Multibreed genomic prediction using multitrait genomic residual maximum likelihood and multitask Bayesian variable selection

Calus, M. P. L., Goddard, M. E., Wientjes, Y. C. J., Bowman, P. J., & Hayes, B. J.

Multibreed genomic prediction using multitrait GREML and multitask Bayesian variable selection. ***By Calus et al. (Page 000)*****.**

So far, limited benefits have been observed from combining information on multiple breeds in genomic evaluations. We investigated a model that accumulates evidence for the presence of QTL across breeds, while computing SNP effects within breeds. This model was slightly outperformed by a simple pooling strategy where information on Holsteins and Jerseys was analyzed without considering the differences between breeds. The most likely explanation is that, in the case of larger QTL effects, which are the main drivers of genomic prediction across breeds, the pooling strategy's assumption that SNP effects are the same across breeds is indeed appropriate.

Multibreed genomic prediction using multitrait GREML and multitask Bayesian variable selection

M. P. L. Calus*[1], M. E. Goddard†,‡, Y. C. J. Wientjes*, P. J. Bowman‡,§, and B. J. Hayes§,#

[*]Wageningen University & Research, Animal Breeding and Genomics, P.O. Box 338, 6700 AH Wageningen, The Netherlands

†Faculty of Veterinary and Agricultural Science, University of Melbourne, Melbourne, Victoria 3010, Australia

‡Agriculture Research, Department of Economic Development, Jobs, Transport and Resources, Melbourne, Victoria, Australia

§School of Applied Systems Biology, La Trobe University, Bundoora, Victoria, Australia

#Queensland Alliance for Agriculture and Food Innovation, Centre for Animal Science, The University of Queensland, St. Lucia, Queensland, Australia

[1]Corresponding author: Mario Calus, mario.calus@wur.nl

Wageningen University & Research

Animal Breeding and Genomics

P.O. Box 338

6700 AH Wageningen

Phone: 0031-317-480501

ABSTRACT

Genomic prediction is applicable to individuals of different breeds. Empirical results to date, however, show limited benefits in using information on multiple breeds in the context of genomic prediction. We investigated a multitask Bayesian model, presented previously by others, implemented in a Bayesian stochastic search variable selection (BSSVS) model. This model allowed for evidence of QTL to be accumulated across breeds or for both QTL that segregate across breeds and breed-specific QTL. In both cases, SNP effects were estimated with information from a single breed. Other models considered were a single-trait and multitrait genomic residual maximum likelihood (GREML) model, with breeds considered as different traits, and a single-trait BSSVS model. All single-trait models were applied to each of the two breeds separately, and to the pooled data of both breeds. The data used included a training data set of 6,278 Holstein and 722 Jersey bulls, and 374 Jersey validation bulls. All animals had genotypes for 474,773 SNPs after editing, and phenotypes for milk, fat and protein yields. Using the same training data, BSSVS consistently outperformed GREML. The multitask BSSVS, however, did not outperform single-trait BSSVS, which used pooled Holstein and Jersey data for training. Thus, the rigorous assumption that the traits are the same in both breeds yielded a slightly better prediction than a model that had to estimate the correlation between the breeds from the data. Adding the Holstein data significantly increased the accuracy of the single-trait GREML and BSSVS in predicting the Jerseys for milk and protein, in line with estimated correlations between the breeds of 0.66 and 0.47 for milk and protein yields, while only the BSSVS model significantly improved the accuracy for fat yield with an estimated correlation between breeds of only 0.05. The relatively high genetic correlations for milk and protein yields, and the superiority of the pooling strategy, is likely the result of the observed admixture between both breeds in our data. The Bayesian model was able to detect several QTLs in Holsteins, which likely enabled it to outperform GREML. The inability of the multitask

Bayesian models to outperform a simple pooling strategy may be explained by the fact that the pooling strategy assumes equal effects in both breeds; furthermore, this assumption may be valid for moderate- to large-sized QTLs, which are important for multibreed genomic prediction.

Keywords: genomic prediction, multibreed, Bayesian variable selection

INTRODUCTION

One of the benefits of genomic prediction is that it can use information across groups of individuals, such as different livestock breeds, which are not connected through any recent pedigree links. Considering the hypothesis that genomic prediction relies on linkage disequilibrium (LD) between SNPs and QTLs (Meuwissen et al., 2001), the expectation was that genomic prediction across breeds would be possible if the SNP density was large enough. This expectation was supported by the supposition that genomic prediction across Holsteins and Jerseys would be possible if the number of SNPs was greater than 300,000 (De Roos et al., 2008). This, however, was based on simulations that assumed that the QTLs underlying the traits of interest are the same and have the same effects among different breeds.

Several empirical studies have shown that the accuracy of multibreed, compared to single-breed, genomic prediction is, at best, slightly higher, but often remains unchanged or is even slightly lower when breeds are distantly related (Erbe et al., 2012; Karoui et al., 2012; Olson et al., 2012; Lund et al., 2014; Zhou et al., 2014). In situations where breeds are closely related, increases in accuracy from multibreed genomic prediction are more easily obtained (Brøndum et al., 2011), especially if the initial training data of the predicted breed is small (Hozé et al., 2014). One possible explanation for the limited success of multibreed genomic prediction is that the genetic basis of traits has evolved, at least to a partially different extent,

in the breeds involved, while the genomic prediction model is not flexible enough to accommodate these differences. Differences in genetic backgrounds may be due, for instance, to only a partial overlap between loci affecting a trait across breeds, to interactions with the genetic background of the breed, and to differences in allele frequencies and LD patterns of loci, which do affect any traits in different breeds.

One proposed strategy to accommodate these differences between breeds is to use multitrait (MT) models, where trait-by-breed combinations are treated as different, but correlated, traits (Karoui et al., 2012; Olson et al., 2012; Huang et al., 2014; Zhou et al., 2014). All these studies applied an MT genomic best linear unbiased prediction (GBLUP) type of model. One important assumption underlying this model is that, across the genome, one single genetic correlation between breeds is considered, which assumes for each SNP, *a priori*, the same covariance structure between effects among different breeds. An alternative model, which has been proposed recently, is the so-called multitask Bayesian learning model for multibreed genomic prediction (Chen et al., 2014), which does not consider the same co-variance structure between breeds across the genome. This is effectively a Bayesian variable selection model, which uses the data on all breeds to decide whether or not a variable is selected into the model. In other words, this model accumulates evidence across breeds in order to determine whether or not a SNP is linked to a QTL. The SNP effects are subsequently estimated separately within each breed, using only phenotypic information on the breed itself. The implementation, as presented by Chen et al. (2014), however, does not explicitly accommodate SNPs linked to a breed-specific QTL. That said, there are indications that modeling both breed-specific and common QTLs is beneficial for multibreed genomic prediction (van den Berg et al., 2016b).

The objective of this paper, therefore, was to expand the multitask Bayesian learning model to allow for SNPs linked to a breed-specific QTL to obtain a large effect in one breed and a small effect in another, as well as to compare this to the originally proposed multitask

Bayesian learning model and several other models. These other models include single-trait and multitrait GBLUP-type models, and a single-trait Bayesian variable selection model. In all single-trait models, either phenotypes of only one of the breeds were used, or phenotypes of different breeds were pooled and analyzed simultaneously, as if the same trait was involved. Analyses were performed on a data set including Holsteins, with a moderate size of training set, and Jerseys, with a small-sized size of training set. Validation was, in all cases, only performed for the Jersey breed.

MATERIAL AND METHODS

Data

Phenotypic data. The data used in this study contained 7,994 Holstein and 1,378 Jersey bulls, which had both genotypes and phenotypes available. The Holstein bulls originated from Australia (35%), New Zealand (15%), and the Netherlands (50%), while the Jersey bulls originated from Australia (43%) and New Zealand (57%). The phenotypes were de-regressed proofs (DRPs) for milk, fat and protein yields, which were derived from international multiple trait across country evaluation (MACE) estimated breeding values (EBVs), as computed by Interbull and converted to the Australian scale. Each DRP had a weight computed as effective daughter equivalents (EDCs), which was derived from the corresponding MACE EBVs. Average reliabilities of the DRPs for the Holstein training bulls, as computed from the EDCs, were 0.81, 0.77 and 0.76, respectively, for milk, fat and protein yields. Average reliabilities of the DRPs for the Jersey training bulls were 0.84 for milk, fat and protein yields.

As the Jersey data set was considerably smaller than the Holstein data set, we only expected improvement in genomic prediction accuracy by adding information from the other breed for Jerseys, while validation of the models described in the next section was only

performed using Jersey validation bulls. The data were split into groups of training and validation bulls by assigning all bulls born prior to January 2004 to the training data set. This yielded an initial training data set containing 6,278 Holstein and 1,004 Jersey bulls, and a validation data set containing 374 Jersey bulls. Analysis of the data revealed that those 374 Jersey bulls had strong relationships with the Jersey training bulls, which likely reduced the potential impact of adding the Holstein training data to a considerable extent. To reduce the relationship with the training data set, close relatives of the 374 Jersey validation bulls were removed from the training data. This included 93 sires, 105 paternal half-sibs (i.e., sons of sires of validation bulls), 4 maternal half-sibs (i.e., sons of dams of validation bulls), 76 paternal and 53 maternal grandsons of sires of validation bulls, and 4 paternal and 16 maternal grandsons of dams of validation bulls. Some bulls appeared in more than one of these categories. For instance, 7 training bulls were both the sire of a validation bull and a paternal half-sib of another validation bull. Finally, the training data set contained 6,278 Holstein and 722 Jersey bulls.

Genotype data. All bulls were initially genotyped either with one of the two custom 50,000 chips used by CRV BV (all Dutch Holstein bulls), or the Illumina BovineSNP50 chip (all other bulls). Genotypes from these custom chips were imputed to the Illumina BovineSNP50, while ~10,000 or 17,000 SNPs were shared with the Illumina BovineSNP50 (Lund et al., 2011). After this imputation step, all bulls had genotypes for 43,990 SNPs. A total of 1,620 Holstein bulls and cows and 125 Jersey bulls were genotyped with the Illumina BovineHD array (~777,000 SNPs). This reference population was then used to impute HD genotypes for all bulls using Beagle version 3.0 (Browning and Browning, 2009). After quality control, in line with Erbe et al. (2012), and the removal of monomorphic SNPs, 600,640 SNPs remained. From any pair of SNPs that had an LD (i.e., r^2) value of 1, only one SNP was retained. This reduced the number of SNPs used for the analyses to 474,773.

156 ***Models***

157 ***Relationship-based models.*** The first model used is termed the pedigree based residual
158 maximum likelihood (PREML) model, since it computes variance components simultaneously
159 with EBVs based on pedigree information using residual maximum likelihood (REML). As the
160 PREML model only used phenotypic information on the Jersey training data set, this is only
161 applied as a single-trait (ST) model. The general PREML model was:

162 $$\mathbf{y} = \mathbf{1}\mu + \mathbf{Zu} + \mathbf{e}$$

163 where $\mathbf{y}$ is a vector with DRPs, $\mathbf{1}$ is a vector of ones, μ is the mean, $\mathbf{Z}$ is a matrix that links
164 records to animals, $\mathbf{u}$ is a vector with breeding values, and $\mathbf{e}$ is a vector with random residuals.
165 The assumed distributions of $\mathbf{u}$ and $\mathbf{e}$ were respectively $N(\mathbf{0}, \mathbf{A}\sigma_u^2)$ and $N(\mathbf{0}, \mathbf{D}\sigma_e^2)$, where $\mathbf{A}$ is
166 the pedigree-based additive genetic relationship matrix, σ_u^2 is the genetic variance, $\mathbf{D}$ is a
167 diagonal matrix containing $1/EDC_{DRP}$ on the diagonals, EDC_{DRP} are the EDCs of the DRP,
168 and σ_e^2 is the residual variance.

169 The second model used is termed GREML, since it is similar to a GBLUP model, but
170 computes variance components simultaneously with the genomic EBVs (GEBVs) using REML.
171 This model was applied both as an ST model and as a MT model. The general MT-GREML
172 model was:

173 $$\mathbf{y_k} = \mathbf{1}\mu_k + \mathbf{Z_k g_k} + \mathbf{e_k}$$

174 where $\mathbf{y_k}$ is a vector with DRPs for animals in breed *k*, *k* takes values of 1 for Holsteins and 2
175 for Jerseys in the MT model, $\mathbf{1}$ is a vector of ones, μ_k is the mean effect of breed *k* (effectively
176 the breed effect in our analyses), $\mathbf{Z_k}$ is a matrix that links records to animals, $\mathbf{g_k}$ is a vector with
177 GEBVs, and $\mathbf{e_k}$ is a vector with random residuals. The assumed distributions of $\begin{bmatrix}\mathbf{g_1}\\ \mathbf{g_2}\end{bmatrix}$ and $\begin{bmatrix}\mathbf{e_1}\\ \mathbf{e_2}\end{bmatrix}$
178 were respectively $N(\mathbf{0}, \mathbf{G_g} \otimes \mathbf{GRM})$ and $N(\mathbf{0}, \begin{bmatrix}\mathbf{D_1}\sigma_{e_1}^2 & 0\\ 0 & \mathbf{D_2}\sigma_{e_2}^2\end{bmatrix})$, where **GRM** is the genomic

relationship matrix, $\mathbf{G}_g$ is the genetic covariance matrix, $\mathbf{D_1}$ ($\mathbf{D_2}$) is a diagonal matrix containing $1/EDC_{DRP}$ on the diagonals for animals in the first (second) breed, EDC_{DRP} are the EDCs of the DRPs of animals in each of the breeds, and $\sigma^2_{e_1}$ ($\sigma^2_{e_2}$) is the residual variance for the first (second) breed. The **GRM** was computed following the description by Erbe et al., (2012), which concerns a multibreed development of the first method proposed by VanRaden, (2008):

$$\mathbf{GRM}_{\mathrm{c}} = \mathbf{WW}'/\mathrm{M}$$

where **W** is calculated as $\mathbf{W} = \mathbf{X} - 2\mathbf{p}$, **X** is a matrix containing genotypes coded as 0, 1 and 2, $\mathbf{p} = \alpha\mathbf{p}_{HOL} + (1-\alpha)\mathbf{p}_{JER}$, $\mathrm{M} = 2\sum_{j=1}^{m} p_j(1-p_j)$, and m is the total number of SNP loci used. Allele frequencies $\mathbf{p}_{HOL}$ and $\mathbf{p}_{JER}$ are averages within Holsteins and Jerseys, and $\alpha = \frac{F_{JER}}{F_{JER}+F_{HOL}}$, where F_{JER} and F_{HOL} are computed as defined below. Finally, the **GRM** was scaled; this means that the inbreeding is relative to the point before breed divergence, which then is the base of the **GRM**. Following Erbe et al. (2012),

$$\mathbf{GRM} = \mathbf{GRM}_{\mathrm{c}}(1-F) + 2F,$$

where F is the inbreeding relative to an *F1* base:

$$F = \frac{F_{JER}F_{HOL}}{F_{JER} + F_{HOL}}$$

$$F_{JER} = 1 - \frac{\sum_{j=1}^{m} 2p_{JER,j}\left(1-p_{JER,j}\right)}{\sum_{j=1}^{m}\left[p_{HOL,j}\left(1-p_{JER,j}\right) + p_{JER,j}\left(1-p_{HOL,j}\right)\right]},$$

and

$$F_{HOL} = 1 - \frac{\sum_{j=1}^{m} 2p_{HOL,j}\left(1-p_{HOL,j}\right)}{\sum_{j=1}^{m}\left[p_{HOL,j}\left(1-p_{JER,j}\right) + p_{JER,j}\left(1-p_{HOL,j}\right)\right]}.$$

Two different applications of the GREML model were used. The first application was an ST model (ST-GREML), which means that, in the above model description, the (co)variance matrices reduce to one scalar value. The ST-GREML used data from one breed or used data

pooled across breeds. When data were pooled across breeds, the assumed genetic correlation between breeds was 1. The second application of GREML was an MT model (MT-GREML), which analyzed the data simultaneously for both breeds by considering the trait to be different, but correlated, between the breeds. In this application, the genetic correlation between the breeds was explicitly estimated in the model, and this estimate is expected to be unbiased when using the **GRM** as outlined above (Wientjes et al., 2017). All PREML and GREML models were run using ASReml (Gilmour et al., 2014).

In compliance with the Bayesian variable selection model, which is explained in the next section, a polygenic effect based on pedigree was initially included in the GREML model. Due to the correlation with the effects modeled using the **GRM** matrix, this led to severe convergence issues; therefore, this pedigree-based polygenic effect was omitted from the GREML models in further analyses.

Bayesian stochastic search variable selection. The third model used is commonly termed Bayesian Stochastic Search Variable Selection (BSSVS) (Verbyla et al., 2009; Calus, 2014). The general BSSVS model used was:

$$\mathbf{y_k} = \mathbf{1}\mu_k + \mathbf{Z_k u_k} + \mathbf{X_k \alpha_k} + \mathbf{e_k}$$

where, for breed *k* (*k* taking values of 1, or 1 and 2 when both breeds are considered), $\mathbf{u_k}$ is a vector with additive genetic polygenic breeding values, $\mathbf{X_k}$ is a matrix with centered and scaled genotypes, and $\boldsymbol{\alpha_k}$ is a vector of allele substitution effects. The assumed distribution of $\mathbf{u_k}$ was $N(\mathbf{0}, \mathbf{A_k}\sigma^2_{u_k})$ in the ST model, while the distribution of $\begin{bmatrix}\mathbf{u_1}\\ \mathbf{u_2}\end{bmatrix}$ was $N(\mathbf{0}, \begin{bmatrix}\mathbf{A_1}\sigma^2_{u_1} & \mathbf{0}\\ \mathbf{0} & \mathbf{A_2}\sigma^2_{u_2}\end{bmatrix})$ when both breeds were considered simultaneously, where $\mathbf{A_k}$ is the pedigree numerator relationship matrix for breed *k*, and $\sigma^2_{u_k}$ is the polygenic variance of breed *k*. Three different implementations of the BSSVS model were used. The first was an ST model (ST-BSSVS) using data from one breed or using data pooled across both breeds. This ST model has also been

termed a single-task model (Chen et al., 2014). This implementation is described elsewhere in more detail (Calus, 2014). It assumes that a certain proportion of all loci, denoted as π, have a zero effect within each of the iterations of the Gibbs sampling scheme. The second application was a multitask (mt) model (mt-BSSVS), which used one π value across breeds, and thus accumulated evidence from across breeds to determine whether an SNP was linked to a QTL, while estimating SNP effects within a breed using only information from the breed itself. This mt-BSSVS implementation was similar to the mt model described by Chen et al. (2014). The third application was also an mt-BSSVS model, which allowed for fitting both SNPs linked to breed-specific QTLs, as well as SNPs linked to QTLs, which were the same across breeds. It also estimated SNP effects within a breed using only information from the breed itself. Both mt-BSSVS models are described in more detail below.

Prior densities. The likelihood of the BSSVS model being conditional on all unknowns is assumed to be normal:

$$p\left(y_{ik}\middle|\mu_k, \mathrm{u}_{ik}, \boldsymbol{\alpha_k}, \sigma^2_{e_k}\right) = N(y_{ik} - \mu_k - \mathrm{u}_{ik} - \mathbf{x}'_{ik}\boldsymbol{\alpha_k}, \sigma^2_{e_k})$$

where $\mathbf{x}_{ik}$ denotes the genotypes of animal *i* of breed *k*. The prior for μ_k was a constant. The residual variance $\sigma^2_{e_k}$ has a scaled inverse-χ^2 prior distribution of $p\left(\sigma^2_{e_k}\right) = \chi^{-2}(-2,0)$, which yields a flat prior.

The prior for α_{jk}, the allele substitution effect of locus *j* in breed *k*, depends on the variance $\sigma^2_{\alpha_k}$ and the indicator variable I_{jk} :

$$\alpha_{jk}|\pi_k, \sigma^2_\alpha = \begin{cases} \sim N\left(0, \frac{\sigma^2_{\alpha_k}}{100}\right) \; when \; I_{jk} = 0 \\ \sim N\left(0, \sigma^2_{\alpha_k}\right) when \; I_{jk} = 1 \end{cases}$$

The prior distribution for the indicator variable I_{jk} is:

$p\left(I_{jk}\right) = \text{Bernoulli}(1 - \pi_k)$,

where π is assigned a value of 0.999 and $\sigma^2_{\alpha_k}$ has a scaled inverse-χ^2 prior distribution of:

$$p\left(\sigma^2_{\alpha_k}\right) = \chi^{-2}\left(\nu_{\alpha_k}, S^2_{\alpha_k}\right)$$

where ν_{α_k} represents the degrees of freedom, set to 4.2, following (Meuwissen et al., 2001; Habier et al., 2011), while the scale parameter $S^2_{\alpha_k}$ is calculated as $S^2_{\alpha_k} = \frac{\tilde{\sigma}^2_{\alpha_k}(\nu_{\alpha_k}-2)}{\nu_{\alpha_k}}$, where $\tilde{\sigma}^2_{\alpha_k}$ is computed in line with (de los Campos et al., 2013):

$$\tilde{\sigma}^2_{\alpha_k} = \left(\frac{100}{100+\pi_k(1-100)}\right)\frac{\sigma^2_{\alpha_k}}{n}$$

where n is the number of loci. The value used for π and the ratio of the variance between the two distributions were the same as those we used in previous studies, where the BSSVS model was shown to be competitive, compared to other models (Daetwyler et al., 2013; Calus et al., 2014a; Calus et al., 2014b).

Conditional posterior densities. The conditional posterior density of α_{jk} is:

$$N\left(\hat{\alpha}_{jk};\ \frac{\omega_{jk}\hat{\sigma}^2_{e_k}}{\mathbf{x}'_{jk}\mathbf{D_k^{-1}}\mathbf{x}_{jk} + \lambda_{jk}}\right)$$

where $\hat{\alpha}_{jk}$ is the conditional mean of the allele substitution effect at locus *j* in breed *k*, computed as:

$$\hat{\alpha}_{jk} = \frac{\mathbf{x}'_{jk}\mathbf{D_k^{-1}}\mathbf{y}^*_{jk}}{\mathbf{x}'_{jk}\mathbf{D_k^{-1}}\mathbf{x}_{jk}+\lambda_{jk}},$$

where $\mathbf{y}^*_{jk}$ are conditional phenotypes for SNP *j*, defined as phenotypes corrected for estimated effects at all other SNP loci, $\lambda_{jk} = \frac{\omega_{jk}\hat{\sigma}^2_{e_k}}{\hat{\sigma}^2_{\alpha_k}}$, and

$$\omega_{jk} = 1 \quad \text{if} \quad I_{jk} = 1$$

$$\omega_{jk} = 100 \quad \text{if} \quad I_{jk} = 0$$

The conditional posterior density of $\sigma^2_{\alpha_k}$ is a scaled inverse-χ^2 distribution:

$$\sigma^2_{\alpha_k}|\alpha_k \sim \chi^{-2}(\nu_{\alpha_k} + n, S^2_{\alpha_k} + \boldsymbol{\omega}'_{\boldsymbol{k}}\hat{\boldsymbol{\alpha}}^{\mathbf{2}}_{\mathbf{k}})$$

where $\widehat{\boldsymbol{\alpha}}_{\mathbf{k}}^{\mathbf{2}}$ is a vector with squares of the current estimates of the allele substitution effects of all loci, that is, weighted by vector $\boldsymbol{\omega}_{\boldsymbol{k}}$, which contains values of 1 or 100 for all loci.

Finally, the conditional posterior distribution of the indicator variable I_{jk}, following the notation in (Jia and Jannink, 2012), was:

$$\Pr(I_{jk}=1)=\frac{\sum_k(\mathrm{f}(r_{jk}|I_{jk}=1)(1-\pi_k))}{\sum_k(\mathrm{f}(r_{jk}|I_{jk}=0)\pi_k+\mathrm{f}(r_{jk}|I_{jk}=1)(1-\pi_k))}$$

where $1-\pi_k$ (π_k) is the prior probability that $I_{jk}=1$ ($I_{jk}=0$), $r_{jk}=\sum_k(\mathbf{x}'_{jk}\mathbf{D}^{-\mathbf{1}}\mathbf{y}^{*}_{\boldsymbol{k}}+\mathbf{x}'_{jk}\mathbf{D}^{-\mathbf{1}}\mathbf{x}_{jk}\widehat{\alpha}_{jk})$, with $\mathbf{y}^{*}_{\boldsymbol{k}}$ representing the conditional phenotypes as defined previously, while $f(r_{jk}|I_{jk}=\delta)$, with δ as either 0 or 1, is proportional to $\frac{1}{\sqrt{v_k}}\mathrm{e}^{-\frac{r_{jk}^2}{2v_k}}$, with $v_k=\sum_k(\mathbf{x}'_{jk}\mathbf{D}_{\mathbf{k}}^{-\mathbf{1}}\mathbf{x}_{jk})^2\frac{\sigma^2_{\alpha_{jk}}}{\omega_{jk}}+\mathbf{x}'_{jk}\mathbf{D}_{\mathbf{k}}^{-\mathbf{1}}\mathbf{x}_{jk}\sigma^2_{e_k}$. It should be noted that v_k depends on I_{jk} through its dependence on ω_{jk}, i.e., if $I_{jk}=0$ ($I_{jk}=1$), then $\omega_{jk}=100$ ($\omega_{jk}=1$).

In all of the above, for the model with one π value across breeds, the "*k*" subscripts could effectively be removed from the parameters $I_{jk,}$ ω_{jk}, $\boldsymbol{\omega}_{\boldsymbol{k}}$, and π_k, given that they have the same values in both breeds. For the model with breed-specific π values, however, these parameters may be different for different breeds. The only other factor that changes in the above for this model is that for breed 1:

$$\Pr(I_{j1}=1)=\frac{\mathrm{f}(r_{jk}|I_{j1}=1)\pi_{11}+\mathrm{f}(r_{jk}|I_{j1}=1)\pi_{10}}{\mathrm{f}(r_{jk}|I_{j1}=0)\pi_{00}+\mathrm{f}(r_{jk}|I_{j1}=0)\pi_{01}+\mathrm{f}(r_{jk}|I_{j1}=1)\pi_{11}+\mathrm{f}(r_{jk}|I_{j1}=1)\pi_{10}}$$

$$=\frac{\mathrm{f}(r_{jk}|I_{j1}=1)(\pi_{11}+\pi_{10})}{\mathrm{f}(r_{jk}|I_{j1}=0)(\pi_{00}+\pi_{01})+\mathrm{f}(r_{jk}|I_{j1}=1)(\pi_{11}+\pi_{10})}$$

and equivalently for breed 2:

$$\Pr(I_{j2}=1)=\frac{\mathrm{f}(r_{jk}|I_{j2}=1)(\pi_{11}+\pi_{01})}{\mathrm{f}(r_{jk}|I_{j2}=0)(\pi_{00}+\pi_{10})+\mathrm{f}(r_{jk}|I_{j2}=1)(\pi_{11}+\pi_{01})}$$

where π_{11} is the prior probability that SNP *j* is linked to a QTL in both breeds, and π_{10} (π_{01}) is the prior probability that SNP *j* is linked to a QTL in breed 1 (2) but not in breed 2 (1), while π_{00} is the prior probability that SNP *j* is not linked to a QTL in both breeds. Here, we assumed that $\pi_{11} = \pi_{10} = \pi_{01} = 0.0005$, such that the total prior probability per breed was still 0.001, assuming that, for an SNP that is linked to a QTL in one breed, it is equally likely to be linked to a QTL in another breed or not. Given that all prior probabilities need to sum up to 1, $\pi_{00} = 0.9985$.

The log-likelihood of $I_{jk} = \delta$ is proportional to:

$$\mathrm{f}\left(r_{jk}\middle|I_{jk} = \delta\right)(pr) =$$

$$\sum_k \left(-\frac{1}{2}\log\left(1 + \mathbf{x}'_{jk}\mathbf{D^{-1}}\mathbf{x}_{jk}\frac{\widehat{\sigma}^2_{\alpha_{jk}}}{\omega_{jk}\widehat{\sigma}^2_{e_k}}\right) + \frac{1}{2}\frac{\left(\mathbf{x}'_{jk}\mathbf{D^{-1}}\mathbf{y}^*_{\boldsymbol{k}}\right)^2}{\sum_k \widehat{\sigma}^2_{e_k}\left(\frac{\omega_{jk}\widehat{\sigma}^2_{e_k}}{\widehat{\sigma}^2_{\alpha_{jk}}} + \mathbf{x}'_{jk}\mathbf{D^{-1}}\mathbf{x}_{jk}\right)} + \log(prior) \right)$$

where $\omega_{jk} = 100$ and $pr = \pi_k$ for $\delta = 0$, and $\omega_{jk} = 1$ and $pr = 1 - \pi_k$ for $\delta = 1$. It should be noted that the terms $\mathbf{x}'_{jk}\mathbf{D^{-1}}\mathbf{x}_{jk}$ can be computed once and stored, while the term $\mathbf{x}'_{jk}\mathbf{D^{-1}}\mathbf{y}^*_{\boldsymbol{k}}$ is equal to the right-hand side of the reduced model for estimating $\hat{\alpha}_{jk}$. Computation of the log-likelihoods is therefore efficient, which means that the implementation of this model using right-hand-side updating (Calus, 2014) is relatively straightforward. Full details on the derivation of the log-likelihood are given in the Appendix. Finally, the conditional posterior density of $\sigma^2_{e_k}$ is a scaled inverse-χ^2 distribution:

$$\sigma^2_{e_k}|\mathbf{e_k} \sim \chi^{-2}(m-2, \mathbf{e_k}'\mathbf{e_k})$$

where m is the number of animals with records, and $\mathbf{e_k}$ is a vector with the current residuals.

The BSSVS models were implemented in a Gibbs sampler, using right-hand-side updating (Calus, 2014). For all applications of the BSSVS model, a Gibbs chain of 100,000 iterations was used, discarding the first 20,000 as burn-in. Hereafter, mt-BSSVS-1π refers to

311 the mt model using a single π value, which considers that an SNP is linked to a QTL in both
312 breeds or not, and mt-BSSVS-2π refers to the mt model using a breed-specific π value, which
313 considers that an SNP was: 1) linked to a QTL in both breeds, 2) linked to a QTL in only one
314 of the breeds, or 3) not linked to a QTL in both breeds.

315 ***Model comparison.*** Accuracy and bias of the predictions were computed for the
316 validation bulls. Accuracy was simply computed as the correlation between the observed DRP
317 and the (G)EBV of the validation bulls. The significance of the differences in this correlation
318 was assessed using the Hotel-Williams test (Steiger, 1980). Bias was assessed by the coefficient
319 of the regression of the observed DRPs in the EBVs of the validation bulls.

320 To assess the underlying differences of the applied BSSVS models (for instance, to
321 assess whether using the Holstein data helped to increase the evidence that certain loci are
322 important for the prediction in Jerseys), posterior probabilities of the same locus obtained with
323 different models were compared. Posterior probabilities were computed as the posterior mean
324 of the QTL indicator I_{jk}. In addition, to visualize the evidence for QTLs being present across
325 the genome, Manhattan plots of the Bayes factors of each of the loci for each of the BSSVS
326 models were created as well. Bayes factors were computed as:

327 $$BF = \frac{Pr(H_1|y)}{1 - Pr(H_1|y)} \div \frac{Pr(H_1)}{1 - Pr(H_1)}$$

328 where H_1 is the hypothesis that the variant has a large effect, $Pr(H_1|y)$ is the posterior probability
329 of the hypothesis, and $Pr(H_1)$ is the prior probability of the hypothesis. (1 - $Pr(H_1|y)$) and (1 -
330 $Pr(H_1)$) represent the posterior and prior probability for the alternative hypothesis, respectively.
331 A high Bayes factor indicates that a variant has a strong association with the trait.

332

333 **RESULTS**

334 ***Accuracy and Bias of Genomic Prediction***

Genomic relationships between Holstein and Jersey were expected to be symmetric around zero. The estimates revealed that some relationships between Holstein and Jersey individuals were higher than expected, showing some admixture in the population (Figure 1). Estimated genetic correlations between Holsteins and Jerseys were 0.66 for milk, 0.05 for fat and 0.47 for protein yields (Table 1). The standard errors indicated that all genetic correlations were significantly smaller than unity, and the genetic correlations for milk and protein were significantly higher than 0.

The accuracies regarding all genomic prediction scenarios are presented in Table 2. Genomic prediction accuracies from all scenarios that included Jersey animals in the training data were higher than those obtained with the standard pedigree-based model, which only used Jersey training data. When only Holstein animals were used as training data, genomic prediction accuracies were always lower than those based on the pedigree-based model, which only used Jersey training data, although the difference was relatively small for milk and protein yields, especially when the ST-BSSVS model was used. Pooling Holstein and Jersey training data significantly improved the accuracy of the predictions of the ST-GREML model for milk and protein yields, but not for fat yield. The MT-GREML model only achieved a significantly higher accuracy for milk yield, compared to the ST-GREML model, when using only Jersey data. Using combined Jersey and Holstein training data, the MT-GREML model produced somewhat lower accuracies than the ST-GREML model with pooled data.

Pooling Holstein and Jersey training data significantly improved the accuracy of the predictions of the ST-BSSVS model for all three traits. The mt-BSSVS models achieved a significantly higher accuracy for milk and fat yields, compared to the ST-BSSVS model using only Jersey data. Using pooled Jersey and Holstein training data, both mt-BSSVS models persistently achieved somewhat lower accuracies than the ST-BSSVS model with pooled data. Accuracies of the mt-BSSVS-1π and mt-BSSVS-2π were very similar.

In almost all cases the BSSVS models yielded higher accuracies than the GREML models when applied to the same data. Differences were most pronounced and most often significant for the ST models using only Holstein or Holstein and Jersey animals in the training data. For the scenario where only Jersey animals were included in the training, the BSSVS model only produced a significant increase in prediction accuracy, compared to GREML, for fat yield.

Bias in the GEBV scale was assessed by the coefficient of the regression of observed DRPs on the GEBV (Table 3), where a value of 1 is expected if the GEBVs are unbiased. The largest deviation from 1, i.e., a regression coefficient of 0.50, was observed for fat yield using ST-GREML, when the training only included Holstein data. When using only Jersey data, the BSSVS model produced less biased predictions than the GREML model. When using both Holstein and Jersey training data, ST-GREML gave less biased predictions than ST-BSSVS, while both mt-BSSVS models gave less biased predictions than MT-GREML.

Visualization of QTL Detection in the Bayesian Models

For each of the five applications of the BSSVS model, Manhattan plots of the Bayes factors were made for the three traits (Figures 2-4). When using only Jersey data, a few QTL peaks were observed for milk yield, but no peaks were detected for fat and protein yield (Figures 2-4). It should be noted that the scenarios that only used Holstein or pooled data in the ST-BSSVS model or used Holstein and Jersey data in mt-BSSVS-1π effectively represent QTLs found in Holsteins. Whenever the Holstein data were used, for each trait, several clear QTL peaks were observed, regardless of which BSSVS model was used or whether the Jersey data were also used or not. In most cases, the peaks observed for the analyses including the Holstein data were the same across models and training data composition. There were, however, a few exceptions. For instance, on BTA 1, an SNP at 49,950,467 bp with a relatively large Bayes

385 factor for milk yield was observed when using only Jersey data or when using one of the mt-
386 BSSVS models, while this SNP did not appear in the other two analyses. Since the other two
387 analyses were mostly driven by Holstein data, this suggests that this particular SNP has an
388 association in Jerseys, but not in Holsteins. In contrast, there were also a few peaks that were
389 clearer when the ST-BSSVS, with the pooled Jersey and Holstein data, was used, compared to
390 using either of the mt-BSSVS models. This was, for instance, the case for a peak at the
391 beginning of BTA 3 for fat yield, and at the end of BTA 3 for protein yield, as well as for an
392 SNP at 28,842,616 bp on BTA 10 for protein yield.

393 To enable a more precise comparison of (trends of) differences in associations at the
394 individual locus level across different models, the underlying posterior probabilities for the
395 same locus obtained with different BSSVS models were plotted against each other
396 (Supplemental Figures S1-S10; http://dx.doi.org/10.3168/jds.20XX-XXXXX). These results
397 confirm that, when using Jersey data alone, there was limited evidence for clearly segregating
398 QTLs, i.e., few posterior probabilities noticeably larger than 0 were observed (Supplemental
399 Figures S1-S4; http://dx.doi.org/10.3168/jds.20XX-XXXXX). All analyses including Holstein
400 data forced the Holstein QTLs in the model, i.e. posterior probabilities were similar, on the one
401 hand, for the ST-BSSVS model using only Holstein data, compared either to the ST-BSSVS
402 model using both Holstein and Jersey data or to the mt-BSSVS-1π model using both Holstein
403 and Jersey data (Supplemental Figures S5 and S6, respectively;
404 http://dx.doi.org/10.3168/jds.20XX-XXXXX). While the same tendency was observed for the
405 mt-BSSVS-2π model, in this case, the posterior probabilities for Jerseys were generally smaller
406 than those obtained with the ST-BSSVS model and Holstein data (Supplemental Figure S7;
407 http://dx.doi.org/10.3168/jds.20XX-XXXXX). Finally, the posterior probabilities of mt-
408 BSSVS-1π were similar to those of ST-BSSVS using pooled Holstein and Jersey data, while

they tended to be smaller for mt-BSSVS-2π (Supplemental Figures S8 and S9 versus S10; http://dx.doi.org/10.3168/jds.20XX-XXXXX).

DISCUSSION

One of the objectives of our study was to compare the predictive ability of different genomic prediction models for a breed with a small training set size (i.e. Jerseys) when supplemented with another breed with a moderate training set size (i.e. Holsteins). Estimated genetic correlations between Holsteins and Jerseys were 0.05 for fat yield and 0.66 and 0.47 for milk and protein yields, respectively, suggesting that milk and protein yields have, at least partially, the same genetic background across Holstein and Jersey cattle. Apart from the low correlation for fat yield, these results are in line with an estimated genetic correlation of 0.79 for milk yields between Montbéliardes and Holsteins (Karoui et al., 2012), and estimated genetic correlations of 0.46, 0.58 and 0.37 for milk, fat and protein yields between Nordic Reds and Holsteins (Zhou et al., 2014), where some admixture between the breeds exists. Genomic relationships between the Holsteins and Jerseys (Figure 1) showed admixture between these breeds in our data, in line with the notion that the Holstein population in Australia is the result of upgrading from Jerseys (Pryce et al., 2011). Thus, the Holsteins could still carry some Jersey chromosome segments, which is the likely explanation for the observed superiority of the pooling strategy over the mt-BSSVS models, and the higher estimates of the genetic correlations compared to those obtained by van den Berg et al. (2016b) between Holsteins and Jerseys, using a model component based on 50,000 SNPs. Our correlations, however, were similar or somewhat lower than those obtained, based on a QTL component that included sequence variants significantly associated in multibreed genome-wide association studies (van

den Berg et al., 2016b). These authors also found the lowest correlation for fat, compared to milk and protein, yields.

The observation that milk and protein had considerably higher genetic correlations than fat is in line with the result that milk and protein yields showed larger gains, compared to fat, when using the pooled training data, instead of only the Jersey training data. This is also in line with the result that using only Holstein training data in the ST-GREML model produced considerably higher accuracies for milk and protein (0.20-0.24), compared to fat (0.10). Finally, using ST-BSSVS and only Holstein data for training yielded considerable accuracies of 0.17-0.31, in line with the observation of admixture between Holsteins and Jerseys in our data.

Comparison of Models

The main objective of our study was to compare the predictive ability of the mt-BSSVS models with ST-GREML, ST-BSSVS and MT-GREML models. Simply pooling the data of multiple breeds into an ST genomic prediction model, may be appropriate when breeds are closely related; in which case, it is expected that the genetic correlation between breeds is high. When genetic correlations between breeds become smaller, it is expected that an MT or mt model may be more appropriate, for example, when a QTL with a large effect in one breed only has a small effect in the other breed. That is, MT and mt models are more flexible than ST models when translating effects across breeds, as these models have varying degrees of opportunities to model breed-specific effects. The mt-BSSVS models are similar to an approach where QTL mapping results obtained from one breed are used as prior information in genomic prediction for a second breed (Brøndum et al., 2012) or when information from another breed is used to select or give higher weights to SNPs (Hoze et al., 2014; Khansefid et al., 2014; van den Berg et al., 2016b). It has been shown that a strategy of “partial pooling”, allowing for the estimation of breed- or (sub)population-specific SNP effects, which are “shrunk” towards

effects across all breeds or (sub)populations, makes optimal use of information on training sets involving different populations (Technow and Totir, 2015). In reality, between two breeds, there may be some QTLs with (large) common effects, while there may be others with breed-specific effects. Our results show that ST-GREML and ST-BSSVS using the pooled training data consistently outperformed MT-GREML and the mt-BSSVS models, respectively. This finding is unexpected, given that the estimated genetic correlations between the breeds are considerably lower than 1. Several reasons may explain this. Firstly, estimating twice as many effects in the MT and mt models may counteract their benefit of being better able to accommodate the lower-than-unity genetic correlation. Secondly, the MT-GREML model has to retrieve information through the genomic relationships across the breeds, which in general are very week. In contrast, the mt-BSSVS models used information from Holsteins to indicate QTLs, but then used only information from Jerseys to estimate the SNP effects, instead of pooling the data to increase power to estimate SNP-effects as the ST models did, while this may have been the crucial benefit given our small Jersey training data. Thirdly, the traits analyzed were all associated with a few QTLs of moderate to large effects in Holsteins. For those QTLs with relatively large effects, the actual effects are likely to be similar in magnitude across breeds, as is, for instance, shown regarding the effect of the *DGAT1* gene (Spelman et al., 2002; Thaller et al., 2003; Maurice-Van Eijndhoven et al., 2015). In other words, for this specific group of QTLs, the genetic correlation between breeds is expected to be close to unity. This is, if the same QTL segregate in both breeds. It has been observed that not all the well-known QTLs in Holsteins also segregate in Jerseys (Kemper et al., 2015b), so the genetic correlations for QTLs with relatively large effects will be lower than unity. For QTLs with much smaller effects, it can be expected that their effects are less consistent across breeds, simply because their effects are less disruptive, meaning that their genetic correlation across breeds is likely to be much smaller. As genomic prediction across breeds largely relies on QTLs of moderate to

large effects (van den Berg et al., 2016b), it can be expected that it is much more important to closely fit the properties of moderate to large QTLs across breeds than it is to fit the properties of small QTLs across breeds.

Our results indicate that the BSSVS model consistently outperformed the GREML model, when Holstein or pooled data were used for training. In addition, the BSSVS model, in almost all cases, yielded significantly higher accuracies when using the pooled training data, while this phenomenon was not so profound for the GREML model. These results are in line with the observations that (Bayesian) variable selection models are better able to pick up QTLs (van den Berg et al., 2015), and that selection of SNPs close to the causative mutations yields more persistent genomic predictions across breeds (van den Berg et al., 2015; van den Berg et al., 2016a), suggesting that the BSSVS model was able to take advantage of the BovineHD SNP data with relatively high SNP density. The similar accuracies of the ST-BSSVS and ST-GREML models when only Jersey data were used, is like due to the inability of the ST-BSSVS model to find QTLs in the Jersey data. Only for fat yield did ST-BSSVS outperform GREML because the former yielded posterior probabilities of ~0.02 for several SNPs near the *DGAT1* gene, while the genome-wide average was only 0.001.

The limited benefit of pooling the training data for fat yield, and the low estimate of the genetic correlation for fat yield, may be partly due to differences in estimated effects for Holstein and Jersey in the detected QTL regions on BTA 5; associated to the *MGST1* gene in Holsteins (Wang et al., 2012; Raven et al., 2014; Kemper et al., 2015a; Maurice-Van Eijndhoven et al., 2015; Littlejohn et al., 2016), and on BTA 14; associated to the *DGAT1* gene which has a strong effect on fat yield (Grisart et al., 2002; Boichard et al., 2003; Schennink et al., 2007). Possible explanations for the differences in estimated effects between Holstein and Jerseys for these regions, are that the SNPs on BTA 5 and 14 with the largest signal in Holstein had very low MAF in Jersey. In addition, the local LD patterns in the *DGAT1* region in the

Jerseys were different than in the Holsteins (for more details, see Supplemental Material; http://dx.doi. org/10.3168/jds.20XX-XXXXX).

The use of QTL information across breeds relies on the LD consistency between SNPs and QTLs across breeds (De Roos et al., 2009; Wientjes et al., 2015). The prior specification of the mt-BSSVS-2π model assumed that, for an SNP linked to a QTL in one breed, it is equally likely to be linked to a QTL in another breed. Another approach would be to use the value of any known genetic correlations between breeds to inform this prior specification. It may also be expected that, while certain genes affect the same trait in different breeds, the causal mutations are not necessarily the same. A well-known example of this phenomenon is double-muscling, which, in different breeds, is caused by different mutations in the myostatin gene (McPherron and Lee, 1997; Grobet et al., 1998). A further extension of the mt-BSSVS models, could therefore be to consider the evidence of QTLs across a sliding window (Wientjes, 2016), or across all variants within each annotated gene, similar to a gene-based genome-wide association study approach (Liu et al., 2010), rather than only on a per variant basis.

Implications

In the Jersey training data used, we deliberately removed close relatives (i.e., sires, sibs, and grandsons of sires and dams) from the validation animals, to create some distance between training and validation animals. In reality, one would use all available information of the predicted breed, and especially information of close relatives of selection candidates, as this is the most powerful information. In this study, we aimed to resemble an ongoing genomic selection program that takes full advantage of shortening the generation intervals of the different selection paths (García-Ruiz et al., 2016). In those programs, it is reasonable to assume that there are at least two generations between the training animals and the selection candidates.

The situation considered in our study, where one breed has a limited training data size, is relevant for numerically small breeds. In our study, using the pooled Holstein and Jersey training data in the ST-BSSVS model gave the largest increase in accuracy. We focused here on production traits with moderate heritability and at least some known QTL with relatively large effects. An important unanswered question is whether the same result is expected for more polygenic traits with a low heritability, and possibly a lower genetic correlation between breeds. We hypothesize that for such traits the assumption made when using pooled training data in an ST model, i.e. the genetic correlation between breeds is 1, may be violated too much, and that, in such situations, the mt-BSSVS models may better fit the characteristics of the data.

CONCLUSIONS

In this study, we investigated the use of an mt Bayesian model, which accumulates evidence across breeds to indicate whether SNPs are linked to a QTL and thus should receive a large effect, while the SNP effects are estimated within breeds. We further developed this model, such that it is able to model breed-specific probabilities for SNPs in order to have a large effect on the trait under study. Both mt Bayesian models, however, were slightly outperformed by a simple pooling strategy, where data on Holsteins and Jerseys were combined in an ST Bayesian model to predict Jerseys. This result may be partly due to the fact that we considered the moderately heritable traits of milk, fat and protein yields, which are affected by some moderate to large QTLs. Milk and protein had moderately estimated genetic correlations (0.66 and 0.47) between Holsteins and Jerseys, in line with the observed increases in accuracy when adding Holsteins to predict Jerseys with both the ST-GREML (0.06 vs. 0.04) and the ST-BSSVS model (0.10 vs. 0.05). The relatively high genetic correlations for milk and protein yields, and the superiority of the pooling strategy in terms of prediction accuracy, is likely the result of the observed admixture between both breeds in our data. Fat yield had an estimated

genetic correlation of only 0.05, in line with the observed more limited increases in accuracy with both the ST-GREML (0.01) and the ST-BSSVS model (0.04). The comparison between the Bayesian model and the GREML model shows there is some scope for multibreed genomic prediction, especially if the model used is able to pinpoint underlying QTLs.

ACKNOWLEDGMENTS

The first author acknowledges financial support from CRV BV (Arnhem, the Netherlands) and the Dutch Ministry of Economic Affairs (TKI Agri & Food project 12018; "Breed4Food" public-private partnership code BO-22.04-011-001-ASG-LR-3), as well as EU FP7 IRSES SEQSEL (Grant no. 317697), which funded the research visit of Mario Calus at AgriBio, the Centre For AgriBioscience of the Department of Environmental and Primary Industries and La Trobe University, in Victoria, Australia. Both CRV BV and the Australian Dairy Herd Improvement Scheme (ADHIS) are acknowledged for providing data. Iona MacLeod is acknowledged for performing initial data edits. The use of the HPC cluster has been made possible by CAT-AgroFood (Shared Research Facilities Wageningen UR).

APPENDIX

The log-likelihood of $I_{jk} = \delta$ is computed across breeds *k* as:

$$\mathrm{f}\left(r_{jk}\middle|I_{jk}=\delta\right)(pr) = \sum\nolimits_{k}\left(-\frac{1}{2}\log\left(\left|\mathbf{V}_{\boldsymbol{y}_{\boldsymbol{k}}^{*}}\right|\right)-\frac{1}{2}y_{k}^{*}\mathbf{V}_{\boldsymbol{y}_{\boldsymbol{k}}^{*}}^{-\mathbf{1}}y_{k}^{*}+\log(pr\boldsymbol{ior})\right)$$

where, for breed *k*:

$$\mathbf{V}_{\boldsymbol{y}_{\boldsymbol{k}}^{*}} = \mathbf{D}\widehat{\sigma}_{e_k}^{2}+\mathbf{x}_{jk}\mathbf{x}_{jk}^{\prime}\frac{\widehat{\sigma}_{\alpha_{jk}}^{2}}{\omega_{jk}}$$

Considering that:

$$\mathbf{V}_{\boldsymbol{y}_{\boldsymbol{k}}^{*}} = \mathbf{D}\widehat{\sigma}_{e_k}^2 + \mathbf{x}_{jk}\mathbf{x}'_{jk}\frac{\widehat{\sigma}_{\alpha_{jk}}^2}{\omega_{jk}} = \left(\mathbf{D}\frac{\omega_{jk}\widehat{\sigma}_{e_k}^2}{\widehat{\sigma}_{\alpha_{jk}}^2} + \mathbf{x}_{jk}\mathbf{x}'_{jk}\right)\frac{\omega_{jk}}{\widehat{\sigma}_{\alpha_{jk}}^2}$$

$\left|\mathbf{V}_{\boldsymbol{y}_{\boldsymbol{k}}^{*}}\right|$ can be obtained using the matrix determinant lemma:

$$\left|\mathbf{V}_{\boldsymbol{y}_{\boldsymbol{k}}^{*}}\right| = \left|\left(\mathbf{D}\frac{\omega_{jk}\widehat{\sigma}_{e_k}^2}{\widehat{\sigma}_{\alpha_{jk}}^2} + \mathbf{x}_{jk}\mathbf{x}'_{jk}\right)\frac{\omega_{jk}}{\widehat{\sigma}_{\alpha_{jk}}^2}\right| = \left|\left(\mathbf{D}\frac{\omega_{jk}\widehat{\sigma}_{e_k}^2}{\widehat{\sigma}_{\alpha_{jk}}^2} + \mathbf{x}_{jk}\mathbf{x}'_{jk}\right)\right|\left|\frac{\omega_{jk}}{\widehat{\sigma}_{\alpha_{jk}}^2}\right|$$

$$= \left(1 + \mathbf{x}'_{jk}\mathbf{D}^{-\mathbf{1}}\frac{\widehat{\sigma}_{\alpha_{jk}}^2}{\omega_{jk}\widehat{\sigma}_{e_k}^2}\mathbf{x}_{jk}\right)\left|\frac{\omega_{jk}}{\widehat{\sigma}_{\alpha_{jk}}^2}\right|$$

$\mathbf{V}_{\boldsymbol{y}_{\boldsymbol{k}}^{*}}^{-\mathbf{1}}$ can be computed using Woodbury's matrix identity:

$$\mathbf{V}_{\boldsymbol{y}_{\boldsymbol{k}}^{*}}^{-\mathbf{1}} = \left(\mathbf{D}\widehat{\sigma}_{e_k}^2 + \mathbf{x}_{jk}\mathbf{x}'_{jk}\frac{\widehat{\sigma}_{\alpha_{jk}}^2}{\omega_{jk}}\right)^{-\mathbf{1}}$$

$$= \mathbf{D}^{-\mathbf{1}}\widehat{\sigma}_{e_k}^{-2} - \mathbf{D}^{-\mathbf{1}}\widehat{\sigma}_{e_k}^{-2}\mathbf{x}_{jk}\left(\frac{\omega_{jk}}{\widehat{\sigma}_{\alpha_{jk}}^2} + \mathbf{x}'_{jk}\mathbf{D}^{-\mathbf{1}}\widehat{\sigma}_{e_k}^{-2}\mathbf{x}_{jk}\right)^{-\mathbf{1}}\mathbf{x}'_{jk}\mathbf{D}^{-\mathbf{1}}\widehat{\sigma}_{e_k}^{-2}$$

$$= \mathbf{D}^{-\mathbf{1}}\widehat{\sigma}_{e_k}^{-2} - \mathbf{D}^{-\mathbf{1}}\mathbf{x}_{jk}\mathbf{x}'_{jk}\mathbf{D}^{-\mathbf{1}}\widehat{\sigma}_{e_k}^{-2}\left(\frac{\omega_{jk}\widehat{\sigma}_{e_k}^2}{\widehat{\sigma}_{\alpha_{jk}}^2} + \mathbf{x}'_{jk}\mathbf{D}^{-\mathbf{1}}\mathbf{x}_{jk}\right)^{-\mathbf{1}}$$

Thus:

$$y_k^{*'}\mathbf{V}_{\boldsymbol{y}_{\boldsymbol{k}}^{*}}^{-\mathbf{1}}y_k^{*} = y_k^{*'}\mathbf{D}^{-\mathbf{1}}\widehat{\sigma}_{e_k}^{-2} - y_k^{*'}\mathbf{D}^{-\mathbf{1}}\mathbf{x}_{jk}\mathbf{x}'_{jk}\mathbf{D}^{-\mathbf{1}}y_k^{*}\widehat{\sigma}_{e_k}^{-2}\left(\frac{\omega_{jk}\widehat{\sigma}_{e_k}^2}{\widehat{\sigma}_{\alpha_{jk}}^2} + \mathbf{x}'_{jk}\mathbf{D}^{-\mathbf{1}}\mathbf{x}_{jk}\right)^{-\mathbf{1}}$$

$$= y_k^{*'}\mathbf{D}^{-\mathbf{1}}\widehat{\sigma}_{e_k}^{-2} - \frac{\left(\mathbf{x}'_{jk}\mathbf{D}^{-\mathbf{1}}\mathbf{y}_{\boldsymbol{k}}^{*}\right)^{\mathbf{2}}}{\widehat{\sigma}_{e_k}^2\left(\frac{\omega_{jk}\widehat{\sigma}_{e_k}^2}{\widehat{\sigma}_{\alpha_{jk}}^2} + \mathbf{x}'_{jk}\mathbf{D}^{-\mathbf{1}}\mathbf{x}_{jk}\right)}$$

As such, after dropping terms that are equivalent for $I_{jk} = 0$ and $I_{jk} = 1$ (i.e., $y_k^{*'}\mathbf{D}^{-\mathbf{1}}\widehat{\sigma}_{e_k}^{-2}$), the log-likelihood becomes:

$$\mathrm{f}\left(r_{jk}\middle|I_{jk} = \delta\right)(pr) =$$

$$\sum_k\left(-\frac{1}{2}\log\left(1 + \mathbf{x}'_{jk}\mathbf{D}^{-\mathbf{1}}\mathbf{x}_{jk}\frac{\widehat{\sigma}_{\alpha_{jk}}^2}{\omega_{jk}\widehat{\sigma}_{e_k}^2}\right) + \frac{1}{2}\frac{\left(\mathbf{x}'_{jk}\mathbf{D}^{-\mathbf{1}}\mathbf{y}_{\boldsymbol{k}}^{*}\right)^{\mathbf{2}}}{\sum_k\widehat{\sigma}_{e_k}^2\left(\frac{\omega_{jk}\widehat{\sigma}_{e_k}^2}{\widehat{\sigma}_{\alpha_{jk}}^2} + \mathbf{x}'_{jk}\mathbf{D}^{-\mathbf{1}}\mathbf{x}_{jk}\right)} + \log(prior)\right)$$

593

REFERENCES

Boichard, D., C. Grohs, F. Bourgeois, F. Cerqueira, R. Faugeras, A. Neau, R. Rupp, Y. Amigues, M. Y. Boscher, and H. Levéziel. 2003. Detection of genes influencing economic traits in three French dairy cattle breeds. Genet. Sel. Evol. 35:77-102.

Brøndum, R. F., E. Rius-Vilarrasa, I. Strandén, G. Su, B. Guldbrandtsen, W. F. Fikse, and M. S. Lund. 2011. Reliabilities of genomic prediction using combined reference data of the Nordic Red dairy cattle populations. J. Dairy Sci. 94:4700-4707.

Brøndum, R. F., G. Su, M. S. Lund, P. J. Bowman, M. E. Goddard, and B. J. Hayes. 2012. Genome position specific priors for genomic prediction. BMC Genomics. 13:543.

Browning, B. L. and S. R. Browning. 2009. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. Am. J. Hum. Genet. 84:210-223.

Calus, M. P. L. 2014. Right-hand-side updating for fast computing of genomic breeding values. Genet. Sel. Evol. 46:24.

Calus, M. P. L., H. Huang, A. Vereijken, J. Visscher, J. Ten Napel, and J. J. Windig. 2014a. Genomic prediction based on data from three layer lines: a comparison between linear methods. Genet. Sel. Evol. 46:57.

Calus, M. P. L., C. Schrooten, and R. F. Veerkamp. 2014b. Genomic prediction of breeding values using previously estimated SNP variances. Genet. Sel. Evol. 46:52.

Chen, L., C. Li, S. Miller, and F. Schenkel. 2014. Multi-population genomic prediction using a multi-task Bayesian learning model. BMC Genet. 15:53.

Daetwyler, H. D., M. P. L. Calus, R. Pong-Wong, G. de los Campos, and J. M. Hickey. 2013. Genomic prediction in animals and plants: simulation of data, validation, reporting, and benchmarking. Genetics. 193:347-365.

de los Campos, G., J. M. Hickey, R. Pong-Wong, H. D. Daetwyler, and M. P. L. Calus. 2013. Whole-genome regression and prediction methods applied to plant and animal breeding. Genetics. 193:327-345.

De Roos, A. P. W., B. J. Hayes, and M. E. Goddard. 2009. Reliability of genomic predictions across multiple populations. Genetics. 183:1545-1553.

De Roos, A. P. W., B. J. Hayes, R. J. Spelman, and M. E. Goddard. 2008. Linkage disequilibrium and persistence of phase in Holstein-Friesian, Jersey and Angus cattle. Genetics. 179:1503-1512.

Erbe, M., B. J. Hayes, L. K. Matukumalli, S. Goswami, P. J. Bowman, C. M. Reich, B. A. Mason, and M. E. Goddard. 2012. Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. J. Dairy Sci. 95:4114-4129.

García-Ruiz, A., J. B. Cole, P. M. VanRaden, G. R. Wiggans, F. J. Ruiz-López, and C. P. Van Tassell. 2016. Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection. Proc. Nat. Acad. Sci. 113:E3995-E4004.

Gilmour, A. R., B. J. Gogel, B. R. Cullis, S. J. Welham, and R. Thompson. 2014. ASReml User Guide Release 4.0. VSN International Ltd, Hemel Hempstead, UK.

Grisart, N., W. Coppieters, F. Farnir, L. Karim, C. Ford, P. Berzi, N. Cambisamo, M. Mni, P. Simon, R. Spelman, M. Georges, and R. Snell. 2002. Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. Genome Res. 12:222-231.

Grobet, L., D. Poncelet, L. J. Royo, B. Brouwers, D. Pirottin, C. Michaux, F. Ménissier, M. Zanotti, S. Dunner, and M. Georges. 1998. Molecular definition of an allelic series of

642 mutations disrupting the myostatin function and causing double-muscling in cattle.
643 Mamm. Genome. 9:210-213.
644 Habier, D., R. L. Fernando, K. Kizilkaya, and D. J. Garrick. 2011. Extension of the Bayesian
645 alphabet for genomic selection. BMC Bioinformatics. 12:186.
646 Hoze, C., S. Fritz, F. Phocas, D. Boichard, V. Ducrocq, and P. Croiseau. 2014. Genomic
647 evaluation using combined reference populations from Montbéliarde and French
648 Simmental breeds. Pages 17-22 in 10th World Congress of Genetics Applied to Livestock
649 Production.
650 Hozé, C., S. Fritz, F. Phocas, D. Boichard, V. Ducrocq, and P. Croiseau. 2014. Efficiency of
651 multi-breed genomic selection for dairy cattle breeds with different sizes of reference
652 population. J. Dairy Sci. 97:3918-3929.
653 Huang, H., J. J. Windig, A. Vereijken, and M. P. L. Calus. 2014. Genomic prediction based on
654 data from three layer lines using non-linear regression models. Genet. Sel. Evol. 46:75.
655 Jia, Y. and J.-L. Jannink. 2012. Multiple-trait genomic selection methods increase genetic value
656 prediction accuracy. Genetics. 192:1513-1522.
657 Karoui, S., M. J. Carabano, C. Diaz, and A. Legarra. 2012. Joint genomic evaluation of French
658 dairy cattle breeds using multiple-trait models. Genet. Sel. Evol. 44:39.
659 Kemper, K., C. Reich, P. Bowman, C. vander Jagt, A. Chamberlain, B. Mason, B. Hayes, and
660 M. Goddard. 2015a. Improved precision of QTL mapping using a nonlinear Bayesian
661 method in a multi-breed population leads to greater accuracy of across-breed genomic
662 predictions. Genet. Sel. Evol. 47:29.
663 Kemper, K. E., B. J. Hayes, H. D. Daetwyler, and M. E. Goddard. 2015b. How old are
664 quantitative trait loci and how widely do they segregate? J. Anim. Breed. Genet. 132:121-
665 134.

Khansefid, M., J. E. Pryce, S. Bolormaa, S. P. Miller, Z. Wang, C. Li, and M. E. Goddard. 2014. Estimation of genomic breeding values for residual feed intake in a multibreed cattle population. J. Anim. Sci. 92:3270-3283.

Littlejohn, M. D., K. Tiplady, T. A. Fink, K. Lehnert, T. Lopdell, T. Johnson, C. Couldrey, M. Keehan, R. G. Sherlock, C. Harland, A. Scott, R. G. Snell, S. R. Davis, and R. J. Spelman. 2016. Sequence-based association analysis reveals an MGST1 eQTL with pleiotropic effects on bovine milk composition. Sci. Rep. 6:25376.

Liu, J. Z., A. F. McRae, D. R. Nyholt, S. E. Medland, N. R. Wray, K. M. Brown, N. K. Hayward, G. W. Montgomery, P. M. Visscher, N. G. Martin, and S. Macgregor. 2010. A versatile gene-based test for genome-wide association studies. Am. J. Hum. Genet. 87:139-145.

Lund, M., S. de Roos, A. de Vries, T. Druet, V. Ducrocq, S. Fritz, F. Guillaume, B. Guldbrandtsen, Z. Liu, R. Reents, C. Schrooten, F. Seefried, and G. Su. 2011. A common reference population from four European Holstein populations increases reliability of genomic predictions. Genet. Sel. Evol. 43:43.

Lund, M. S., G. Su, L. Janss, B. Guldbrandtsen, and R. F. Brøndum. 2014. Invited review: Genomic evaluation of cattle in a multi-breed context. Livest. Sci. 166:101-110.

Maurice-Van Eijndhoven, M. H. T., H. Bovenhuis, R. F. Veerkamp, and M. P. L. Calus. 2015. Overlap in genomic variation associated with milk fat composition in Holstein Friesian and Dutch native dual-purpose breeds. J. Dairy Sci. 98:6510-6521.

McPherron, A. C. and S.-J. Lee. 1997. Double muscling in cattle due to mutations in the myostatin gene. Proc. Nat. Acad. Sci. USA. 94:12457-12461.

Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics. 157:1819-1829.

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

691 Pryce, J. E., B. Gredler, S. Bolormaa, P. J. Bowman, C. Egger-Danner, C. Fuerst, R. Emmerling,
692 J. Solkner, M. E. Goddard, and B. J. Hayes. 2011. Short communication: Genomic
693 selection using a multi-breed, across-country reference population. J. Dairy Sci. 94:2625-
694 2630.
695 Raven, L.-A., B. G. Cocks, and B. J. Hayes. 2014. Multibreed genome wide association can
696 improve precision of mapping causative variants underlying milk production in dairy
697 cattle. BMC Genomics. 15:62.
698 Schennink, A., W. M. Stoop, M. H. P. W. Visker, J. M. L. Heck, H. Bovenhuis, J. J. Van Der
699 Poel, H. J. F. Van Valenberg, and J. A. M. Van Arendonk. 2007. DGAT1 underlies large
700 genetic variation in milk-fat composition of dairy cows. Anim. Genet. 38:467-473.
701 Spelman, R. J., C. A. Ford, P. McElhinney, G. C. Gregory, and R. G. Snell. 2002.
702 Characterization of the DGAT1 gene in the New Zealand dairy population. J. Dairy Sci.
703 85:3514-3517.
704 Steiger, J. H. 1980. Tests for comparing elements of a correlation matrix. Psych. Bull. 87:245.
705 Technow, F. and L. R. Totir. 2015. Using Bayesian multilevel whole genome regression models
706 for partial pooling of training sets in genomic prediction. G3 Genes Genom. Genet.
707 5:1603-1612.
708 Thaller, G., W. Krämer, A. Winter, B. Kaupe, G. Erhardt, and R. Fries. 2003. Effects of DGAT1
709 variants on milk production traits in German cattle breeds. J. Anim. Sci. 81:1911-1918.
710 van den Berg, I., D. Boichard, B. Guldbrandtsen, and M. S. Lund. 2016a. Using sequence
711 variants in linkage disequilibrium with causative mutations to improve across-breed
712 prediction in dairy cattle: a simulation study. G3 Genes Genom. Genet. 6:2553-2561.
713 van den Berg, I., D. Boichard, and M. S. Lund. 2016b. Sequence variants selected from a multi-
714 breed GWAS can improve the reliability of genomic predictions in dairy cattle. Genet.
715 Sel. Evol. 48:83.

van den Berg, S., M. P. L. Calus, T. H. E. Meuwissen, and Y. C. J. Wientjes. 2015. Across population genomic prediction scenarios in which Bayesian variable selection outperforms GBLUP. BMC Genet. 16:146.

VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91:4414-4423.

Verbyla, K. L., B. J. Hayes, P. J. Bowman, and M. E. Goddard. 2009. Accuracy of genomic selection using stochastic search variable selection in Australian Holstein Friesian dairy cattle. Genet. Res. 91:307-311.

Wang, X., C. Wurmser, H. Pausch, S. Jung, F. Reinhardt, J. Tetens, G. Thaller, and R. Fries. 2012. Identification and dissection of four major QTL affecting milk fat content in the German Holstein-Friesian population. PLoS ONE. 7:e40711.

Wientjes, Y. C. J. 2016. Multi-population genomic prediction. PhD thesis, Wageningen University, Wageningen.

Wientjes, Y. C. J., P. Bijma, J. Vandenplas, and M. P. L. Calus. 2017. Multi-population genomic relationships for estimating current genetic variances within and genetic correlations between populations. Genetics. 207:503-515.

Wientjes, Y. C. J., M. P. L. Calus, M. E. Goddard, and B. J. Hayes. 2015. Impact of QTL properties on the accuracy of multi-breed genomic prediction. Genet. Sel. Evol. 47:42.

Zhou, L., M. S. Lund, Y. Wang, and G. Su. 2014. Genomic predictions across Nordic Holstein and Nordic Red using the genomic best linear unbiased prediction model with different genomic relationship matrices. J. Anim. Breed. Genet. 131:249-257.

Table 1. Estimated genetic correlations between milk production traits in Jerseys and Holsteins using GREML.

Trait	Genetic correlation (SE)

Milk	0.661 (0.143)
Fat	0.050 (0.158)
Protein	0.470 (0.157)

740

741

742 **Table 2**. GEBV accuracies estimated with either single-trait implementations of PREML (a
743 pedigree-based model using REML), GREML (a genomic relationship-based model using
744 REML) or BSSVS (Bayesian Stochastic Search Variable Selection). For GREML, a multi-trait
745 model (MT-GREML) is also used, while, for BSSVS, two mt Bayesian learning models (mt-
746 BSSVS) are used, which either have one π value for both breeds (mt-BSSVS-1π) or a different
747 π value separately for each breed (mt-BSSVS-2π). Models were trained on data from Jerseys
748 (J), Holsteins (H) or both (H/J).

Model	Milk	Fat	Protein
ST-PREML (J)	0.326	0.582	0.317
ST-GREML (H)	0.204***	0.100***	0.237***
ST-GREML (J)	0.461^{-}	0.609^{-}	0.484^{-}
ST-GREML (H/J)	0.525**	0.614^{ns}	0.524*
MT-GREML (H/J)	0.493**	0.609^{ns}	0.496†
ST-BSSVS (H)	$0.311^{**;***}$	$0.173^{***;*}$	$0.265^{***;\dagger}$
ST-BSSVS (J)	$0.465^{-;ns}$	$0.628^{-;***}$	$0.490^{-;ns}$
ST-BSSVS (H/J)	$0.561^{***;**}$	$0.665^{**;***}$	$0.538^{**;\dagger}$
mt-BSSVS-1π (H/J)	$0.498^{***;ns}$	$0.637^{*;***}$	$0.500^{ns;ns}$
mt-BSSVS-2π (H/J)	$0.493^{**;ns}$	$0.636^{\dagger;***}$	$0.494^{ns;ns}$

749 The first superscript denotes, within-trait and -model (GREML or BSSVS), whether the
750 accuracy is significantly lower (for animals using only H data) or higher (for models using J
751 and H data), compared to the standard (single-trait) model, which uses only J data; – indicates
752 the standard model, ns indicates *P*-values ≥ 0.10, † indicates *P*-values < 0.10, * indicates *P*-
753 values < 0.05, ** indicates *P*-values < 0.01, and *** indicates *P*-values < 0.001.
754 The second superscript for the BSSVS models denotes whether their accuracy is significantly
755 larger than the accuracy of their GREML counterpart, i.e., ST-BSSVS (H) vs. ST-GREML (H),

ST-BSSVS (J) vs. ST-GREML (J), ST-BSSVS (H/J) vs. ST-GREML (H/J), mt-BSSVS (H/J) vs. MT-GREML (H/J); ns indicates *P*-values ≥ 0.10, † indicates *P*-values < 0.10, * indicates *P*-values < 0.05, ** indicates *P*-values < 0.01, *** indicates *P*-values < 0.001.

Table 3. Coefficients of the regression of de-regressed EBVs on (G)EBVs estimated with either single-trait (ST) implementations of PREML (a pedigree-based model using REML), GREML (a genomic relationship-based model using REML) or BSSVS (Bayesian Stochastic Search Variable Selection). For GREML, a multi-trait model (MT-GREML) is also used, while, for BSSVS, two multi-task Bayesian learning models (mt-BSSVS) are used, which either have one π value for both breeds (mt-BSSVS-1π) or a different π value separately for each breed (mt-BSSVS-2π). Models were trained on data from Jerseys (J), Holsteins (H) or both (H/J).

Model	Milk	Fat	Protein
ST PREML (J)	0.801	1.097	0.924
ST-GREML (H)	0.806	0.504	1.034
ST-GREML (J)	0.804	0.928	0.903
ST-GREML (H/J)	0.963	0.979	0.998
MT-GREML (H/J)	0.832	0.927	0.899
ST-BSSVS (H)	0.989	0.844	1.151
ST-BSSVS (J)	0.882	1.009	1.020
ST-BSSVS (H/J)	1.027	1.124	1.066
mt-BSSVS-1π (H/J)	0.895	0.991	0.944
mt-BSSVS-2π (H/J)	0.888	0.988	0.932

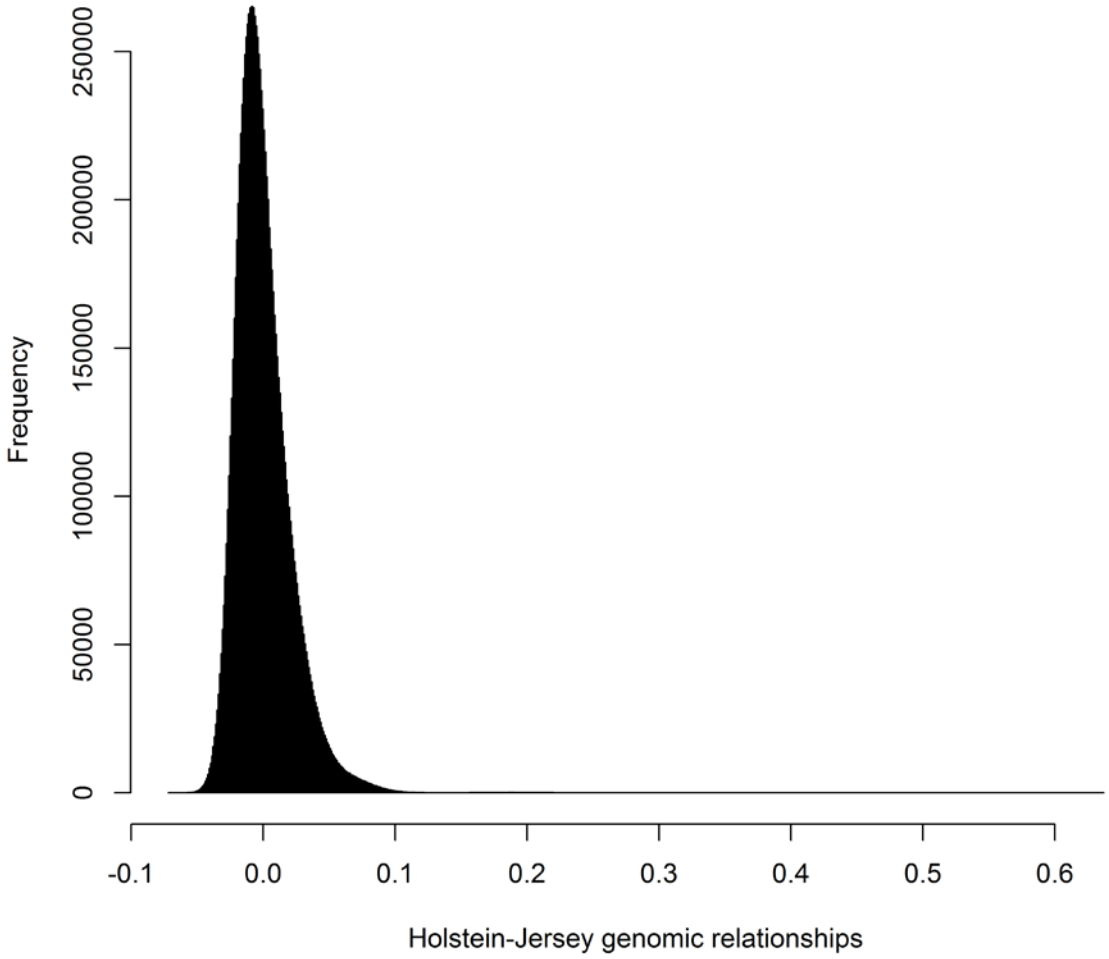
Figure 1. Distribution of genomic relationships between Holstein and Jersey animals.

Figure 2. Manhattan plot of the Bayes factors for milk yield obtained from five different analyses, including ST-BSSVS, using only Jersey (J), only Holstein (H), or H and J data, and multi-task BSSVS (mt) using H and J data, and either using the same π value (mt-BSSVS-1π) or different π values (mt-BSSVS-2π). For the last model, the Bayes factors for Jerseys are presented.

Figure 3. Manhattan plot of the Bayes factors for fat yield obtained from five different analyses, including single-trait BSSVS (ST-BSSVS) using only Jersey (J), only Holstein (H), or H and J data, and multi-task BSSVS (mt) using H and J data, and either using the same π value (mt-BSSVS-1π) or different π values (mt-BSSVS-2π). For the last model, the Bayes factors for Jerseys are presented.

Figure 4. Manhattan plot of the Bayes factors for protein yield obtained from five different analyses, including single-trait BSSVS (ST-BSSVS) using only Jersey (J), only Holstein (H), or H and J data, and multi-task BSSVS (mt) using H and J data, and either using the same π value (mt-BSSVS-1π) or different π values (mt-BSSVS-2π). For the last model, the Bayes factors for Jerseys are presented.

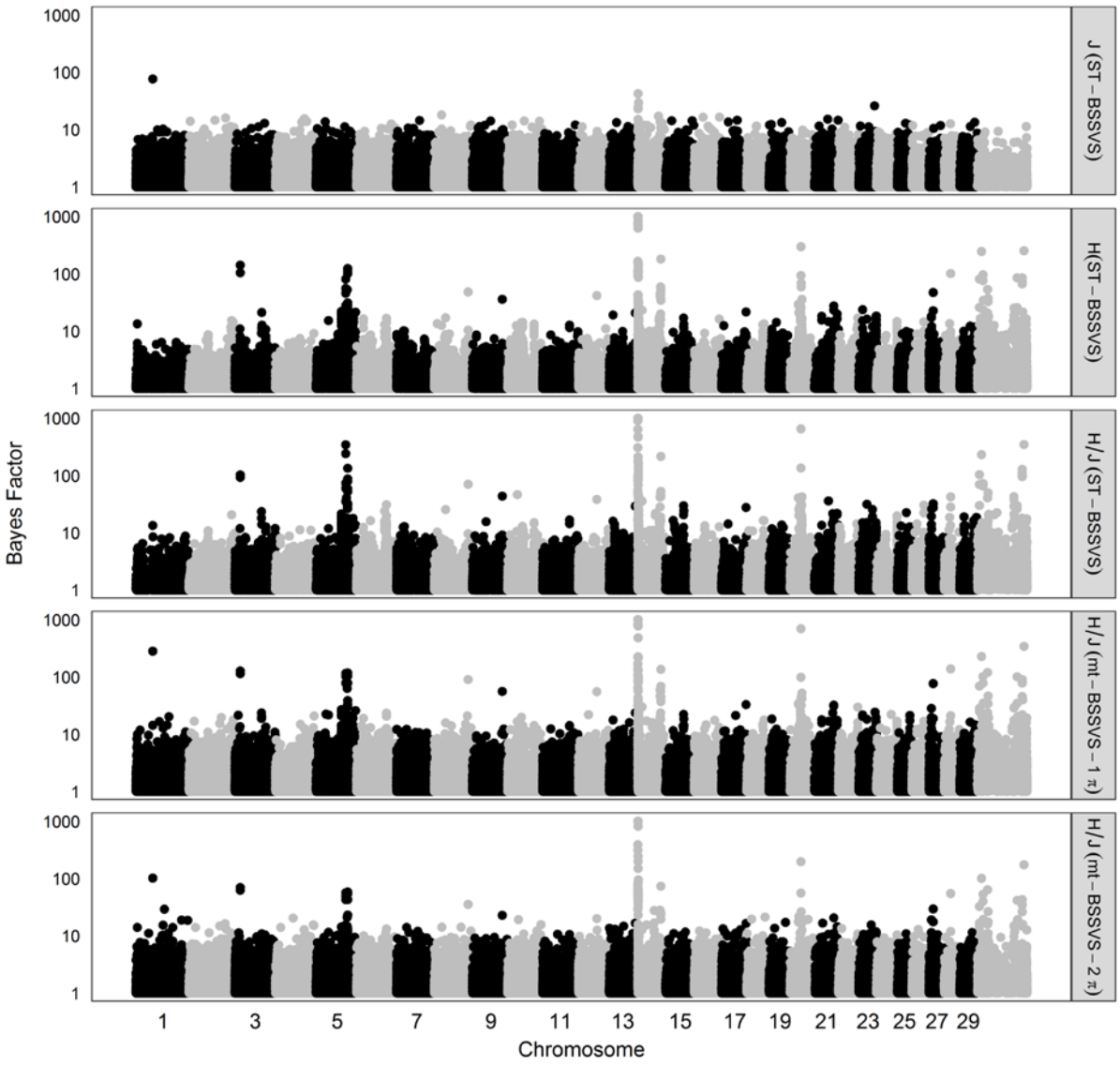
788 Figure 1



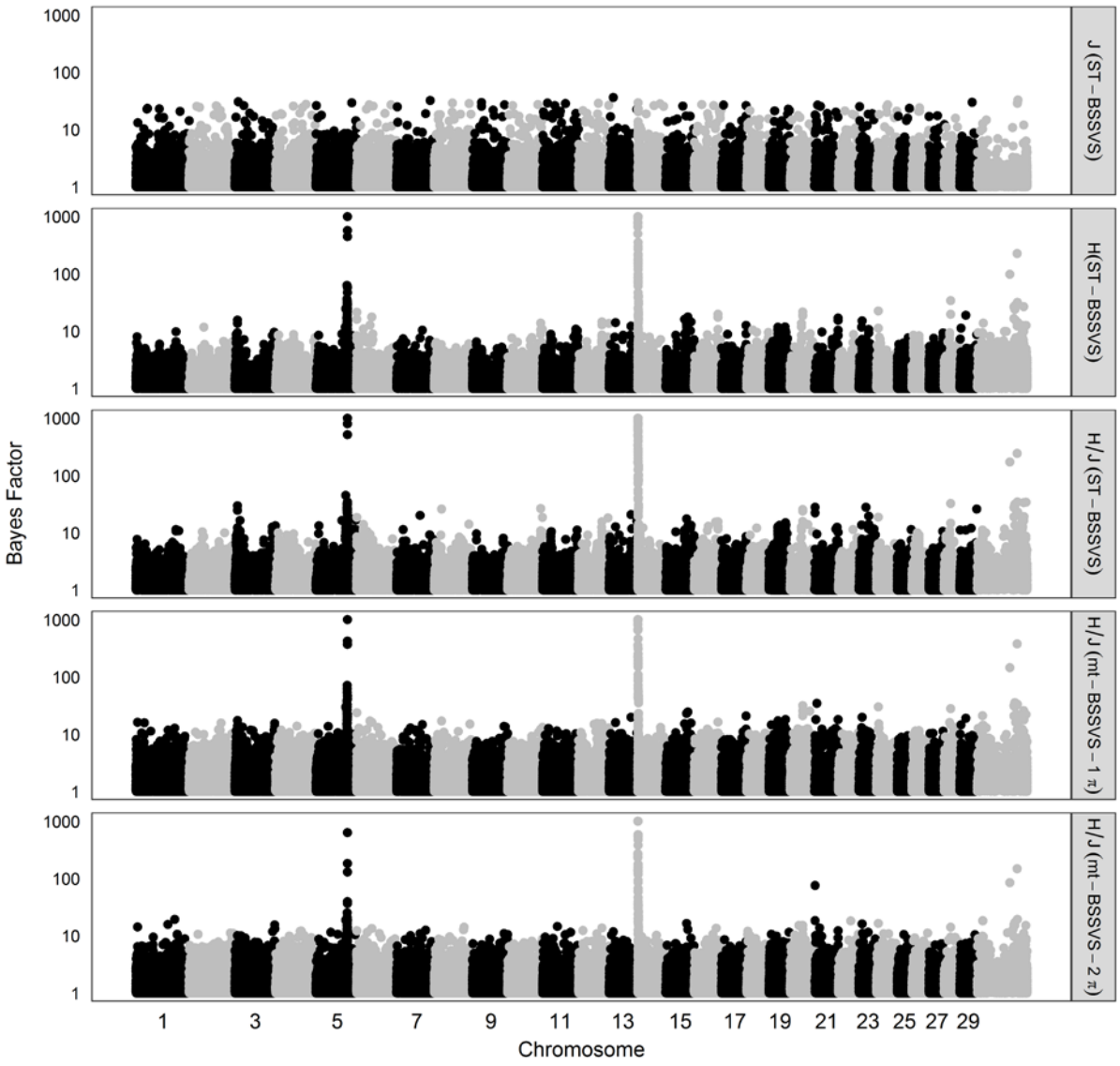
789

790

Figure 2


Bayes Factor
1000
100
10
1
J (ST – BSSVS)
H(ST – BSSVS)
H/J (ST – BSSVS)
H/J (mt – BSSVS – 1 π)
H/J (mt – BSSVS – 2 π)
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29
Chromosome

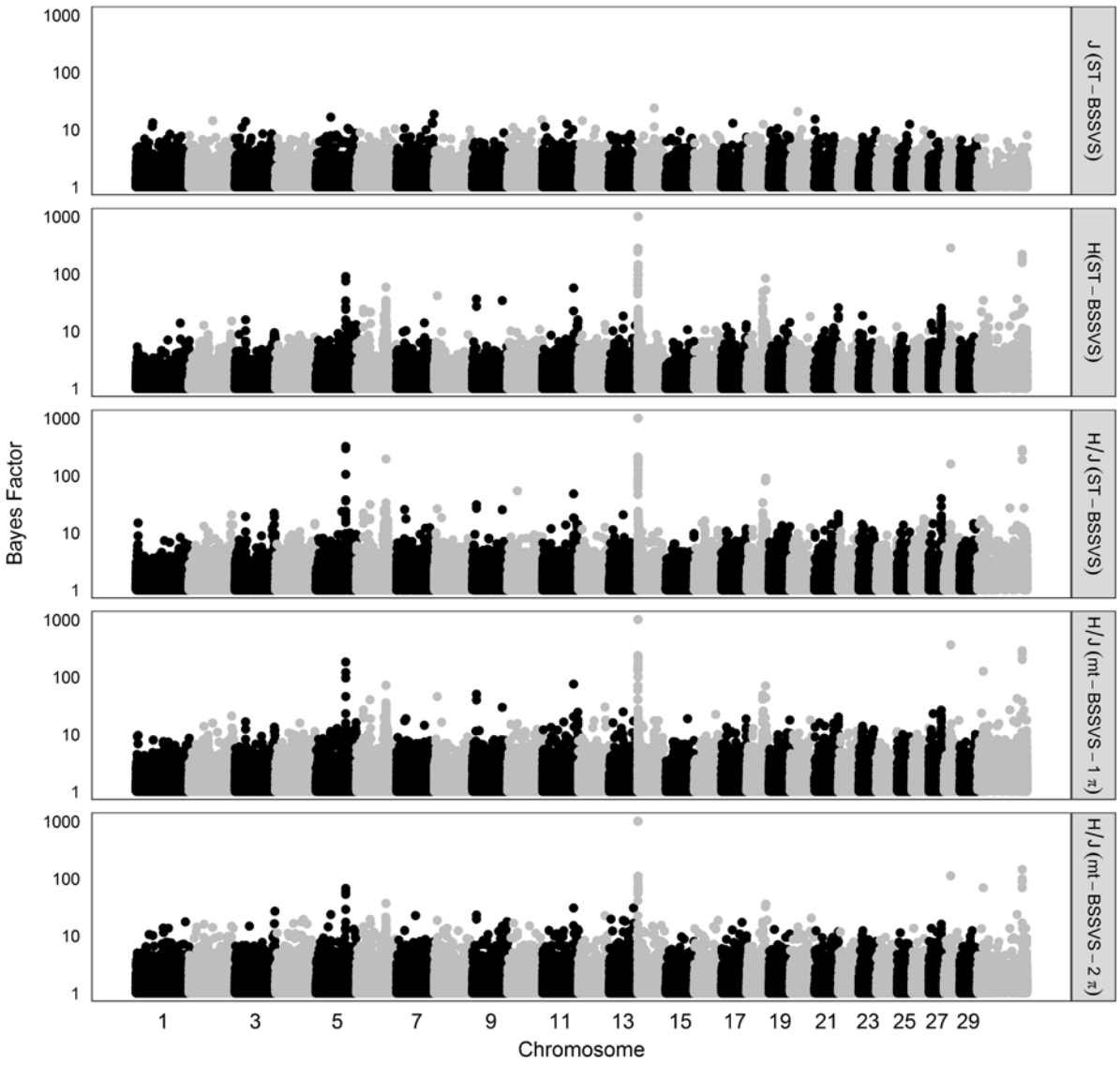
794 Figure 3



795

796

797 Figure 4



798

799