Journal of Theoretical Biology 417 (2017) 131–141

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

Joint genome-wide prediction in several populations accounting for randomness of genotypes: A hierarchical Bayes approach. II: Multivariate spike and slab priors for marker effects and derivation of approximate Bayes and fractional Bayes factors for the complete family of models

Carlos Alberto Martínez[a,b,*], Kshitij Khare[b], Arunava Banerjee[c], Mauricio A. Elzo[a]

[a] Department of Animal Sciences, University of Florida, Gainesville, FL, USA
[b] Department of Statistics, University of Florida, Gainesville, FL, USA
[c] Department of Computer and Information Science and Engineering, University of Florida, Gainesville, FL, USA

ARTICLE INFO

Keywords:
Bayesian whole-genome regressions
Finite mixture priors
Genetic heterogeneity
Laplace approximation
Multi-population genome-enabled prediction

ABSTRACT

This study corresponds to the second part of a companion paper devoted to the development of Bayesian multiple regression models accounting for randomness of genotypes in across population genome-wide prediction. This family of models considers heterogeneous and correlated marker effects and allelic frequencies across populations, and has the ability of considering records from non-genotyped individuals and individuals with missing genotypes in any subset of loci without the need for previous imputation, taking into account uncertainty about imputed genotypes. This paper extends this family of models by considering multivariate spike and slab conditional priors for marker allele substitution effects and contains derivations of approximate Bayes factors and fractional Bayes factors to compare models from part I and those developed here with their null versions. These null versions correspond to simpler models ignoring heterogeneity of populations, but still accounting for randomness of genotypes. For each marker loci, the spike component of priors corresponded to point mass at $\mathbf{0}$ in $\mathbb{R}^S$, where S is the number of populations, and the slab component was a S-variate Gaussian distribution, independent conditional priors were assumed. For the Gaussian components, covariance matrices were assumed to be either the same for all markers or different for each marker. For null models, the priors were simply univariate versions of these finite mixture distributions. Approximate algebraic expressions for Bayes factors and fractional Bayes factors were found using the Laplace approximation. Using the simulated datasets described in part I, these models were implemented and compared with models derived in part I using measures of predictive performance based on squared Pearson correlations, Deviance Information Criterion, Bayes factors, and fractional Bayes factors. The extensions presented here enlarge our family of genome-wide prediction models making it more flexible in the sense that it now offers more modeling options.

1. Introduction

The scenario of across population genome-wide prediction accounting for randomness of genotypes was addressed in part I of our series of studies. There, we adopted a hierarchical Bayesian modeling strategy to accommodate heterogeneous and correlated marker effects across subpopulations and random genotypes. In that companion paper we provided a detailed derivation of the joint pmf of the genotypes conditional on pedigree information and allelic frequencies and also discussed some of its properties. Furthermore, the flexibility of hierarchical Bayesian modeling allowed us to account for heterogeneous and correlated allelic frequencies. The "MG-GBLUP" model proposed by Lehermeir et al. (2015) is similar to the models developed in part I of this study, except that they did not consider randomness of genotypes. In addition, they did not consider models with different (heterogeneous) covariance matrices of marker effects. One of the main properties of our models is that individuals with phenotypic records and missing genotypes at any subset of loci (including non-genotyped individuals) can be considered in the analysis without previous imputation. Furthermore, due to the use of a Bayesian approach, uncertainty about imputed genotypes is automatically taken into account.

The so called "spike and slab" priors, are finite mixtures of a continuous distribution (the slab) and a mass point at some constant

* Corresponding author at: Department of Animal Sciences, University of Florida, Gainesville, FL 32611, USA.
E-mail address: carlosmn@ufl.edu (C.A. Martínez).

http://dx.doi.org/10.1016/j.jtbi.2016.12.022
Received 9 September 2016; Received in revised form 30 November 2016; Accepted 28 December 2016
Available online 12 January 2017

(the spike) (Mitchell and Beauchamp, 1988). A particular case of these priors are the zero-inflated priors which have point mass at zero. This sort of priors has been used in high dimensional problems to induce a stronger shrinkage and perform variable selection. In single population analyses, it has been reported that when there are genes with major effects controlling the trait under study or the number of genes controlling the trait is low, Bayesian variable selection models tend to perform better (Daetwyler et al., 2012; Heslot et al., 2012; Gianola and Rosa, 2015). In the case of multiple population analyses, van den Berg et al. (2015) studied scenarios under which Bayesian variable selection models outperformed genomic BLUP (GBLUP). They found that GBLUP was outperformed when the number of QTL was smaller than the number of independent chromosome segments. They also found that the difference in accuracy between these models was larger than in the single population case.

In a Bayesian framework, model comparison can be performed via Bayes factors and some modifications of them known as non-subjective Bayes factors (Ghosh et al., 2006). Bayes factors measure the change in the odds favoring a model once data are observed (Lavine and Schervish, 1999). On the other hand, O'Hagan (1994, 1995) proposed a non-subjective Bayes factor known as fractional Bayes factor which uses a fractional part of the likelihood resulting in a "partial" Bayes factor. Analytical forms of Bayes factors involve integration of the joint distribution of data and parameters over the parameter space of a given model to obtain marginal likelihoods, and even for some simple models these integrals do not have a closed form solution. One option to obtain algebraic approximations is to use the Laplace approximation after arranging the integrand in an appropriate form (Ghosh et al., 2006). Another criterion to compare models is the Deviance Information Criterion (DIC, Spiegelhalter et al., 2002, 2014) which combines measures of model fit and model complexity and, despite some limitations, it has been used in several research areas (Spiegelhalter et al., 2014).

Thus, the objectives of this study were to extend the family of models developed in a companion paper (part I) by considering the so called spike and slab priors for marker effects and to derive approximate expressions for Bayes factors and fractional Bayes factors to compare the proposed models with their corresponding null versions that ignore population structure.

2. Methods

2.1. The models

The complete population or simply the population is defined as the set of individuals with phenotypes considered in the analysis, which is comprised by a set of $\mathcal{S}$ subpopulations defined by some criterion like environment, race, breed, line, etc. Also the following assumptions are made: linkage equilibrium, Hardy-Weinberg equilibrium, no mutation, and starting from the oldest individuals with phenotypes, the pedigree is fully known.

The following is the linear model describing the relationship between records and mappings of marker genotypes: $\boldsymbol{y}=W\boldsymbol{g}+\boldsymbol{e}$, where $\boldsymbol{y}\in\mathbb{R}^n$ is a vector containing response variables (e.g., records corrected for non-genetic factors), W_{nxsm} is an observable random matrix with entries corresponding to a one to one mapping from the set of individual marker genotypes to a subset of the integers (defined later), $\boldsymbol{g}\in\mathbb{R}^{sm}$ is an unknown random vector of average marker allele substitution effects for every population and $\boldsymbol{e}\in\mathbb{R}^n$ is a random vector of residuals. If records are sorted by subpopulation as well as the columns of W and the elements of $\boldsymbol{g}$, then for every $l=1, 2,\ldots,\mathcal{S}$, $\boldsymbol{y}_l=W_l\boldsymbol{g}_l+\boldsymbol{e}_l$, with dimensions: $(\boldsymbol{y}_l)_{n_l\times 1}$, $(W_l)_{n_l\times m}$, $(\boldsymbol{g}_l)_{m\times 1}$ and $(\boldsymbol{e}_l)_{n_l\times 1}$ where n_l is the sample size of subpopulation l, and m is the number of marker loci; therefore, $n=\sum_{l=1}^{\mathcal{S}} n_l$.

In our models, the mapping from the set of genotypes at each locus and each individual into a subset of the integers is defined as follows, biallelic loci are considered. If A and B are the marker alleles at each locus and B is considered the reference allele then:

$$W_l=\{w_{ij}^l\}_{n_l\times m}=\begin{cases}1, & \text{if genotype} = BB\\ 0, & \text{if genotype} = AB\\ -1, & \text{if genotype} = AA\end{cases}.$$

The following is the hierarchical representation of our models. Let $R=(\sigma_{e1}^2,\ldots,\sigma_{e\mathcal{S}}^2)$ and $V=Block\,Diag.\{\sigma_{el}^2 I_{n_l}\}_{l=1}^{\mathcal{S}}$ then

$$\boldsymbol{y}|W,\boldsymbol{g},R\sim MVN(W\boldsymbol{g},\quad V)$$

$$W|\boldsymbol{p}_1^*,\boldsymbol{p}_2^*,\ldots,\boldsymbol{p}_m^*\sim\pi(\bullet|\boldsymbol{p}_1^*,\boldsymbol{p}_2^*,\ldots,\boldsymbol{p}_m^*)$$

$$\boldsymbol{p}_j^*\overset{iid}{\sim}\pi(\boldsymbol{p}^*),\ j=1,2,\ldots,m$$

$$\sigma_{e1}^2,\ldots,\sigma_{e\mathcal{S}}^2\overset{iid}{\sim}Inverse\ Gamma\left(\frac{\tau^2}{2},\frac{\nu}{2}\right):=IG\left(\frac{\tau^2}{2},\frac{\nu}{2}\right)$$

$$\boldsymbol{g}_j|G_j,\pi_0\overset{ind}{\sim}\begin{cases}Point\ mass\ at\ \mathbf{0}\ with\ probability\ \pi_0\\ MVN(\mathbf{0},\ G_j)\ with\ probability\ 1-\pi_0\end{cases}$$

$$G_j\overset{iid}{\sim}Inverse\ \ Wishart(a,\boldsymbol{\Sigma}):=IW(a,\boldsymbol{\Sigma})$$

$$G_j=\begin{bmatrix}\sigma_{j_1}^2 & \sigma_{j_{1,2}} & \cdots & \sigma_{j_{1,\mathcal{S}}}\\ & \sigma_{j_2}^2 & \cdots & \sigma_{j_{2,\mathcal{S}}}\\ & & \ddots & \vdots\\ sym & & & \sigma_{j_{\mathcal{S}}}^2\end{bmatrix}$$

where σ_{el}^2 is the residual variance in subpopulation l, $\sigma_{j_l}^2$ is the variance of the effect of the j^{th} marker in the l^{th} subpopulation, $\sigma_{j_{l,l'}}$ is the covariance between effects of marker j in subpopulations l and l', $\boldsymbol{p}_j^*$ is a parameter related to allelic frequencies of the j^{th} marker in each subpopulation and $\pi(\boldsymbol{p}^*)$ is its probability density function (pdf). This set of parameters and their pdf are described in part I of this series of papers. Here, parameter π_0 was assumed to be known.

The model presented above assumed a different covariance matrix for the vector of allele substitution effects for each marker in the slab component of the mixture distribution and consequently this sort of models will be referred to as heterogeneous marker effects covariance matrix models. On the other hand, models with $G_1=\ldots=G_m=G^0$ will be referred to as homogeneous marker effects covariance matrix models. Moreover, the special case $\sigma_{e1}^2=\ldots=\sigma_{e\mathcal{S}}^2=\sigma^2$ corresponds to that of models with homoscedastic residuals.

In part I, it was discussed that the scenario of completely (i.e., at all loci) or partially missing genotypes can be handled because of the use of the pmf $\pi(W|P^*)$, $P^*=(\boldsymbol{p}_1^*,\boldsymbol{p}_2^*,\ldots,\boldsymbol{p}_m^*)$ and the fact that these missing genotypes are regarded as model parameters. There, it was also shown that the likelihood can be written as $f(\boldsymbol{y},W^o|W^N,\boldsymbol{g},R,P^*)=f(\boldsymbol{y}|W,\boldsymbol{g},R)f(W^o|W^N,P^*)$ where W^o is the fraction of W corresponding to observed genotypes, W^N the fraction corresponding to missing genotypes, and $f(\boldsymbol{y}|W,\boldsymbol{g},R)$ and $f(W^o|W^N,P^*)$ are referred to as the $\boldsymbol{y}$ component and the W component of the likelihood.

The conditional prior for $\boldsymbol{g}_j$ can be written as:

$$\pi(\boldsymbol{g}_j|G_j,\pi)=\pi_0 I_{\{\boldsymbol{g}_j=\mathbf{0}\}}+(1-\pi_0)MVN(\boldsymbol{g}_j;0,\quad G_j)I_{\{\boldsymbol{g}_j\neq\mathbf{0}\}}$$

where $I_{\{\bullet\}}$ is the indicator function. This form is more convenient from the algebraic point of view because it allows carrying out computations and writing expressions for the joint conditional prior in an easier way. Under the conditional independence assumption, the joint conditional prior for $\boldsymbol{g}$ is:

$$\pi(\boldsymbol{g}|G,\pi_0)=\prod_{j=1}^{m}\left\{\pi_0 I_{\{\boldsymbol{g}_j=\mathbf{0}\}}+(1-\pi_0)MVN(\boldsymbol{g}_j;0,\quad G_j)I_{\{\boldsymbol{g}_j\neq\mathbf{0}\}}\right\}.$$

An explicit form of this prior pdf can be found as follows. Let $i=0,1,\ldots,m$ be the number of markers having a null effect.

Consequently, when expanding the product above, for each i there are $\binom{m}{i}$ combinations of i markers with null effect chosen from m markers. For $l = 1,2,\ldots,\binom{m}{i}$, let δ_{il} denote the event that the l^{th} subset of i markers (i.e., the l^{th} combination of i markers with null effect chosen from the total set of m markers) have null effect and $I_{\delta_{il}}$ the indicator function of this event. Thus, there are $\binom{m}{i}$ terms in the expansion with π_0 appearing exactly i times; therefore, each one of these $\binom{m}{i}$ terms is of the form:

$$I_{\delta_{il}}\,\pi_0^i(1-\pi_0)^{m-i}\prod_{k:g_k\in\delta_{il0}^c} MVN(\boldsymbol{g}_k;\mathbf{0},G_k)$$

where δ_{il0} is the set of marker loci with null effects given δ_{il}, and δ_{il0}^c is its complement, i.e., the set of $m-i$ markers with non-null effect under δ_{il}. Therefore when expanding $\pi(\boldsymbol{g}|G,\pi_0)$ for the heterogeneous marker effect covariance matrix model:

$$\pi(\boldsymbol{g}|G,\pi_0)=\sum_{i=0}^{m}\pi_0^i(1-\pi_0)^{m-i}\sum_{l=1}^{\binom{m}{i}}I_{\delta_{il}}\prod_{k:g_k\in\delta_{il0}^c}MVN(\boldsymbol{g}_k;\mathbf{0},G_k),$$

while for the homogeneous marker effect covariance matrix model the expression is the same except that $G_j=G^0\,\forall\, j=1,2\ldots,m$.

Regarding the marginal priors, under homogeneous covariance matrix of marker effects:

$$\begin{aligned}\pi(\boldsymbol{g}|\pi_0)&\propto\sum_{i=0}^{m}\pi_0^i(1-\pi_0)^{m-i}\sum_{l=1}^{\binom{m}{i}}I_{\delta_{il}}\\&\times\int_{\mathcal{P}_{\mathcal{S}}^+}|G^0|^{\frac{-a+\mathcal{S}+m-i+1}{2}}\exp\left(\frac{-1}{2}tr\left(\left(\boldsymbol{\Sigma}+\sum_{k:\boldsymbol{g}_k\in\delta_{il0}^c}\boldsymbol{g}_k\boldsymbol{g}_k'\right)(G^0)^{-1}\right)\right)dG^0\\&\propto\sum_{i=0}^{m}\pi_0^i(1-\pi_0)^{m-i}\sum_{l=1}^{\binom{m}{i}}I_{\delta_{il}}2^{\mathcal{S}(m-i)/2}\Gamma_{\mathcal{S}}\left(\frac{a+m-i}{2}\right)\\&\times\left|\boldsymbol{\Sigma}+\sum_{k:g_k\in\delta_{il0}^c}\boldsymbol{g}_k\boldsymbol{g}_k'\right|^{-\left(\frac{a+m-i}{2}\right)}.\end{aligned}$$

Hence, marker effects are not marginally independent *a priori* and their joint marginal prior distribution is a mixture of non-standard distributions with mixing probabilities $\pi_0^i(1-\pi_0)^{m-i}$.

For heterogeneous marker effect covariance matrix model:

$$\begin{aligned}\pi(\boldsymbol{g}|\pi_0)&\propto\sum_{i=0}^{m}\pi_0^i(1-\pi_0)^{m-i}\sum_{l=1}^{\binom{m}{i}}I_{\delta_{il}}2^{-\mathcal{S}i/2}\Gamma_{\mathcal{S}}\left(\frac{a}{2}\right)^i\Gamma_{\mathcal{S}}\left(\frac{a+1}{2}\right)^{m-i}\\&\times\prod_{k:g_k\in\delta_{il0}^c}\frac{1}{\left|1+\frac{\boldsymbol{g}_k'\boldsymbol{\Sigma}_*^{-1}\boldsymbol{g}_k}{a+1-\mathcal{S}}\right|^{\left(\frac{a+1}{2}\right)}}.\end{aligned}$$

This is a mixture distribution with mixing probabilities $\pi_0^i(1-\pi_0)^{m-i}$. Each component in the mixture is a sum of $\binom{m}{i}$ elements. Each one of these elements is the product of $m-i$ multivariate t distributions with scale matrix $\boldsymbol{\Sigma}_*=\frac{1}{a+1-\mathcal{S}}\boldsymbol{\Sigma}$ and degrees of freedom $a+1-\mathcal{S}$ for non-null vectors of markers effects, and point mass at zero for i null vectors of marker effects, under event δ_{il}. In this case, marker effects are marginally independent *a priori*.

2.2. Full conditionals

Only full conditionals that change with respect to those considered in part I are presented.

$$\pi(\boldsymbol{g}|Else)=$$

$$\sum_{i=0}^{m}\pi_0^i(1-\pi_0)^{m-i}\sum_{l=1}^{\binom{m}{i}}I_{\delta_{il}}MVN\left(\boldsymbol{g}_{\delta_{il0}^c};\left(\frac{W_{\delta_{il0}^c}'W_{\delta_{il0}^c}}{\sigma^2}+G_{\delta_{il0}^c}^{-1}\right)^{-1}\frac{W_{\delta_{il0}^c}'\boldsymbol{y}}{\sigma^2},\left(\frac{W_{\delta_{il0}^c}'W_{\delta_{il0}^c}}{\sigma^2}+G_{\delta_{il0}^c}^{-1}\right)^{-1}\right)$$

where $\boldsymbol{g}_{\delta_{il0}^c}=(\boldsymbol{g}_{k_1}'\ \cdots\ \boldsymbol{g}_{k_{m-i}}')'$, $k{:}\boldsymbol{g}_k\in\delta_{il0}^c$, corresponds to the vector of dimension $\mathcal{S}(m-i)$ with the non-null marker effects under δ_{il}, $W_{\delta_{il0}^c}$ is the submatrix of the design matrix corresponding to $\boldsymbol{g}_{\delta_{il0}^c}$ and $G_{\delta_{il0}^c}^{-1}=I_{m-i}\bigotimes G_0^{-1}, i=0,1,\ldots,m$.

Remark 1 Notice that each element in the summation above corresponds to a multivariate normal distribution of dimension $\mathcal{S}(m-i)$ for those markers in δ_{il0}^c and point mass at zero for those markers in δ_{il0}. Thus, in each term, the multivariate normal corresponds to the distribution of the effects of the subset of markers with non-null effects given δ_{il}. Therefore, this joint full conditional distribution of $\boldsymbol{g}$ suggests that for each marker, the full conditional distribution of $\boldsymbol{g}_j$ (given data, and other parameters in the model including the remaining components of $\boldsymbol{g}$) is a spike and slab distribution. Note that it is easier to deal with $\pi(\boldsymbol{g}_j|Else)$ than with $\pi(\boldsymbol{g}|Else)$. The full conditional $\pi(\boldsymbol{g}_j|Else)$ can be found from $\pi(\boldsymbol{g}|Else)$ using the Bayes theorem. However, this could be complex because it requires identifying all the cases in which $\boldsymbol{g}_j=0$ and all the cases in which $\boldsymbol{g}_j\neq 0$. An easier way is to derive it using the conditional prior for $\boldsymbol{g}_j$. Details are presented in Appendix A. The final result is:

$$\begin{aligned}\pi(\boldsymbol{g}_j|Else)=&\\p(\boldsymbol{g}_j=0|Else)I_{\{\boldsymbol{g}_j=\mathbf{0}\}}&+(1-p(\boldsymbol{g}_j=0|Else))MVN\\&\left(G_{Fj}^{-1}\frac{W_j'}{\sigma^2}(\boldsymbol{y}-W_{(-j)}\boldsymbol{g}_{(-j)}),\,G_{Fj}^{-1}\right)I_{\{\boldsymbol{g}_j\neq\mathbf{0}\}}\end{aligned}$$

$$G_{Fj}=\frac{W_j'W_j}{\sigma^2}+(G^0)^{-1}$$

$$\begin{aligned}p(\boldsymbol{g}_j=0|Else)&\\=\frac{\pi_0}{\pi_0+(1-\pi_0)(|G_{Fj}||G^0|)^{-1/2}\exp\left(\frac{1}{2\sigma^2}\|G_{*Fj}^{-1/2}W_j'(\boldsymbol{y}-W_{(-j)}\boldsymbol{g}_{(-j)})\|_2^2\right)},&\end{aligned}$$

where $G_{*Fj}=\sigma^2G_{Fj}=W_j'W_j+\sigma^2(G^0)^{-1}$. Thus, the full conditional distribution of $\boldsymbol{g}_j$ is a spike and slab distribution where the slab component is a $MVN(G_{Fj}^{-1}W_j'(\boldsymbol{y}-W_{(-j)}\boldsymbol{g}_{(-j)}),\,G_{Fj}^{-1})$ and the spike is a point mass at 0 in $\mathbb{R}^{\mathcal{S}}$. On the other hand,

$$\pi(G^0|Else)\propto\sum_{i=0}^{m}\boldsymbol{\pi}^i(1-\boldsymbol{\pi})^{m-i}\sum_{l=1}^{\binom{m}{i}}I_{\delta_{il}}|G^0|^{-\frac{(m-i+a+\mathcal{S}+1)}{2}}$$

$$\times\exp\left(\frac{-1}{2}tr\left(\left(\boldsymbol{\Sigma}+\sum_{k:g_k\in\delta_{il0}^c}\boldsymbol{g}_k\boldsymbol{g}_k'\right)(G^0)^{-1}\right)\right),$$

a mixture of sums of inverse Wishart distributions with mixing probabilities $\boldsymbol{\pi}^i(1-\boldsymbol{\pi})^{m-i}, i=0,1,\ldots,m$. The i^{th} component of the mixture is the sum of $\binom{m}{i}$ inverse Wishart distributions with parameters $\left(m-i+a,\,\boldsymbol{\Sigma}+\sum_{k:g_k\in\delta_{il0}^c}\boldsymbol{g}_k\boldsymbol{g}_k'\right)$, $l=1,2,\ldots,\binom{m}{i}$.

For the heterogeneous marker effect covariance matrix model the full conditional $\pi(\boldsymbol{g}_j|Else)$ has the same form as for the homogeneous marker effect covariance matrix model except that now $G_{Fj}=\frac{W_j'W_j}{\sigma^2}+G_j^{-1}$ and $G_{*Fj}=W_j'W_j+\sigma^2G_j^{-1}$ and

$$\pi(G_j|Else) = \begin{cases} IW(a+1, \boldsymbol{\Sigma} + \boldsymbol{g}_j\boldsymbol{g}_j^{'}), & if\ \boldsymbol{g}_j \neq 0 \\ IW(a, \boldsymbol{\Sigma}), & if\ \boldsymbol{g}_j = 0 \end{cases}$$

The expressions for models with heteroscedastic residuals are very similar and therefore these are omitted. Such expressions can be found in Appendix A along with joint posterior distributions.

2.3. Model comparison

2.3.1. Theoretical approximation to model comparison via Bayes factors and fractional Bayes factors

Here, the term null model refers to simplified versions of the proposed models in two scenarios. The first one corresponds to the case in which all data are pooled and the factor splitting the complete population into subpopulations is ignored. In the second scenario, the complete population is split into subpopulations and each one of them is analyzed independently. The null model corresponding to the first scenario was already presented in part I, and for the second scenario, the model for each subpopulation is the same, but only considering data from the corresponding subpopulation. This model is referred to as independent subpopulations model.

In order to find some theoretical approach to compare the full models with their null versions, approximate Bayes factors and fractional Bayes factors are derived in this section. To this end, analytical approximations of multivariate integrals that have to be solved to find marginal likelihoods are derived. The Laplace approximation (Ghosh et al., 2006) is used to solve some of these multivariate integrals. As will be shown in this section, the use of the Laplace approximation requires the matrix W to be of full column rank. This assumption does not hold in many real life situations where $m > n$ and therefore this matrix cannot be of full column rank. However, as more individuals are genotyped, this situation can be found more frequently, especially for chips of intermediate density. Notice that for matrix W to be of full rank, the number of observations in each subpopulation cannot be smaller than m; therefore, the requirement is that $n_l \geq m\ \forall\ l = 1,2,\ldots,\mathcal{S}$. As a matter of fact, in countries like the US there exist data sets where the number of genotyped animals exceeds the number of molecular markers in chips like the Illumina 50k (CDCB, 2016). Moreover, in certain cases, some filtering or preselection criteria reduces the set of markers to be included in the analyses and for populations with a large amount of genotyped individuals this could also lead to the full rank scenario. More comments on this will be made in the discussion. Therefore, in real life situations like across country or across breed analysis, the situation $n_l \geq m\ \forall\ l = 1,2,\ldots,\mathcal{S}$ could be observed, thus the assumption of matrix W being of full column rank could be satisfied. Of course, $n_l \geq m\ \forall\ l = 1,2,\ldots,\mathcal{S}$ is not a sufficient condition for matrix W to be of full column rank, but given the structure of this matrix, this would generally be the case except in certain situations, for example, having clones in the same subpopulation.

2.3.1.1. Bayes factors. Bayes factors have generally been interpreted as measures of support in favor of a model provided by data. Lavine and Schervish (1999) showed that Bayes factors are actually measuring the change in the odds favoring a model once data are observed. The Bayes factor comparing two models denoted as M_1 and M_0 is defined as:

$$BF_{10} = \frac{f(\boldsymbol{y}|M_1)}{f(\boldsymbol{y}|M_0)}$$

$$= \frac{\int_{\Theta_1} \pi_1(\boldsymbol{\theta}_1) f_1(\boldsymbol{y}|\boldsymbol{\theta}_1) d\boldsymbol{\theta}_1}{\int_{\Theta_0} \pi_0(\boldsymbol{\theta}_0) f_0(\boldsymbol{y}|\boldsymbol{\theta}_0) d\boldsymbol{\theta}_0}$$

where $\boldsymbol{\theta}_i$, $\pi_i(\boldsymbol{\theta}_i)$, $f_i(\boldsymbol{y}|\boldsymbol{\theta}_i)$ and Θ_i are the parameters, prior, likelihood and parametric space under model i, respectively, $i = 1,2$.

Approximate Bayes factors comparing homogenous marker effect covariance matrix models (Gaussian and spike and slab priors, homoscedastic residuals) and heterogeneous marker effect covariance matrix models (Gaussian and spike and slab priors, homoscedastic residuals) to their null versions were derived. Also, an approximate Bayes factor comparing the heterogeneous marker effect covariance matrix model with heteroscedastic residuals with the independent subpopulations model was found. These approximate Bayes factors were conditioned on the genotypes (i.e., conditioned on W and W_0). Therefore, the $\boldsymbol{y}$ component of the likelihood is used. The case when a part of W is not observed is treated at the end of this section.

A brief outline of the derivation of these approximate Bayes factors is presented. In each case, model sub-index 1 corresponds to the full model while sub-index 0 denotes the null model. The Bayes factor comparing homogeneous marker effect covariance matrix models with its null version is denoted BF_{10W} when a Gaussian prior is posed over $\boldsymbol{g}$ and residuals are homoscedastic. Whenever residuals are heteroscedastic the letter H appears in the subindex and when the prior posed over $\boldsymbol{g}$ is spike and slab the letter G is replaced by SS. Moreover, the superindex * is used to identify models with heterogeneous marker effect covariance matrices. The same subindex notation is used for fractional Bayes factors.

In general, let:

$$BF_{10W} = \frac{f(\boldsymbol{y}|W, M_1)}{f(\boldsymbol{y}|W_0, M_0)}.$$

For the homogeneous marker effect covariance matrix model $\boldsymbol{\theta}_1 := (\boldsymbol{\theta}, \boldsymbol{\phi}) = (\{\boldsymbol{g}, \sigma^2, W\}, \{G^0, P^*\})$ and $\boldsymbol{\theta}_0 := (\boldsymbol{\theta}_0^*, \boldsymbol{\phi}_0) = (\{\boldsymbol{g}_0, \sigma^2, W_0\}, \{\sigma_g^2, \boldsymbol{p}_0\})$. Let $\mathbb{R}_+$ denote the positive reals. Then:

$$\pi(\boldsymbol{y}, \boldsymbol{\theta}_1) = f(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}, \boldsymbol{\phi})$$

$$= f(\boldsymbol{y}|\boldsymbol{g}, \sigma^2, W)\pi(\boldsymbol{g}, G^0)\pi(\sigma^2)\pi(W, P^*)$$

then

$$\int_{\mathcal{P}_S^+} \int_{\mathbb{R}^{m\mathcal{S}}} \int_{\mathbb{R}_+} f(\boldsymbol{y}|\boldsymbol{g}, \sigma^2, W)\pi(\boldsymbol{g}, G^0)\pi(\sigma^2) d\sigma^2 d\boldsymbol{g} dG^0 = f(\boldsymbol{y}|W)$$

$$= \int_{\mathcal{P}_S^+} \pi(G^0) \left(\int_{\mathbb{R}^{m\mathcal{S}}} \int_{\mathbb{R}_+} f(\boldsymbol{y}|\boldsymbol{g}, \sigma^2, W)\pi(\boldsymbol{g}|G^0)\pi(\sigma^2) d\sigma^2 d\boldsymbol{g} \right) dG^0$$

Thus, the previous multiple integral has to be solved in order to find $f(\boldsymbol{y}|W)$. An analytic expression for the inner integral $\int_{\mathbb{R}^{mn}} \int_{\mathbb{R}_+} f(\boldsymbol{y}|\boldsymbol{g}, \sigma^2, W)\pi(\boldsymbol{g}|G^0)\pi(\sigma^2) d\sigma^2 d\boldsymbol{g}$ is approximated using the Laplace approximation (Ghosh et al., 2006). As shown in appendix B, after obtaining this approximation, the external integral can be found in a closed form. The Laplace method is based on a second order Taylor series expansion and allows finding an approximation to integrals of the form:

$$I = \int_{\mathbb{R}^p} q(\boldsymbol{\theta}) e^{nh(\boldsymbol{\theta})} d\boldsymbol{\theta},$$

where q and h are smooth functions of $\boldsymbol{\theta}$ and h has a unique maximum at $\hat{\boldsymbol{\theta}}$. In Bayesian statistics, $nh(\boldsymbol{\theta})$ is usually taken to be the log-likelihood or the log of the unnormalized posterior. Hence, $\hat{\boldsymbol{\theta}}$ can be the MLE or the posterior mode when the posterior is unimodal. The Laplace approximation has the form (Ghosh et al., 2006):

$$I = e^{nh(\hat{\boldsymbol{\theta}})} (2\pi)^{p/2} n^{-p/2} |\Delta_h(\hat{\boldsymbol{\theta}})|^{-1/2} q(\hat{\boldsymbol{\theta}})(1 + O(n^{-1})),$$

where $p = \dim(\boldsymbol{\theta})$ and $|\Delta_h(\hat{\boldsymbol{\theta}})|$ is the determinant of the Hessian matrix of $-h$ evaluated at $\hat{\boldsymbol{\theta}}$. The inner integral in $f(\boldsymbol{y}|W)$ can be written as:

$$\int_{\mathbb{R}^{m\mathcal{S}}} \int_{\mathbb{R}_+} \pi(\boldsymbol{g}|G^0)\pi(\sigma^2) e^{\ln f(\boldsymbol{y}|\boldsymbol{g}, \sigma^2, W)} d\sigma^2 d\boldsymbol{g} := \int_{\mathbb{R}^{m\mathcal{S}+1}} q(\boldsymbol{\theta}^*) e^{nh(\boldsymbol{\theta}^*)} d\boldsymbol{\theta}^*,$$

where $\boldsymbol{\theta}^* := (\boldsymbol{g}, \sigma^2)$.

Under the assumption that $f(\boldsymbol{y}|\boldsymbol{g}, \sigma^2, W)$ has a unique maximum at

$\hat{\boldsymbol{\theta}}^*:=(\hat{\boldsymbol{g}}, \hat{\sigma}^2)$, Laplace approximation can be used. The $\boldsymbol{y}$ component of the likelihood function is a $MVN(W\boldsymbol{g},\sigma^2 I)$. Therefore, following standard results from linear models theory, if W is of full column rank then, $\hat{\boldsymbol{g}}=(W'W)^{-1}W'\boldsymbol{y}$ is the MLE of $\boldsymbol{g}$, and $\hat{\sigma}^2=\frac{\|\boldsymbol{y}-W\hat{\boldsymbol{g}}\|^2}{n}=\frac{\boldsymbol{y}'(I-H_W)\boldsymbol{y}}{n}=\frac{(n-r)}{n}S^2$ is the MLE of σ^2, where $S^2=\frac{\boldsymbol{y}'(I-H_W)\boldsymbol{y}}{n-r}$ is the least squares estimator of σ^2, $r=rank(W'W)=m\mathcal{S}$ and $H_W=W(W'W)^{-1}W'$ is the projection matrix onto the column space of W.

After computing all the required expressions and making algebraic simplifications (see Appendix B), it follows that:

$$BF_{10GW}$$

$$\approx\left(\frac{|\boldsymbol{\Sigma}|}{b}\right)^{\frac{a}{2}}\left(\frac{\left|\boldsymbol{\Sigma}+\sum_{j=1}^{m}\hat{\boldsymbol{g}}_j\hat{\boldsymbol{g}}_j'\right|}{b+\sum_{j=1}^{m}\hat{g}_{0j}^2}\right)^{-\left(\frac{a+m}{2}\right)}\left(\frac{SSR}{SSR_0}\right)^{-\left(\frac{n+\nu+2}{2}\right)}\frac{SSR^{(m\mathcal{S}+2)/2}}{SSR_0^{(m+2)/2}}\left(\frac{|W_0'W_0|}{|W'W|}\right)^{\frac{1}{2}}$$

$$\times\exp\left(\frac{-n\tau^2}{2}\left(\frac{1}{SSR}-\frac{1}{SSR_0}\right)\right)\left(\frac{2}{n}\right)^{\frac{m(\mathcal{S}-1)}{2}}\left(\prod_{l=2}^{\mathcal{S}}\frac{\Gamma(\frac{a+m+1-l}{2})}{\Gamma(\frac{a+1-l}{2})}\right)$$

where $SSR=\boldsymbol{y}'(I-H_W)\boldsymbol{y}$, $SSR_0=\boldsymbol{y}'(I-H_{W_0})\boldsymbol{y}$, $H_{W_0}=W_0(W_0'W_0)^{-1}W_0'$, $S_0^2=\frac{\|\boldsymbol{y}-W_0\hat{\boldsymbol{g}}_0\|^2}{n-r_0}$, $r_0=rank(W_0'W_0)=m$, $\hat{\boldsymbol{g}}_0=(W_0'W_0)^{-1}W_0'\boldsymbol{y}$.

Following similar steps (see Appendix B),

$$BF_{10GW}^*=\frac{f(\boldsymbol{y}W,M_1^*)}{f(\boldsymbol{y}W_0,M_0^*)}$$

$$\approx\left(\frac{|\boldsymbol{\Sigma}|}{b}\right)^{\frac{am}{2}}\prod_{j=1}^{m}\left(\frac{|\boldsymbol{\Sigma}+\hat{\boldsymbol{g}}_j\hat{\boldsymbol{g}}_j'|}{b+\hat{g}_{0j}^2}\right)^{-\left(\frac{a+1}{2}\right)}\left(\frac{SSR}{SSR_0}\right)^{-\left(\frac{n+\nu+2}{2}\right)}\frac{SSR^{(m\mathcal{S}+2)/2}}{SSR_0^{(m+2)/2}}\left(\frac{|W_0'W_0|}{|W'W|}\right)^{\frac{1}{2}}$$

$$\times\exp\left(\frac{-n\tau^2}{2}\left(\frac{1}{SSR}-\frac{1}{SSR_0}\right)\right)\left(\frac{2}{n}\right)^{\frac{m(\mathcal{S}-1)}{2}}\left(\prod_{l=2}^{\mathcal{S}}\frac{\Gamma(\frac{a+2-l}{2})}{\Gamma(\frac{a+1-l}{2})}\right)^{m}.$$

$$BF_{10GWH}^*\approx\left(\frac{|\boldsymbol{\Sigma}|}{b^{\mathcal{S}}}\right)^{\frac{am}{2}}\prod_{j=1}^{m}\left(\frac{|\boldsymbol{\Sigma}+\hat{\boldsymbol{g}}_j\hat{\boldsymbol{g}}_j'|}{\prod_{l=1}^{\mathcal{S}}(b+\hat{g}_{0jl}^2)}\right)^{-\left(\frac{a+1}{2}\right)}\left(\prod_{l=2}^{\mathcal{S}}\frac{\Gamma(\frac{a+2-l}{2})\Gamma(\frac{a}{2})}{\Gamma(\frac{a+1-l}{2})\Gamma(\frac{a+1}{2})}\right)^{m}$$

$$BF_{10SSW}\approx(2\pi)^{m(\mathcal{S}+1)/2}\left(\frac{SSR}{SSR_0}\right)^{-\left(\frac{n+\nu+2}{2}\right)}\frac{SSR^{(m\mathcal{S}+2)/2}}{SSR_0^{(m+2)/2}}\left(\frac{|W_0'W_0|}{|W'W|}\right)^{\frac{1}{2}}$$

$$\times\exp\left(\frac{-n\tau^2}{2}\left(\frac{1}{SSR}-\frac{1}{SSR_0}\right)\right)\left(\frac{2}{n}\right)^{\frac{m(\mathcal{S}-1)}{2}}\left(\frac{|\boldsymbol{\Sigma}|}{b}\right)^{\frac{a}{2}}\frac{\Gamma(\frac{a}{2})}{\Gamma_{\mathcal{S}}(\frac{a}{2})}$$

$$\times\frac{\sum_{i=0}^{m}\sum_{l=1}^{\binom{m}{i}}I_{\delta_{il}}\boldsymbol{\pi}^i(1-\boldsymbol{\pi})^{m-i}2^{\mathcal{S}(m-i)/2}\Gamma_{\mathcal{S}}(\frac{a+m-i}{2})\left|\boldsymbol{\Sigma}+\sum_{k:g_k\in\delta_{l0}^c}\hat{\boldsymbol{g}}_k\hat{\boldsymbol{g}}_k'\right|^{-\left(\frac{a+m-i}{2}\right)}}{\sum_{i=0}^{m}\sum_{l=1}^{\binom{m}{i}}I_{\delta_{il}}\boldsymbol{\pi}^i(1-\boldsymbol{\pi})^{m-i}2^{(m-i)/2}\Gamma(\frac{a+m-i}{2})\left|b+\sum_{k:g_k\in\delta_{l0}^c}\hat{g}_{k0}^2\right|^{-\left(\frac{a+m-i}{2}\right)}}$$

$$BF_{10SSW}^*\approx(2\pi)^{m(\mathcal{S}+1)/2}\left(\frac{SSR}{SSR_0}\right)^{-\left(\frac{n+\nu+2}{2}\right)}\frac{SSR^{(m\mathcal{S}+2)/2}}{SSR_0^{(m+2)/2}}\left(\frac{|W_0'W_0|}{|W'W|}\right)^{\frac{1}{2}}$$

$$\times\exp\left(\frac{-n\tau^2}{2}\left(\frac{1}{SSR}-\frac{1}{SSR_0}\right)\right)\left(\frac{2}{n}\right)^{\frac{m(\mathcal{S}-1)}{2}}\left(\frac{|\boldsymbol{\Sigma}|}{b}\right)^{\frac{-1}{2}}\left(\frac{\Gamma(\frac{a}{2})}{\Gamma_{\mathcal{S}}(\frac{a}{2})}\right)^{m}$$

$$\times\frac{\sum_{i=0}^{m}\sum_{l=1}^{\binom{m}{i}}I_{\delta_{il}}\boldsymbol{\pi}^i(1-\boldsymbol{\pi})^{m-i}2^{-\mathcal{S}i/2}(\Gamma_{\mathcal{S}}(\frac{a}{2}))^i(\Gamma_{\mathcal{S}}(\frac{a+1}{2}))^{m-i}}{\sum_{i=0}^{m}\sum_{l=1}^{\binom{m}{i}}I_{\delta_{il}}\boldsymbol{\pi}^i(1-\boldsymbol{\pi})^{m-i}2^{-i/2}(\Gamma(\frac{a}{2}))^i(\Gamma(\frac{a+1}{2}))^{m-i}}$$

$$\times\frac{\prod_{k:g_k\in\delta_{l0}^c}\frac{1}{|1+\hat{\boldsymbol{g}}_k\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{g}}_k'|^{\left(\frac{a+1}{2}\right)}}}{\prod_{k:g_k\in\delta_{l0}^c}\frac{1}{|1+\hat{g}_{0k}^2b^{-1}|^{\left(\frac{a+1}{2}\right)}}}$$

Before presenting fractional Bayes factors, the following result comparing SSR_0 and SSR in the particular case of our models is presented and proven. This result will be used in the discussion section to help in the interpretation of Bayes factors and fractional Bayes factors.

Result 1. For the models considered in this study, the following inequality holds: $SSR_0\geq SSR$.

Proof.

Let $SSM_1=\boldsymbol{y}'H_W\boldsymbol{y}$ and $SSM_0=\boldsymbol{y}'H_{W_0}\boldsymbol{y}$. Thus, proving that $SSR_0\geq SSR$ is equivalent to prove that $SSM_1\geq SSM_0$. Let $C(W_0)$ be the column space of W_0 and $C(W)$ the column space of W. Now, it is proven that $C(W_0)\preccurlyeq C(W)$, where the notation "$C(W_0)\preccurlyeq C(W)$" means that $C(W_0)$ is a subspace of $C(W)$. Let $z\in C(W_0)$, then $\exists\ \boldsymbol{a}\in\mathbb{R}^m$ such that $z=W_0\boldsymbol{a}$, that is,

$$z=\begin{bmatrix}W_1\boldsymbol{a}\\ \vdots\\ W_{\mathcal{S}}\boldsymbol{a}\end{bmatrix}.$$

Similarly, let $\boldsymbol{w}\in C(W)$, then $\exists\ \boldsymbol{b}\in\mathbb{R}^{m\mathcal{S}}$ such that $\boldsymbol{w}=W\boldsymbol{b}$. Without loss of generality vector $\boldsymbol{b}$ can be partitioned as $\boldsymbol{b}=(\boldsymbol{b}_1,\ldots,\boldsymbol{b}_{\mathcal{S}})$ where $\boldsymbol{b}_l\in\mathbb{R}^m\forall\ l=1,2,\ldots,\mathcal{S}$. Then $\boldsymbol{w}$ is of the form

$$\boldsymbol{w}=\begin{bmatrix}W_1\boldsymbol{b}_1\\ \vdots\\ W_{\mathcal{S}}\boldsymbol{b}_{\mathcal{S}}\end{bmatrix}.$$

In particular, if $\boldsymbol{b}_l=\boldsymbol{a}\forall\ l=1,2,\ldots,\mathcal{S}$, it follows that z also has the form of an element of $C(W)$, that is, $z\in C(W)$. Clearly, $\boldsymbol{w}$ cannot be written as a linear combination of the columns of W_0; therefore, $C(W_0)\preccurlyeq C(W)$. Applying theorem B.47 of Christensen (2011), it follows that $H_W-H_{W_0}$ is an orthogonal projection. By properties of orthogonal projections (Harville, 2000) it follows that $H_W-H_{W_0}$ is a semi-positive definite matrix, and consequently $\boldsymbol{y}'(H_W-H_{W_0})\boldsymbol{y}\geq 0\Longleftrightarrow\boldsymbol{y}'H_W\boldsymbol{y}\geq\boldsymbol{y}'H_{W_0}\boldsymbol{y}$.

2.3.1.2. Fractional Bayes factors. O'Hagan (1994, 1995) proposed a non-subjective Bayes factor known as fractional Bayes factor which uses a fraction c of the likelihood resulting in a "partial" Bayes factor having the following form:

$$FBF_{10}=BF_{10}\frac{\int_{\Theta_0}\pi_0(\boldsymbol{\theta}_0)(f_0(\boldsymbol{y}|\boldsymbol{\theta}_0))^c d\boldsymbol{\theta}_0}{\int_{\Theta_1}\pi_1(\boldsymbol{\theta}_1)(f_1(\boldsymbol{y}|\boldsymbol{\theta}_1))^c d\boldsymbol{\theta}_1}.$$

Thus, given W, the fractional Bayes factor for the homogeneous marker effect covariance matrix model with homoscedastic residuals and Gaussian prior for $\boldsymbol{g}$ has the form:

$$FBF_{10GW}=BF_{10GW}\frac{f_c(\boldsymbol{y}|W_0,M_{0G})}{f_c(\boldsymbol{y}|W,M_{1G})}$$

$$=BF_{10GW}\frac{\int_{\mathbb{R}_+}\pi(\sigma_g^2)\left(\int_{\mathbb{R}^m}\int_{\mathbb{R}_+}(f_0(\boldsymbol{y}|\boldsymbol{g}_0,\sigma_e^2,W))^c\pi(\boldsymbol{g}_0|\sigma_g^2)\pi(\sigma_e^2)d\sigma_e^2d\boldsymbol{g}_0\right)d\sigma_g^2}{\int_{\mathcal{P}_{\mathcal{S}}^+}\pi(G^0)\left(\int_{\mathbb{R}^{m\mathcal{S}}}\int_{\mathbb{R}_+}(f_1(\boldsymbol{y}|\boldsymbol{g},\sigma^2,W))^c\pi(\boldsymbol{g}|G^0)\pi(\sigma^2)d\sigma^2d\boldsymbol{g}\right)dG^0}$$

Hence, $\ln(f_i(\boldsymbol{y}\boldsymbol{\theta}_i))^c, i=0,1$, and their corresponding Hessian matrices evaluated at the MLE have to be found in order to find the Laplace approximation to the integrals inside the brackets in the numerator and denominator of FBF_{10GW}. This is easily done because

$\ln(f_i(\boldsymbol{y}\boldsymbol{\theta}_i))^c = c\ln f_i(\boldsymbol{y}\boldsymbol{\theta}_i)$. The determinants of the negative Hessian matrices are now denoted by $\overline{D}_0$, $\overline{D}_1$ and they satisfy: $\overline{D}_0 = c^{m+1}D_0$ and $\overline{D}_1 = c^{mS+1}D_1$. The approximate FBF_{10GW} is denoted as $\overline{FBF}_{10GW}$.

Fractional Bayes factors derived in this study were $\overline{FBF}_{10SSW}$, $\overline{FBF}^*_{10SSW}$ and $\overline{FBF}^*_{10GHW}$. It turned out that $\overline{FBF}_{10GW} = \overline{FBF}^*_{10GW}$ because the components making $\overline{BF}_{10GW}$ different from $\overline{BF}^*_{10GW}$ cancelled when multiplying them by $\frac{f_c(\boldsymbol{y}\mid W_0, M_0)}{f_c(\boldsymbol{y}\mid W, M_1)}$ and $\frac{f_c(\boldsymbol{y}\mid W_0, M_0^*)}{f_c(\boldsymbol{y}\mid W, M_1^*)}$ respectively. For details on the derivation see Appendix B. Moreover, the same cancellation happened when deriving $\overline{FBF}_{10SSW}$ and $\overline{FBF}^*_{10SSW}$. The resulting expression was:

$$\overline{FBF}_{10GW} = \overline{FBF}^*_{10GW} = \overline{FBF}_{10SSW} = \overline{FBF}^*_{10SSW} := \overline{FBF}_{10W}$$

$$= c^{m(\mathcal{S}-1)/2}\left(\frac{SSR}{SSR_0}\right)^{\frac{n(c-1)}{2}}\frac{SSR^{(m\mathcal{S}+2)/2}}{SSR_0^{(m+2)/2}}.$$

Notice that in the case $m > n$ where W and W_0 are not of full column rank, this expression is invariant to the choice of the generalized inverses $(W'W)^-$ and $(W_0'W_0)^-$. This follows because of the uniqueness of the projection operator onto the column space of W, H_W (Harville, 2000), which implies that SSR and SSR_0 are invariant to the choice of the generalized inverses. The approximate fractional Bayes factor $\overline{FBF}^*_{10GHW}$ was equal to 1 (see Appendix B for details). Thus, it does not provide information for comparing the corresponding models.

Based on the fact that the $\overline{FBF}_{10W}$ is invariant to the choice of generalized inverses of $W'W$ and $W_0'W_0$ when $m > n$, a brief discussion about the possible use of this criterion in the non-full rank case is provided in Appendix C. The issue is that the derivation that led to the fractional Bayes factor in the full rank case cannot be applied to the non-full rank case due to the fact that $|W'W| = |W_0'W_0| = 0$ and $(W'W)^{-1}$ and $(W_0'W_0)^{-1}$ do not exist. Although expressions involving these quantities cancel later on, it is clear that the derivations presented in Appendix B do not justify using $\overline{FBF}_{10W}$ in the non-full rank case.

These Bayes factors are useful for carrying out the conventional model selection conditioned on W, that is, conditioned on the observed genotypes. When part of W is not observed, the joint distribution of $\boldsymbol{y}$ and W^N given W^o can be obtained and then summing over the set $\mathcal{G}^N$ yields Bayes factors and fractional Bayes factors conditioned on W^o. Recall that $BF_{10W} = \frac{f(\boldsymbol{y}\mid W, M_1)}{f(\boldsymbol{y}\mid W_0, M_0)}$, to find $BF_{10W^o} = \frac{f(\boldsymbol{y}|W^o M_1)}{f\left(\boldsymbol{y}|W_0^o M_0\right)}$ the following computation has to be performed:

$$f(\boldsymbol{y}|W^o, M_1) = \sum_{\mathcal{G}^N}\pi(\boldsymbol{y}, W^N|W^o, M_1)$$

$$= \sum_{\mathcal{G}^N} f(\boldsymbol{y}|W, M_1)\pi(W^N|W^o, M_1)$$

$$= \sum_{\mathcal{G}^N}\left\{f(\boldsymbol{y}|W, M_1)\int_{\Omega}\pi(W^N|W^\sigma, P^*)\pi(P^*)dP^*\right\}.$$

For $\boldsymbol{r}$ known:

$$\pi(W^N|W^\sigma, M_1) = \prod_{j=1}^{m}\int_{\Omega_j^{\boldsymbol{r}}}\pi\left(\boldsymbol{w}_j^N|\boldsymbol{w}_j^\sigma, \boldsymbol{p}_j\right)\pi(\boldsymbol{p}_j|\boldsymbol{r})d\boldsymbol{p}_j$$

where $\Omega_j^{\boldsymbol{r}} := \{\boldsymbol{p}_j \in \mathbb{R}^{\mathcal{S}} | 0 < p_{lj} \le r_l \quad \forall \quad l, \quad \sum_{l=1}^{\mathcal{S}} r_l = 1\}$ and $\Omega = \Omega_1^{\boldsymbol{r}} \times \cdots \times \Omega_m^{\boldsymbol{r}}$ is the support of the distribution of P. For all j, the pmf $\pi\left(\boldsymbol{w}_j^N|\boldsymbol{w}_j^o, \boldsymbol{p}_j\right)$ can be found using Bayes theorem as $\pi\left(\boldsymbol{w}_j^N|\boldsymbol{w}_j^o, \boldsymbol{p}_j\right) = \pi(\boldsymbol{w}_j|\boldsymbol{p}_j)/\pi(\boldsymbol{w}_j^o|\boldsymbol{p}_j)$, but computing $\pi(\boldsymbol{w}_j^o|\boldsymbol{p}_j)$ requires $\sum_{\mathcal{G}^N}\pi(\boldsymbol{w}_j|\boldsymbol{p}_j)$ which can be unfeasible from the computational point of view. Alternatively, $\pi(W^N|W^o, P^*)$ can be derived from first principles by noticing that the dependence on W^o comes from the term where genotypes of individuals are conditioned on parental genotypes and then proceeding as in section 2.1.1 of part I. Using the expressions derived in Section 2.2.1 of part I and assuming $\boldsymbol{r}$ known:

$$\pi(W^N|W^o, M_1) \propto$$

$$2^{n_N^H}\prod_{j=1}^{m}\int_{\Omega_j^{\boldsymbol{r}}} p_{(\mathcal{S}+1)j}^{\alpha_{\mathcal{S}+1}-1}\prod_{l=1}^{\mathcal{S}}\left\{\frac{1}{2f_{lj_N}}p_{lj}^{n_{lN}^{Bj}+\alpha_l-1}(r_l - p_{lj})^{n_{lN}^{Aj}}\prod_{i'=f_{lj_N}+1}^{n_{lj_N}}\pi(w_{i'j}^l|w_{S_{i'j}}, w_{D_{i'j}})\right\}d\boldsymbol{p}_j$$

where f_{lj_N} is the number of founders with missing genotypes at locus j in subpopulation l, n_{lj_N} is the total number of individuals with missing genotypes at locus j in subpopulation l. Given that $\prod_{i'=f_{lj_N}+1}^{n_{lj_N}}\pi(w_{i'j}^l|w_{S_{i'j}}, w_{D_{i'j}})$ does not depend on P^*, the problem of finding $\pi(W^N|W^o, M_1)$ involves the evaluation of m integrals of the form:

$$\int_{\Omega_j^{\boldsymbol{r}}} p_{(\mathcal{S}+1)j}^{\alpha_{\mathcal{S}+1}-1}\prod_{l=1}^{\mathcal{S}}\left\{p_{lj}^{n_{lN}^{Bj}+\alpha_l-1}(r_l - p_{lj})^{n_{lN}^{Aj}}\right\}d\boldsymbol{p}_j,$$

this integral corresponds to the expectation of the function $\prod_{l=1}^{\mathcal{S}}(r_l - p_{lj})^{n_{lN}^{Aj}}$ of the random vector $\boldsymbol{p}_j$ taken over $\pi(\boldsymbol{p}_j|\boldsymbol{r})$. It does not have a closed form solution, but these integrals could be evaluated numerically in order to find a numerical approximation to $\pi(W^N|W^o, M_1)$. A similar situation occurs when $\boldsymbol{r}$ is not known, that is, integrals with no closed form solutions have to be evaluated in order to find $\pi(W^N|W^o, M_1)$.

Notice that matrices W and W_0 contain the same random variables but in different arrays. Consequently, W^N and W^o are the same in both cases and the analytic form of $\pi(W_0|\boldsymbol{p}_0)$, can be easily derived from $\pi(W|P^*)$ by setting $\mathcal{S}=1$ and taking into account that the prior posed over $\boldsymbol{p}_0$ is the product of m $Beta(\alpha, \beta)$ densities.

$$f(\boldsymbol{y}|W^o, M_0) = \sum_{\mathcal{G}^N} f(\boldsymbol{y}|W_0, M_0)\pi(W^N|W^o, M_0)$$

where

$$\pi(W^N|W^\sigma, M_0) = \int_{\Omega_0}\pi(W^N|W^\sigma, \boldsymbol{p}_0)\pi(\boldsymbol{p}_0)d\boldsymbol{p}_0, \Omega_0 = [0,1]\times[0,1]\times\cdots\times[0,1],$$

an $m-$dimensional unit hypercube.

$$\pi(W^N|W^o, \boldsymbol{p}_0) = 2^{n_N^H}\prod_{j=1}^{m}\left\{p_j^{n_N^{Bj}}(1-p_j)^{n_N^{Aj}}\prod_{i'=f_{Nj}+1}^{n_{Nj}}\pi(w_{i'j}|w_{S_{i'j}}, w_{D_{i'j}})\right\},$$

then, using the fact that $n_N^{Bj} + n_N^{Aj} = 2f_{Nj}$ (which is twice the total number of founders with missing genotypes at locus j), it follows that:

$$\pi(W^N|W^o, M_0) = \frac{2^{n_N^H}}{B(\alpha, \beta)^m}\prod_{j=1}^{m}\prod_{i'=f_{Nj}+1}^{n_{Nj}}\pi(w_{i'j}|w_{S_{i'}}, w_{D_{i'}}) \times \int_0^1 p_j^{n_N^{Bj}+\alpha-1}(1-p_j)^{n_N^{Aj}+\beta-1}dp_j$$

$$= \frac{2^{n_N^H}}{B(\alpha, \beta)^m}\prod_{j=1}^{m}\left\{\frac{\Gamma(n_N^{Bj}+\alpha)\Gamma(n_N^{Aj}+\beta)}{\Gamma(2f_{Nj}+\alpha+\beta)}\prod_{i'=f_{Nj}+1}^{n_{Nj}}\pi(w_{i'j}|w_{S_{i'j}}, w_{D_{i'j}})\right\},$$

where n_{Nj} is the total number of individuals with missing genotypes at locus j. Applying properties of the Gamma function (Casella and Berger, 2002; Kosmala, 2004) this can be reduced to (see Appendix A):

$$\pi(W^N|W^o, M_0)$$

$$=2^{n_N^H}\prod_{j=1}^{m}\left\{\frac{\prod_{k=1}^{n_N^{B_j}}(n_N^{B_j}-k+\alpha)\prod_{k=1}^{n_N^{A_j}}(n_N^{A_j}-k+\beta)}{\prod_{k=1}^{2f_{Nj}}(2f_{Nj}-k+\alpha+\beta)}\right.$$

$$\left.\prod_{i'=f_{Nj}+1}^{n_{Nj}}\pi(w_{i'j}|w_{S_{i'j}},w_{D_{i'j}})\right\}.$$

Therefore, in the case $\mathcal{S}=1$, there is an explicit expression for $\pi(W^N|W^o, M_0)$.

Notice that obtaining an approximation to the pdf $f(\boldsymbol{y}|W^o, M_1)$ involves computation of $SSR_{\hat{\boldsymbol{g}}}$, $\left|\boldsymbol{\Sigma}+\sum_{j=1}^{m}\hat{\boldsymbol{g}}_j\hat{\boldsymbol{g}}_j'\right|^{(a+m)/2}$ and $|W'W|$ for every possible value of W^N. Thus, this could be computationally unfeasible even for small or moderate sample sizes and chip densities.

Regarding interpretation of Bayes factors, their values can be classified according to the recommendations of Raftery (1996). This author proposed a scale to interpret Bayes factors based on a previous scale proposed by Jeffreys (1961); however, Raftery's scale is more granular and more conservative (Raftery, 1996). The scale is as follows: if $BF_{10}<1$, the evidence is negative (i.e., against model 1), values between 1 and 3 indicate that evidence for model 1 is not worth more than a bare mention, values between 3 and 20 indicate positive evidence in favor of model 1, values between 20 and 150 indicate strong evidence in favor of model 1 and values greater than 150 suggest very strong evidence for model 1.

2.3.2. Deviance information criterion

As in part I, another criterion used to compare models is the Deviance Information Criterion (DIC; Spiegelhalter et al., 2002). It combines a measure of goodness of fit based on the posterior distribution with a penalty for model complexity. In part I it was shown that for our family of models *DIC* can be written as the sum of two components, one computed from the $\boldsymbol{y}$ component of the likelihood and the other from the W component of the likelihood:

$$DIC=-2\log f(\boldsymbol{y}|W^o,\hat{W}_B^N,\hat{\boldsymbol{g}}_B,\hat{R}_B)+2p_{DIC-\boldsymbol{y}}-2\log f(W^o|\hat{W}_B^N,\hat{P}_B^*)+2p_{DIC-W}$$

$$:\quad =DIC_{\boldsymbol{y}}+DIC_W$$

where $p_{DIC-\boldsymbol{y}}=2(\log f(\boldsymbol{y}|W^o,\hat{W}_B^N,\hat{\boldsymbol{g}}_B,\hat{R}_B)-E_{W^N,\boldsymbol{g},R,P^*|\boldsymbol{y},W^o}[\log f(\boldsymbol{y}|W,\boldsymbol{g},R)])$ and $p_{DIC-W}=2(f(W^o|\hat{W}_B^N,\hat{P}_B^*)-E_{W^N,P^*|\boldsymbol{y},W^o}[f(W^o|W^N,P^*)])$.

2.4. Analysis of simulated data

With the aim of providing an example of the implementation of some of the proposed models and to compare their performance, the two small simulated datasets described in part I were used here as well. For the sake of completeness some minor details about the simulation are provided. After simulating a historical population using a forward-in-time approach, subpopulations were created using individuals pertaining to the historical population as founders. Each subpopulation had different selection criteria, selection pressures, and mating systems. Dataset 1 was comprised of three subpopulations with different number of generations, migration was allowed and the heritability of the trait was high. Dataset 2 consisted of two subpopulations with two generations each, migration was not allowed and the heritability of the trait was low (see Table 2 of companion paper for further details). These simulations were performed using the software QMSIm (Sargolzaei and Schenkel, 2013). For further details, see part I.

These datasets were used to carry out analyses using the following models. Spike and slab prior and heterogeneous marker effect covariance matrices with π_0=0.5,π_0=0.9 and π_0=0.2 and their null versions. All models assumed homoscedastic residuals. In the results and discussion sections, results from the models fitted to these datasets in part I will also be considered. Models fit in part I were Multivariate Gaussian prior and homogeneous marker effect covariance matrices, Multivariate Gaussian prior and heterogeneous marker effect covariance matrices, both with homoscedastic residuals. Not all models were used to analyze these data because of the following reasons. Firstly, taking into account that simulations did not consider heteroscedastic residuals, only models with homoscedastic residuals were fit. Secondly, some models have computational issues that make their implementation intractable. This is the case of models with a spike and slab prior over $\boldsymbol{g}$ with homogeneous marker effect covariance matrices. In these models, the full conditional distribution of the covariance matrix G^0 involves all the combinations of i out of m markers with null effects for $i=0,1,\ldots,m$; therefore, it is not easy to sample from $\pi(G^0|Else)$ due to the number of combinations being exponential in m. As shown in Section 2.2.2, for the model with heterogeneous marker effect covariance matrices, it is easy to sample from the full conditional distribution of the covariance matrix of each marker locus which makes its implementation possible.

Data were analyzed using the MCMC algorithm described in part I assuming that $\boldsymbol{r}=\left(\frac{1}{\mathcal{S}},\ldots,\frac{1}{\mathcal{S}}\right)$ and using the product of $\mathcal{S}$ independent uniform $\left(0,\frac{1}{\mathcal{S}}\right)$ distributions as proposal for $\pi(P|Else)$. The following criteria for model comparison were computed: approximate Bayes factors and fractional Bayes factors derived in Section 2.3.1, the squared correlation between predicted breeding values and phenotypes in the testing populations (predictive ability), squared correlations between true and predicted breeding values in the testing and training populations (accuracy) and DIC.

The hyper-parameter π_0 was assumed to be given. In practice, values close to 1 are used reflecting the belief that many of the SNP do not have an effect. Alternatively, this hyperparameter can be tuned or a prior can be posed over it in order to reflect uncertainty. Here, three values of this parameter were implemented in the analyses, 0.9, 0.5 and 0.2. This does not correspond to a tuning procedure; it was done only for illustrative purposes. The three values were chosen to reflect situations in which the prior belief is that a high proportion of marker loci do not have an effect (π_0=0.9), approximately half of them have an effect (π_0=0.5), and a high proportion of markers have an effect (π_0=0.2). In dataset 2, the full genotypes of three individuals (one founder from each subpopulation and a non-founder from subpopulation 1) were not included in the analysis in order to simulate the case of missing genotypes.

For each analysis, 20.000 iterations were run, considering the first 10.000 as burn-ins. In-house R scripts (R Core Team, 2015) were created to accommodate spike and slab priors and to compute Bayes factors and Fractional Bayes factors as well as DIC. Analyses were performed using the University of Florida's high performance computing cluster.

3. Results

3.1. Bayes factors

Using the expressions derived in Section 2.3.1, approximate Bayes factors and fractional Bayes factors were computed for dataset 1. Recall that $\overline{FBF}_{10GW}=\overline{FBF}^*_{10GW}=\overline{FBF}_{10SSW}=\overline{FBF}^*_{10SSW}$; therefore, the same expression permits the comparison of models M_{1G}, M^*_{1G}, M_{1SS} and M^*_{1SS} with their corresponding null models. Because of the same reason that makes the sampling from the full conditional distribution of G^0 under model M_{1SS} difficult, approximate Bayes factors for models with spike and slab priors were not computed. According to the Raftery's scale, $\overline{BF}_{10GW}$ and $\overline{BF}^*_{10GW}$ suggested very strong evidence in favor of all full models (they were greater than 150) in dataset 1. The same result was found when using the fractional Bayes factor which was computed with $c=0.5$.

In dataset 2, computation of Bayes factors was not possible because $m>n_1$. Furthermore, even though only three individuals were assumed

Table 1
$\boldsymbol{y}$ component and total DIC for dataset 1.

Model	DIC_y	Total DIC
M_{1G}	33702.55	4751373.55
M^*_{1G}	11599.05	4729270.05
$M^*_{1SS0.5}$	11604.09	4729275.09
$M^*_{1SS0.9}$	11648.94	4729319.94
$M^*_{1SS0.2}$	11437.05	4729108.05
M_{0G}	15396.32	6604501.32
M^*_{0G}	13008.42	6602113.42
$M^*_{0SS0.5}$	12502.17	6601607.17
$M^*_{0SS0.9}$	12625.29	6601730.29
$M^*_{0SS0.2}$	12137.88	6601242.88

to be non-genotyped and the number of markers was small, computation of the fractional Bayes factor was not performed due to its computational demands. All evidence provided by the approximate fractional Bayes factors computed using the posterior means of W^N (which could be seen as a sort of plug-in criteria) was against the full models, that is, all fractional Bayes factors were smaller than 1.

3.2. DIC, predictive ability and accuracies of predicted breeding values

In dataset 1, DIC_W is common to all full models and to all null models, i.e., there are only two values. It is due to the fact that there were no missing genotypes (see part I for details). The values were 4717671 for full models, and 6589105 for null models, that is, information coming from observed genotypes provided evidence in favor of the full models. It means that in this population, genotypic data provided support for the assumption of heterogeneous and correlated allelic frequencies when comparing it with the competing assumption that allelic frequencies are the same in all subpopulations.

Tables 1 and 2 contain DIC values for datasets 1 and 2 respectively, whereas Table 3 shows predictive abilities and accuracies for the two datasets. For Tables 1–3, the following is the meaning of abbreviations for the different models fitted to datasets 1 and 2: M_{1G}= full model with Multivariate Gaussian prior and homogeneous marker effect covariance matrices, M^*_{1G} = full model with Multivariate Gaussian prior and heterogeneous marker effect covariance matrices, $M^*_{1SS0.5}$ = full model with spike and slab prior, π_0=0.5 and heterogeneous marker effect covariance matrices, $M^*_{1SS0.9}$ = full model with spike and slab prior, π_0=0.9 and heterogeneous marker effect covariance matrices, $M^*_{1SS0.2}$ = full model with spike and slab prior, π_0=0.2 and heterogeneous marker effect covariance matrices. The remaining models with subindex 1 replaced by 0 correspond to null versions of the corresponding full models.

Therefore, according to the component of total DIC computed from the $\boldsymbol{y}$ component of the likelihood, except for the models with homogeneous marker effect covariance matrices (variances), full models should be preferred over their null versions in this dataset. When considering total DIC, all full models had a smaller DIC. Additionally, the model with the smallest DIC, and therefore the one to be preferred was model $M_{0SSH0.2}$ followed by model M_{1GH}. Notwithstanding, the DIC values for models M_{1GH}, $M_{1SSH0.5}$, $M_{1SSH0.9}$ and $M_{1SSH0.2}$ were close.

Table 2
$\boldsymbol{y}$ component, W component and total DIC for dataset 2.

Model	DIC_y	DIC_W	Total DIC
M_{1G}	1314.0	38367.4	39681.4
M^*_{1G}	1328.8	38356.4	39684.2
$M^*_{1SS0.5}$	1313.6	38394.9	39708.5
$M^*_{1SS0.9}$	1304.8	38382.7	39687.5
$M^*_{1SS0.2}$	1323.4	38373.8	39697.2
M_{0G}	1365.6	38180.3	39545.9
M^*_{0G}	1370.1	38179.0	39549.1
$M^*_{0SS0.5}$	1350.4	38173.4	39523.8
$M^*_{0SS0.9}$	1361.2	38195.8	39557.0
$M^*_{0SS0.2}$	1245.5	38178.4	39432.9

Table 3
Predictive abilities and accuracies in datasets 1 and 2.

Model	Predictive Ability		Accuracy in testing population		Accuracy in Training population	
	Dataset 1	Dataset 2	Dataset 1	Dataset 2	Dataset 1	Dataset 2
M_{1G}	0.29	0.019	0.27	0.04	0.32	0.17
M^*_{1G}	0.76	0.016	0.83	0.03	0.94	0.21
$M^*_{1SS0.5}$	0.81	0.017	0.88	0.04	0.92	0.19
$M^*_{1SS0.9}$	0.81	0.018	0.88	0.04	0.90	0.14
$M^*_{1SS0.2}$	0.79	0.016	0.86	0.03	0.94	0.20
M_{0G}	0.53	0.004	0.50	0.07	0.55	0.24
M^*_{0G}	0.83	0.013	0.88	0.05	0.88	0.23
$M^*_{0SS0.5}$	0.72	0.003	0.77	0.06	0.86	0.24
$M^*_{0SS0.9}$	0.69	0.008	0.76	0.05	0.85	0.20
$M^*_{0SS0.2}$	0.72	0.009	0.79	0.05	0.79	0.24

In this dataset the two components of the DIC values and therefore DIC values were similar for all models. The $\boldsymbol{y}$ components of DIC were smaller for the full models except for the model with spike and slab prior for $\boldsymbol{g}$ and π_0=0.2. Conversely, the W components were smaller for null models as well as total DIC values.

According to the behavior of predictive abilities in dataset 1, the performance of the different models was similar except for M_{1G}. The model with the best predictive ability was model M^*_{0G} while model M_{1G} had the worst. The accuracies in testing dataset 1 showed a pattern similar to that followed by predictive abilities. The performance of the models was similar except for model M_{1G} which made the poorest job when predicting breeding values and model M_{0G} which had the worst performance of all null models. The highest accuracies of predicted breeding values in testing population 1 were observed for models $M^*_{1SS0.5}$, $M^*_{1SS0.2}$,and M^*_{0G}. Finally, the accuracies of predicted breeding values in the training population showed the same behavior than the other measures, a poorer performance for models with homogeneous covariance matrix (or variance for null models) of marker effects with model M_{1G} having the smallest accuracy. Models with the highest accuracies were M^*_{1G} and $M^*_{1SS0.2}$.

For dataset 2, predictive abilities and accuracies in the testing sets were very low. Accuracies in training set were slightly larger. All these measures based on squared correlations did not show marked differences between models. Full models had higher predictive abilities and smaller accuracies in testing and training sets.

4. Discussion

4.1. General features of the models

The set of hierarchical Bayesian linear regression models for simultaneous genome-wide prediction in several subpopulations accounting for randomness of genotypes developed in part I was extended by incorporating spike and slab priors. The slab components of the conditional priors for marker effects were $\mathcal{S}$-variate Gaussian distributions considering homogeneous or heterogeneous covariance matrices (or variances) and the spike component was multivariate mass at zero for full models and univariate mass at zero for null models. Then, in order to provide general criteria for comparison of the proposed models with some null versions of them, approximate

Bayes factors and fractional Bayes factors were derived under the assumption that $n_l \geq m \ \forall \ l = 1,2,\ldots,\mathcal{S}$ and the possible use of fractional Bayes factors for the case $m > n$ was briefly discussed. These Bayes factors and fractional Bayes factors were approximations because some of the multiple integrals required to find the marginal distribution of data given a model were approximated via the Laplace method.

Spike and slab priors assign positive mass at zero; therefore, models considering this class of priors can be used for variable selection and they induce a stronger shrinkage towards zero (Gianola, 2013; Xu and Ghosh, 2015). Our spike and slab models can perform variable selection at the marker level, that is, it is assumed that either a given marker has effects in all subpopulations or it does not have effect in any subpopulation. In statistics, this is known as sparsity at the group level (Xu and Ghosh, 2015). Xu and Ghosh (2015) reparametrized the coefficients of the multiple linear regression as the product of a positive diagonal matrix and a vector, i.e., $g_j := V_g b_j, j = 1,2,\ldots,m$. Then, they posed independent univariate spike and slab priors for the elements of the positive diagonal matrix and independent multivariate spike and slab priors for b_j. This strategy permits to induce two kinds of sparsity, at group level and within group. Thus, an extension of our models that would induce sparsity at the group (i.e., marker) and within group levels would be to consider conditional priors similar those developed in Xu and Ghosh (2015). Therefore, a given marker would have positive probability of having null effects only in a proper subset of subpopulations.

Uncertainty on the hyper-parameter π_0 can be accounted for by posing a prior over it. A usual choice is a Beta distribution or its special case the Uniform(0,1). Implementation of this approach in the models presented here is straightforward. It implies adding one more level in the hierarchy. In this case, the question arising is the impact of this on inferences. Using the Kullback-Leibler divergence, Lehmann and Casella (1998, Theorem 5.7) provide theoretical justification for the idea that parameters that are in lower levels of the hierarchy have a smaller impact on inference. Notwithstanding, this does not mean that the impact of this extra level in the hierarchy is negligible and therefore, if the prior knowledge about π_0 is poor or null it may be worth to account for uncertainty. As mentioned before, alternatively this parameter can be tuned.

Regarding approximate Bayes factors and fractional Bayes factor, those derived here were obtained via Laplace approximation which has an error of order $O(n^{-1})$ (Ghosh et al., 2006). This means that the error of approximation is bounded from above by a constant times n^{-1}. There is a refinement based on the Laplace method that allows obtaining an approximation with error of order $O(n^{-2})$ when $q(\boldsymbol{\theta})$ is a positive function (Tierney and Kadane, 1986), which is always satisfied in the context of this study (see Section 2.3.1). This refinement could be implemented to obtain more accurate approximations of Bayes factors and fractional Bayes factors.

Other authors, e.g., Raftery (1996) and Lewis and Raftery (1997) have also used the Laplace method or modifications of it (DiCiccio et al., 1997) to derive approximate Bayes factors. The following comments regarding the algebraic expressions of Bayes factors and fractional Bayes factors are made for a given dataset, that is, given $\boldsymbol{y}$,n, m and W^o. It is well known that for nested models (i.e., the null model corresponds to the full model with some parameters set to zero) $SSR_0 > SSR$ (Searle, 1971). In this case the models are not nested; therefore, this standard result cannot be used. However, Result 1 establishes the relationship between SSR_0 and SSR for our models.

Thus, by Result 1, the following component of the algebraic expression for BF_{10GW} is always greater or equal than 1: $\left(\frac{SSR}{SSR_0}\right)^{-\left(\frac{n+\nu+2}{2}\right)}$ and as a consequence it never provides evidence against model 1. Conversely, for $n \geq 2$ the following component is always smaller or equal than 1, that is, it never provides evidence in favor of model 1: $\exp\left(\frac{-n\tau^2}{2}\left(\frac{1}{SSR}-\frac{1}{SSR_0}\right)\right)\left(\frac{2}{n}\right)^{\frac{m(\mathcal{S}-1)}{2}}$. Of course, the strength of the evidence in favor or against model 1 (when $SSR_0 > SSR$) depends on the observed data. Both expressions depend on the data and the hyper-parameters assigned to the residual variance. On the other hand, the following expression

$$\left(\frac{|W_0{}'W_0|}{|W'W|}\right)^{\frac{1}{2}}=\left(\frac{|W_1{}'W_1+W_2{}'W_2+\cdots+W_{\mathcal{S}}{}'W_{\mathcal{S}}|}{|W_1{}'W_1||W_2{}'W_2|\cdots|W_{\mathcal{S}}{}'W_{\mathcal{S}}|}\right)^{\frac{1}{2}},$$

depends only on the data. However, there are no general results establishing the relationship between the determinants inside the parenthesis and this is why it cannot be established if this component is always smaller or greater than 1. Of course, these determinants are always positive because of the assumption that all submatrices $W_1,\ldots,W_{\mathcal{S}}$ are of full column rank. Thus, if this component favors model 1 or not depends on each dataset. The following component depends on both, the priors and the data:

$$\frac{SSR^{(m\mathcal{S}+2)/2}}{SSR_0^{(m+2)/2}}\left(\prod_{l=2}^{\mathcal{S}}\frac{\Gamma\left(\frac{a+m+1-l}{2}\right)}{\Gamma\left(\frac{a+1-l}{2}\right)}\right)\left(\frac{\left|\boldsymbol{\Sigma}+\sum_{j=1}^{m}\hat{\boldsymbol{g}}_j\hat{\boldsymbol{g}}_j'\right|}{b+\sum_{j=1}^{m}\hat{g}_{0j}^2}\right)^{-\left(\frac{a+m}{2}\right)}.$$

The relative value of this component with respect to 1 cannot be established. Thus, as with the previous component, its contribution to the evidence in favor or against model 1 varies with each dataset. A similar situation occurs with BF^*_{10GW} and BF^*_{10GWH}, while for BF_{10SSW} and BF^*_{10SSW} there are new terms induced by the spike and slab priors posed over $\boldsymbol{g}$ and $\boldsymbol{g}_0$ whose relative value with respect to 1 depends on the observed data. However, the following statement can be made for the term involving gamma functions. In its positive domain, the Gamma function has a minimum point at approximate coordinates (1.461,0.885) (Kosmala, 2004), this implies that after 1.461 the function is increasing. Furthermore as $x \downarrow 0, \Gamma(x) \to \infty$. Note that for $l = 2,3,\ldots,\mathcal{S}$ and $a > \mathcal{S}-1$ (recall that the inverse Wishart distribution requires this condition) $\frac{a+m+1-l}{2} > \frac{m}{2}$ and $\frac{a+1-l}{2} > 0$. Therefore, given that in genome-wide prediction m has order of magnitude of at least 10^2, for values of a such that $(a+1-\mathcal{S})/2 > 1.461$ this term is always greater than 1.

Regarding fractional Bayes factors, as mentioned before,

$$FBF_{10GW}=FBF^*_{10GW}=FBF_{10SSW}=FBF^*_{10SSW}$$

$$=c^{m(\mathcal{S}-1)/2}\left(\frac{SSR}{SSR_0}\right)^{\frac{n(c-1)}{2}}\frac{SSR^{(m\mathcal{S}+2)/2}}{SSR_0^{(m+2)/2}}$$

due to cancellation of terms making approximate Bayes factors different. Recall that $c \in (0,1)$. As $c \uparrow 1$ and m and n remain constant the fractional Bayes factor approaches $\frac{SSR^{(m\mathcal{S}+2)/2}}{SSR_0^{(m+2)/2}}$. For $c \in (0,1)$ the exponent $\frac{n(c-1)}{2}$ is always negative and therefore $\left(\frac{SSR}{SSR_0}\right)^{\frac{n(c-1)}{2}}$ never provides evidence against model 1. On the contrary, $c^{m(\mathcal{S}-1)/2}$ provides evidence against model 1; however, as noted before, given m and $\mathcal{S}$, as $c \uparrow 1$ the evidence provided by this component is negligible because the whole expression approaches 1.

Some recommendations to choose the value of c are given in O'Hagan (1994) and Ghosh et al. (2006). Finally, the behavior of $\frac{SSR^{(m\mathcal{S}+2)/2}}{SSR_0^{(m+2)/2}}$ depends on the magnitude of the difference between SSR and SSR_0 and the number of subpopulations.

An important aspect of these approximations is that they require $n_l \geq m \ \forall \ l = 1,2,\ldots,\mathcal{S}$. As discussed in Section 2.3.1, the fast growth in the number of genotyped individuals may make this assumption possible for SNP chips of moderate size (i.e., 50 to 100k). However, the availability of denser chips and full sequences implies that m also

grows. On one hand, it is said that the higher the number of SNP the better the accuracy of genome-wide predictions because more LD between markers and QTL is "captured". On the other hand, some studies with real data such as Vázquez et al. (2010) in Holstein cattle and de los Campos et al. (2013) in humans have found that using subsets of SNP yields reasonable accuracy of genome-wide predictions. Moreover, the curve relating accuracy to marker density has been reported to reach a plateau for traits as height in humans (Vázquez et al., 2012) which suggests that in some cases not too much accuracy is lost when selecting subsets of SNP using some criteria.

Finally, the ability of our models to include non-genotyped individuals allows having a larger n, which combined with the factors mentioned before, increases the likelihood of having situations with $n_l \geq m \ \forall \ l = 1,2,\ldots,\mathcal{S}$. The approximate fractional Bayes factor $\overline{FBF}_{10W}$ could be used for the case $m > n$ but there is no formal mathematical justification for it. A brief discussion with an outline of the steps required to justify its use in such case is provided in Appendix C. Thus, the use of this expression for model comparison in the non-full rank case has to be seen as an *ad hoc* approach because there is no formal proof of its validity yet. Therefore, the question if the approximate fractional Bayes factor derived here is also valid for the non-full rank case remains to be formally answered. Thus, refuting this result or establishing a rigorous proof of it is an open problem.

4.2. Simulation results

Our small simulations correspond to two populations comprised by three and two subpopulations respectively. One trait per population was simulated. In both cases subpopulations had different mating designs, selection criteria, selection pressures and heritabilities. However, these populations display two contrasting scenarios. The first one (dataset 1) corresponded to a population comprised of three subpopulations that diverged by several generations, heritabilities were high, migration was allowed, the number of individuals in each subpopulation was larger than the number of SNP and there were no missing genotypes. Conversely, the second scenario (dataset 2) considered a population comprised by two subpopulations that diverged by only two generations, trait heritabilities were low, there was no migration, the number of individuals was smaller than the number of SNP in one subpopulation (hence the model was not of full rank) and there were missing genotypes.

In dataset 1, predictive ability did not suggest a superior predictive capability of full models, that is, models accounting for potential heterogeneity induced by the existence of subpopulations. As shown in Table 3, its values were very similar across models (except for the model with a homogeneous covariance matrix of marker effects which had considerably lower predictive ability). In this dataset, the number of marker loci considered in the analysis was equal to the number of QTL; therefore, it could be expected that the smallest value of π_0 had the best performance. The different squared correlations between predicted and observed values yielded similar results for the three values of π_0 used here with a slightly better performance for the model with the smallest value of π_0. While this set of correlations did not provide conclusive evidence in favor of the full models, the DIC, Bayes factors and fractional Bayes factors favored the full models.

Due to the low heritabilities in the two subpopulations forming dataset 2, predictive ability and accuracies were very low (Table 3). In this dataset full models had slightly higher predictive abilities than their null versions. Conversely, accuracies of predicted breeding values in training and validation datasets suggested a tiny superiority of null models. Total DIC and DIC_W provided evidence in favor of null models, but differences were not substantial. In addition, the "plug-in" fractional Bayes factors also gave evidence in favor of null models. As in part I, the performance of the fitted models was more similar in dataset 2 than in dataset 1.

A broad observation is that when combining the results obtained here with those obtained in the companion paper, the overall behavior observed in part I was kept. In general, what was observed in these small simulations was that under the biological scenario simulated in dataset 1, full models tended to have better performance, whereas in the setting simulated in dataset 2, null models tended to outperform full models. In all cases differences were small (except for models M_{1G} and M_{0G} in dataset 1). Therefore, after including the outputs of the spike and slab models, our results are still in agreement with those found by Olson et al. (2012), Makgahlela et al. (2013), de los Campos et al. (2015) and Lehermeir et al. (2015).

5. Conclusions

This study enlarges the family of hierarchical Bayesian models for across population genome-wide prediction accounting for randomness of genotypes derived in the companion paper (part I) by considering the so called spike and slab priors (multivariate and univariate) for marker allele substitution effects. This class of priors allows a stronger shrinkage towards zero and variable selection at group level. This development concedes even more flexibility to our family of models because the user will have more modeling options that permit to cope with a wider spectrum of biological scenarios. For example, for traits controlled by genes with major effects or controlled by a small number of genes, using spike and slab priors is theoretically advantageous.

The approximate Bayes factors and fractional Bayes factors derived here can be used to complement other criteria such as measures of accuracy of predicted breeding values and correlations between predicted breeding values and phenotypes when comparing models. These criteria were derived under the assumption of a full rank model which is currently satisfied in certain populations and we believe that it will become an increasingly more frequent situation as more individuals are genotyped. The invariance of our approximate fractional Bayes factor to the choice of the generalized inverses of $W'W$ and $W_0'W_0$ seems promising because it allows the use of this criterion in the non-full rank case. However, a formal justification or rejection of this criterion remains an open problem. For now, this criterion might be used *ad hoc*, keeping always in mind the risks that it implies.

In addition to all the possible extensions and refinements of our models discussed in the companion paper, the modification of the spike and slab priors presented here to allow sparsity within group (marker) is another aspect that opens a path for further research.

Author contributions

C.A. Martínez developed modeling strategies, carried out the derivations, wrote the R scripts, designed and made the simulations and wrote the paper. K. Khare advised modeling strategies, reviewed, corrected and discussed the derivations and the statistical aspects of the paper. A. Banerjee advised modeling strategies, reviewed, corrected and discussed the derivations and the statistical aspects of the paper. M.A. Elzo designed the simulation, reviewed, corrected and discussed the genetic aspects of the paper.

Acknowledgments

Authors acknowledge Dr. Malay Ghosh from the Department of Statistics of the University of Florida for useful comments and discussions, and for pointing out relevant references. C. A. Martínez also thanks Fulbright Colombia and "Departamento Adiministrativo de Ciencia, Tecnología e Innovación" COLCIENCIAS for supporting his PhD and Master programs at the University of Florida through a scholarship, and Bibiana Coy for her love, support and constant encouragement.

Appendices. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jtbi.2016.12.022.

References

Casella, G., Berger, R., 2002. Statistical Inference 2nd ed. Duxbury, Pacific Grove, CA, USA.

CDCB Council on Dairy Cattle Breeding, 2016. Genetic evaluations-genotype counts. Retrieved January 10, 2016 from ⟨https://www.cdcb.us/Genotype/counts.html⟩.

Christensen, R., 2011. Plane Answers to Complex Questions the Theory of Linear Models 4th ed. Springer, New York, NY, USA.

de los Campos, G., Vázquez, A.I., Fernando, R., Klimentidis, Y.C., Sorensen, D., 2013. Prediction of complex human traits using the genomic best linear unbiased predictor. PLOS Genet. 9 (7), e1003608.

de los Campos, G., Veturi, Y., Vázquez, A.I., Lehermeier, C., Pérez-Rodríguez, P., 2015. Incorporating genetic heterogeneity in whole-genome regressions using interactions. J. Agric., Biol. Environ. Stat. 20 (4), 467–490.

Daetwyler, H.D., Pong-Wong, R., Villanueva, B., Woolliams, J.A., 2012. The impact of genetic architecture on genome-wide evaluation methods. Genetics 185, 1021–1031.

DiCiccio, T.J., Kass, R.E., Raftery, A., Wasserman, L., 1997. Computing Bayes factors by combining simulations and Asymptotic approximations. J. Am. Stat. Assoc. 92 (439), 903–915.

Ghosh, J.K., Delampady, M., Samanta, T., 2006. An Introduction to Bayesian Analysis Theory and Methods. Springer, New York, NY, USA.

Gianola, D., 2013. Priors in whole-genome regression: the Bayesian alphabet returns. Genetics 194, 573–596.

Gianola, D., Rosa, G., 2015. One hundred years of statistical developments in animal breeding. Annu. Rev. Anim. Biosci. 3, 19–56.

Harville, D., 2000. Matrix Algebra from a Statistician's Perspective. Springer, New York, NY, USA.

Heslot, N., Yang, H.P., Sorrells, M.E., Jannink, J.L., 2012. Genomic selection in plant breeding: a comparison of models. Crop Sci. 52, 146–160.

Jeffreys, H., 1961. Theory of Probability 3rd ed. Oxford University Press, Oxford, UK.

Kosmala, W.A.J., 2004. A Friendly Introduction to Analysis Single and Multivariate 2nd ed. Pearson, Upper Saddle River, NJ, USA.

Lavine, M., Schervish, M.T., 1999. Bayes factors: what they are and what they are not. Am. Stat. 53 (2), 119–122.

Lehermeir, C., Schon, C., de los Campos, G., 2015. Assessment of genetic heterogeneity in structured plant populations using multivariate whole-genome regression models. Genetics 201, 323–337.

Lehmann, E.L., Casella, G., 1998. Theory of Point Estimation. Springer, New York, NY, USA.

Lewis, S.M., Raftery, A.E., 1997. Estimating Bayes factors via posterior simulation with the Laplace-Metropolis estimator. J. Am. Stat. Assoc. 92 (438), 648–655.

Makgahlela, M.L., Mantysaari, E.A., Stranden, I., Koivula, M., Nielsen, U.S., Sillanpaa, M.J., Juga, J., 2013. Across breed multi-trait random regression genomic predictions in the Nordic Red dairy cattle. J. Anim. Breed. Genet. 130, 10–19.

Mitchell, T.J., Beauchamp, J.J., 1988. Bayesian variable selection in linear regression. J. Am. Stat. Assoc. 83 (404), 1023–1032.

O'Hagan, A., 1994. Bayesian InferenceKendall's Advanced Theory of Statistics 2B. Halsted Press, New York, NY, USA.

O'Hagan, A., 1995. Fractional Bayes factors for model comparison. *Journal of the Royal Statistical Society*. Ser. B 57 (1), 99–138.

Olson, K.M., VanRaden, P.M., Tooker, M.E., 2012. Multibreed genomic evaluations using purebred Holsteins, Jerseys, and Brown Swiss. J. Dairy Sci. 95, 5378–5383.

Core Team, R., 2015. R: A language And Environment for Statistical Computing. R foundation for statistical computing, Vienna, Austria, (URL)⟨https://www.R-project.org/⟩.

Raftery, A.E., 1996. Approximate Bayes factors and accounting for model uncertainty in generalised linear models. Biometrika 86 (2), 251–266.

Sargolzaei, M., Schenkel, F.S., 2013. QMSim User's Guide Version 1.10. Centre for Genetic Improvement of Livestock, Department of Animal and Poultry Science, University of Guelph, Guelph, Canada.

Searle, S.R., 1971. Linear Models. John Wiley & Sons, New York, NY, USA.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., van der Linde, A., 2002. Bayesian measures of model complexity and fit (with discussion). J. R. Stat. Soc. Ser. B 64, 583–639.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., van der Linde, A., 2014. The deviance information criterion: 12 years on. J. R. Stat. Soc. Ser. B 76, 485–493.

Tierney, L., Kadane, J.B., 1986. Accurate approximations for posterior moments. J. Am. Stat. Assoc. 81, 82–86.

van den Berg, S., Calus, M.P.L., Meuwissen, T.H.E., Wientjes, Y.C.J., 2015. Across population genomic prediction scenarios in which Bayesian variable selection outperforms GBLUP. BMC Genet. 16, 416.

Vázquez, A.I., Rosa, G.J.M., Weigle, K.A., de los Campos, G., Gianola, D., Allison, D.B., 2010. Predictive ability of subsets of single nucleotide polymorphisms with and without parent average in US Holsteins. J. Dairy Sci. 93 (12), 5942–5949.

Vázquez, A.I., de los Campos, G., Klimentidis, Y.C., Rosa, G.J.M., Gianola, D., Yi, N., Allison, D.B., 2012. A comprehensive genetic approach for improving prediction of skin cancer risk in humans. Genetics 192, 1493–1502.

Xu, X., Ghosh, M., 2015. Bayesian variable selection and estimation for group Lasso. Bayesian Anal. 10 (4), 909–936.