. P. L. and R. F. Veerkamp (2003). "Estimation of environmental sensitivity of genetic merit for milk production traits using a random regression model." Journal of Dairy Science **86**(11): 3756-3764.

Calus, M. P. L. and R. F. Veerkamp (2007). "Accuracy of breeding values when using and ignoring the polygenic effect in genomic breeding value estimation with a marker density of one SNP per cM." Journal of Animal Breeding and Genetics **124**(6): 362-368.

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

Cardoso, F. F. and R. J. Tempelman (2012). "Linear reaction norm models for genetic merit prediction of Angus cattle under genotype by environment interaction." Journal of Animal Science **90**(7): 2130-2141.

Carlin, B. P. and T. A. Louis (2008). Bayesian Methods for Data Analysis. Boca Raton, FL, CRC Press.

Chan, J. C. C. and I. Jeliaskov (2009). "MCMC Estimation of Restricted Covariance Matrices." Journal of Computational and Graphical Statistics **18**(2): 457-480.

Chib, S. and E. Greenberg (1995). "Understanding the Metropolis-Hastings algorithm." American Statistician **49**(4): 327-335.

Choi, I., J. P. Steibel, et al. (2010). "Application of alternative models to identify QTL for growth traits in an F-2 Duroc x Pietrain pig resource population." Bmc Genetics **11**.

Coster, A., J. W. M. Bastiaansen, et al. (2010). "QTLMAS 2009: simulated dataset." BMC Proceedings **4**(Suppl 1): S3.

Coster, A., J. W. M. Bastiaansen, et al. (2010). "Sensitivity of methods for estimating breeding values using genetic markers to the number of QTL and distribution of QTL variance." Genetics Selection Evolution **42**: 9.

Daetwyler, H. D. (2009). Genome-wide evaluation of populations. Wageningen, Netherlands, Wageningen Universiteit (Wageningen University).

Daetwyler, H. D., M. P. L. Calus, et al. (2013). "Genomic Prediction in Animals and Plants: Simulation of Data, Validation, Reporting, and Benchmarking." Genetics **193**(2): 347-+.

Daetwyler, H. D., R. Pong-Wong, et al. (2010). "The Impact of Genetic Architecture on Genome-Wide Evaluation Methods." Genetics **185**(3): 1021-1031.

Daniels, M. J. and M. Pourahmadi (2002). "Bayesian analysis of covariance matrices and dynamic models for longitudinal data." Biometrika **89**(3): 553-566.

De Donato, M., S. O. Peters, et al. (2013). "Genotyping-by-Sequencing (GBS): A Novel, Efficient and Cost-Effective Genotyping Method for Cattle Using Next-Generation Sequencing." PLoS ONE **8**(5): e62137.

de Jong, G. (1995). "Phenotypic plasticity as a product of selection in a variable environment." American Naturalist **145**(4): 493-512.

de los Campos, G. and D. Gianola (2007). "Factor analysis models for structuring covariance matrices of additive genetic effects: a Bayesian implementation." Genetics Selection Evolution **39**(5): 481 - 494.

de los Campos, G., D. Gianola, et al. (2010). "Predicting genetic predisposition in humans: the promise of whole-genome markers." Nature Reviews Genetics **11**(12): 880-886.

de los Campos, G., J. M. Hickey, et al. (2012). "Whole Genome Regression and Prediction Methods Applied to Plant and Animal Breeding." Genetics.

de los Campos, G., J. M. Hickey, et al. (2013). "Whole-Genome Regression and Prediction Methods Applied to Plant and Animal Breeding." Genetics **193**(2): 327-+.

de los Campos, G., H. Naya, et al. (2009). "Predicting Quantitative Traits With Regression Models for Dense Molecular Markers and Pedigree." Genetics **182**(1): 375-385.

De Roos, A. P. W., B. J. Hayes, et al. (2008). "Linkage disequilibrium and persistence of phase in Holstein Friesian, Jersey and Angus cattle." Genetics **179**: 1503 - 1512.

Deeb, N. and A. Cahaner (2001). "Genotype-by-environment interaction with broiler genotypes differing in growth rate. 1. The effects of high ambient temperature and naked-neck genotype on lines differing in genetic background." Poultry Science **80**(6): 695-702.

Du, F. X., A. C. Clutter, et al. (2007). "Characterizing linkage disequilibrium in pig populations." International Journal of Biological Sciences **3**(3): 166-178.

Duarte, J. L. G., R. O. Bates, et al. (2013). "Genotype imputation accuracy in a F2 pig population using high density and low density SNP panels." Bmc Genetics **14**.

Edwards, D. B., C. W. Ernst, et al. (2008). "Quantitative trait loci mapping in an F-2 Duroc x Pietrain resource population: I. Growth traits." Journal of Animal Science **86**(2): 241-253.

Falconer, D. S. (1952). "The problem of environment and selection." The American Naturalist **86**: 293-298.

Gelman, A. (2006). "Prior distributions for variance parameters in hierarchical models." Bayesian Analysis **1**(3): 515-533.

Gelman, A., J. B. Carlin, et al. (2003). Bayesian Data Analysis. Boca Raton, FL., CRC Press.

Gianola, D. (2013). "Priors in whole-genome regression: the bayesian alphabet returns." Genetics **194**(3): 573-596.

Gianola, D., G. de los Campos, et al. (2009). "Additive Genetic Variability and the Bayesian Alphabet." Genetics **183**(1): 347-363.

Gianola, D., M. Perez-Enciso, et al. (2003). "On marker-assisted prediction of genetic value: Beyond the ridge." Genetics **163**(1): 347-365.

Gianola, D. and D. Sorensen (2004). "Quantitative Genetic Models for Describing Simultaneous and Recursive Relationships Between Phenotypes." Genetics **167**(3): 1407-1424.

Gianola, D., X.-L. Wu, et al. (2010). "A non-parametric mixture model for genome-enabled prediction of genetic value for a quantitative trait." Genetica **138**(9): 959-977.

Gilmour, A. R., B. J. Gogel, et al. (2009). "2009 ASReml User Guide Release 3.0 " VSN International Ltd, Hemel Hempstead, HP1 1ES, UK.

Goddard, M. E. and B. J. Hayes (2009). "Mapping genes for complex traits in domestic animals and their use in breeding programmes." Nature Reviews Genetics **10**(6): 381-391.

Goddard, M. E., N. R. Wray, et al. (2009). "Estimating Effects and Making Predictions from Genome-Wide Marker Data." Statistical Science **24**(4): 517-529.

Grapes, L., J. C. M. Dekkers, et al. (2004). "Comparing linkage disequilibrium-based methods for fine mapping quantitative trait loci." Genetics **166**(3): 1561-1570.

Habier, D., R. L. Fernando, et al. (2007). "The Impact of Genetic Relationship Information on Genome-Assisted Breeding Values." Genetics **177**(4): 2389-2397.

Habier, D., R. L. Fernando, et al. (2011). "Extension of the Bayesian alphabet for genomic selection." Bmc Bioinformatics **12**: 186.

Hadjipavlou, G. and S. C. Bishop (2009). "Age-dependent quantitative trait loci affecting growth traits in Scottish Blackface sheep." Animal Genetics **40**(2): 165-175.

Hayes, B. and M. E. Goddard (2001). "The distribution of the effects of genes affecting quantitative traits in livestock." Genetics Selection Evolution **33**(3): 209-229.

Hayes, B. J., P. J. Bowman, et al. (2009). "Invited review: Genomic selection in dairy cattle: Progress and challenges." J Dairy Sci **92**(2): 433 - 443.

Henderson, C. R. (1976). "A Simple Method for Computing the Inverse of a Numerator Relationship Matrix Used in Prediction of Breeding Values." Biometrics **32**(1): 69-83.

Henderson, C. R. (1984). Applications of Linear Models in Animal Breeding. Guelph, Canada, University of Guelph.

Hickey, J. M. and G. Gorjanc (2012). "Simulated Data for Genomic Selection and Genome-Wide Association Studies Using a Combination of Coalescent and Gene Drop Methods." G3-Genes Genomes Genetics **2**(4): 425-427.

Hill, W. G., M. E. Goddard, et al. (2008). "Data and Theory Point to Mainly Additive Genetic Variance for Complex Traits." PLoS Genet **4**(2): e1000008.

Hoggart, C. J., J. C. Whittaker, et al. (2008). "Simultaneous Analysis of All SNPs in Genome-Wide and Re-Sequencing Association Studies." PLoS Genetics **4**(7): e1000130.

Jarmila, B., M. Sargolzaei, et al. (2010). "Characteristics of linkage disequilibrium in North American Holsteins." BMC Genomics **11**: 11.

Jia, Y. and J. L. Jannink (2012). "Multiple-Trait Genomic Selection Methods Increase Genetic Value Prediction Accuracy." Genetics **192**(4): 1513-+.

Karkkainen, H. P. and M. J. Sillanpaa (2012). "Back to Basics for Bayesian Model Building in Genomic Selection." Genetics **112**(139014).

Kizilkaya, K., P. Carnier, et al. (2003). "Cumulative t-link threshold models for the genetic analysis of calving ease scores." Genetics Selection Evolution **35**(6): 489 - 512.

Kizilkaya, K. and R. J. Tempelman (2005). "A general approach to mixed effects modeling of residual variances in generalized linear mixed models." Genetics Selection Evolution **37**(1): 31-56.

Knap, P. W. and G. Su (2008). "Genotype by environment interaction for litter size in pigs as quantified by reaction norms analysis." Animal **2**(12): 1742-1747.

Lee, S. H., J. H. van der Werf, et al. (2008). "Predicting unobserved phenotypes for complex traits from whole-genome SNP data." PLoS Genet **4**(10): e1000231.

Legarra, A. and V. Ducrocq (2012). "Computational strategies for national integration of phenotypic, genomic, and pedigree data in a single-step best linear unbiased prediction." Journal of Dairy Science **95**(8): 4629-4645.

Legarra, A. and I. Misztal (2008). "Technical note: Computing strategies in genome-wide selection." Journal of Dairy Science **91**(1): 360-366.

Legarra, A., C. Robert-Granie, et al. (2008). "Performance of Genomic Selection in Mice." Genetics **180**(1): 611-618.

Lillehammer, M., M. Arnyasi, et al. (2007). "A genome scan for quantitative trait locus by environment interactions for production traits." Journal of Dairy Science **90**(7): 3482-3489.

Lillehammer, M., M. E. Goddard, et al. (2008). "Quantitative trait locus-by-environment interaction for milk yield traits on Bos taurus autosome 6." Genetics **179**(3): 1539-1546.

Lillehammer, M., B. J. Hayes, et al. (2009). "Gene by environment interactions for production traits in Australian dairy cattle." Journal of Dairy Science **92**(8): 4008-4017.

Lillehammer, M., J. Odegard, et al. (2007). "Random regression models for detection of gene by environment interaction." Genetics Selection Evolution **39**(2): 105-121.

Liu, J. S. (1994). "The Collapsed Gibbs Sampler in Bayesian Computations with Applications to a Gene-Regulation Problem." Journal of the American Statistical Association **89**(427): 958-966.

Logsdon, B., G. Hoffman, et al. (2010). "A variational Bayes algorithm for fast and accurate multiple locus genome-wide association analysis." BMC Bioinformatics **11**(1): 58.

Lorenz, A. J., S. M. Chao, et al. (2011). Genomic Selection in Plant Breeding: Knowledge and Prospects. Advances in Agronomy, Vol 110. San Diego, Elsevier Academic Press Inc. **110**: 77-123.

Mattar, M., L. O. C. Silva, et al. (2011). "Genotype x environment interaction for long-yearling weight in Canchim cattle quantified by reaction norm analysis." Journal of Animal Science **89**(8): 2349-2355.

Meuwissen, T. and M. Goddard (2010). "Accurate Prediction of Genetic Values for Complex Traits by Whole-Genome Resequencing." Genetics **185**(2): 623-631.

Meuwissen, T. H. E., B. J. Hayes, et al. (2001). "Prediction of total genetic value using genome-wide dense marker maps." Genetics **157**(4): 1819-1829.

Misztal, I., S. E. Aggrey, et al. (2013). "Experiences with a single-step genome evaluation." Poultry Science **92**(9): 2530-2534.

Misztal, I., A. Legarra, et al. (2009). "Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information." Journal of Dairy Science **92**(9): 4648-4655.

Muller, P. (1991). "A generic approach to posterior integration and Gibbs sampling. Technical Report 91-09." Retrieved March 12, 2013, from http://www.stat.purdue.edu/research/technical_reports/1991-tr.html.

Munilla, S. and R. J. C. Cantet (2012). "Bayesian conjugate analysis using a generalized inverted Wishart distribution accounts for differential uncertainty among the genetic parameters - an application to the maternal animal model." Journal of Animal Breeding and Genetics **129**(3): 173-187.

Musani, S. K., H. G. Zhang, et al. (2006). "Principal component analysis of quantitative trait loci for immune response to adenovirus in mice." Hereditas **143**(1): 189-197.

Ntzoufras, I. (2011). Bayesian Modeling Using WinBugs, John Wiley & Sons.

O'Hara, R. B. and M. J. Sillanpaa (2009). "A Review of Bayesian Variable Selection Methods: What, How and Which." Bayesian Analysis **4**(1): 85-117.

Pinheiro, J. C., C. H. Liu, et al. (2001). "Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate t distribution." Journal of Computational and Graphical Statistics **10**(2): 249-276.

Plummer, M., N. Best, et al. (2006). "CODA: convergence diagnostics and output analysis for MCMC." R News **6**(1): 7-11.

Pourahmadi, M. (1999). "Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation." Biometrika **86**(3): 677-690.

Resende, M. F. R., P. Munoz, et al. (2012). "Accuracy of Genomic Selection Methods in a Standard Data Set of Loblolly Pine (*Pinus taeda* L.)." Genetics **190**(4): 1503-+.

Riedelsheimer, C., F. Technow, et al. (2012). "Comparison of whole-genome prediction models for traits with contrasting genetic architecture in a diversity panel of maize inbred lines." BMC Genomics **13**(1): 452.

Shariati, M. and D. Sorensen (2008). "Efficiency of alternative MCMC strategies illustrated using the reaction norm model." Journal of Animal Breeding and Genetics **125**(3): 176-186.

Shepherd, R. K., T. H. Meuwissen, et al. (2010). "Genomic selection and complex trait prediction using a fast EM algorithm applied to genome-wide markers." BMC Bioinformatics **11**: 529.

Sorensen, D. and D. Gianola (2002). Likelihood, Bayesian, and MCMC methods in quantitative genetics. New York, Springer-Verlag.

Strandén, I. and D. J. Garrick (2009). "Technical note: Derivation of equivalent computing algorithms for genomic predictions and reliabilities of animal merit." Journal of Dairy Science **92**(6): 2971-2975.

Streit, M., F. Reinhardt, et al. (2012). "Reaction norms and genotype-by-environment interaction in the German Holstein dairy cattle." Journal of Animal Breeding and Genetics **129**(5): 380-389.

Streit, M., R. Wellmann, et al. (2013). "Using Genome-Wide Association Analysis to Characterize Environmental Sensitivity of Milk Traits in Dairy Cattle." G3-Genes Genomes Genetics **3**(7): 1085-1093.

Su, G., P. Madsen, et al. (2006). "Bayesian analysis of the linear reaction norm model with unknown covariates." Journal of Animal Science **84**(7): 1651-1657.

Technow, F. (2012). "R Package Hypred: Simulation of Genomic Data in Applied Genetics."

Technow, F. and A. Melchinger (2013). "Genomic prediction of dichotomous traits with Bayesian logistic models." Theoretical and Applied Genetics: 1-11.

Technow, F., C. Riedelsheimer, et al. (2012). "Genomic prediction of hybrid performance in maize with models incorporating dominance and population specific marker effects." Theoretical and Applied Genetics **125**(6): 1181-1194.

Valdar, W., L. C. Solberg, et al. (2006). "Genome-wide genetic association of complex traits in heterogeneous stock mice." Nature Genetics **38**(8): 879-887.

Valdar, W., L. C. Solberg, et al. (2006). "Genetic and Environmental Effects on Complex Traits in Mice." Genetics **174**(2): 959-984.

van Binsbergen, R., R. F. Veerkamp, et al. (2012). "Makeup of the genetic correlation between milk production traits using genome-wide single nucleotide polymorphism information." Journal of Dairy Science **95**(4): 2132-2143.

VanRaden, P. M. (2008). "Efficient Methods to Compute Genomic Predictions." Journal of Dairy Science **91**(11): 4414-4423.

Vaughan, L. K., J. Divers, et al. (2009). "The use of plasmodes as a supplement to simulations: A simple example evaluating individual admixture estimation methodologies." Computational Statistics & Data Analysis **53**(5): 1755-1766.

Vazquez, A. I., G. de los Campos, et al. (2012). "A Comprehensive Genetic Approach for Improving Prediction of Skin Cancer Risk in Humans." Genetics **192**(4): 1493-1502.

Verbyla, K. L., B. J. Hayes, et al. (2009). "Accuracy of genomic selection using stochastic search variable selection in Australian Holstein Friesian dairy cattle." Genetics Research **91**(5): 307-311.

Vichi, M. and G. Saporta (2009). "Clustering and disjoint principal component analysis." Computational Statistics & Data Analysis **53**(8): 3194-3208.

Villumsen, T. M., L. Janss, et al. (2008). "The importance of haplotype length and heritability using genomic selection in dairy cattle." Journal of Animal Breeding and Genetics **126**: 3-13.

Waagepetersen, R., N. Ibanez-Escriche, et al. (2008). "A comparison of strategies for Markov chain Monte Carlo computation in quantitative genetics." Genetics Selection Evolution **40**(2): 161-176.

Wang, C. L., X. D. Ding, et al. (2013). "Bayesian methods for estimating GEBVs of threshold traits." Heredity **110**(3): 213-219.

Wang, C. S., J. J. Rutledge, et al. (1994). "Bayesian-Analysis of Mixed Linear-Models Via Gibbs Sampling with an Application to Litter Size in Iberian Pigs." Genetics Selection Evolution **26**(2): 91-115.

- Wang, H., I. Misztal, et al. (2012). "Genome-wide association mapping including phenotypes from relatives without genotypes." Genetics Research **94**(2): 73-83.
- Weller, J. I., G. R. Wiggans, et al. (1996). "Application of a canonical transformation to detection of quantitative trait loci with the aid of genetic markers in a multi-trait experiment." Theoretical and Applied Genetics **92**(8): 998-1002.
- Wiggans, G. R., P. M. VanRaden, et al. (2011). "The genomic evaluation system in the United States: Past, present, future." Journal of Dairy Science **94**(6): 3202-3211.
- Wimmer, V., C. Lehermeier, et al. (2013). "Genome-Wide Prediction of Traits with Different Genetic Architecture Through Efficient Variable Selection." Genetics **195**(2): 573-+.
- Yang, W. and R. J. Tempelman (2012). "A Bayesian antedependence model for whole genome prediction." Genetics **190**(4): 1491-1501.
- Yi, N. and S. Xu (2008). "Bayesian Lasso for quantitative trait loci mapping." Genetics **179**(2): 1045 - 1055.
- Zhu, W. S. and H. P. Zhang (2009). "Rejoinder: Why do we test multiple traits in genetic association studies?" Journal of the Korean Statistical Society **38**(1): 25-27.
- Zimmerman, D. L. and V. A. Núñez-Antón (2010). Antedependence Models for Longitudinal Data, Chapman and Hall/CRC.