



# Genomic Prediction Using Bayesian Regression Models With Global–Local Prior

Shaolei Shi<sup>1</sup>, Xiujin Li<sup>2</sup>, Lingzhao Fang<sup>3</sup>, Aoxing Liu<sup>4</sup>, Guosheng Su<sup>4</sup>, Yi Zhang<sup>1</sup>, Basang Luobu<sup>5</sup>, Xiangdong Ding<sup>1\*</sup> and Shengli Zhang<sup>1\*</sup>

<sup>1</sup> National Engineering Laboratory for Animal Breeding, College of Animal Science and Technology, China Agricultural University, Beijing, China, <sup>2</sup> Guangdong Provincial Key Laboratory of Waterfowl Healthy Breeding, College of Animal Science and Technology, Zhongkai University of Agriculture and Engineering, Guangzhou, China, <sup>3</sup> Medical Research Council Human Genetics Unit, Institute of Genetics and Molecular Medicine, The University of Edinburgh, Edinburgh, United Kingdom, <sup>4</sup> Center for Quantitative Genetics and Genomics, Aarhus University, Tjele, Denmark, <sup>5</sup> Shannan Animal Husbandry and Veterinary Master Station, Shannan, China

## OPEN ACCESS

### Edited by:

Fabyano Fonseca Silva,  
Universidade Federal de Viçosa, Brazil

### Reviewed by:

Hao Cheng,  
University of California, Davis,  
United States  
Miriam Piles,  
Institute of Agrifood Research  
and Technology (IRTA), Spain  
Idalmo Pereira,  
Federal University of Minas Gerais,  
Brazil

### \*Correspondence:

Xiangdong Ding  
xding@cau.edu.cn  
Shengli Zhang  
zhangslcau@cau.edu.cn

### Specialty section:

This article was submitted to  
Livestock Genomics,  
a section of the journal  
Frontiers in Genetics

Received: 11 November 2020

Accepted: 02 March 2021

Published: 15 April 2021

### Citation:

Shi S, Li X, Fang L, Liu A, Su G,  
Zhang Y, Luobu B, Ding X and  
Zhang S (2021) Genomic Prediction  
Using Bayesian Regression Models  
With Global–Local Prior.  
Front. Genet. 12:628205.  
doi: 10.3389/fgene.2021.628205

Bayesian regression models are widely used in genomic prediction for various species. By introducing the global parameter  $\tau$ , which can shrink marker effects to zero, and the local parameter  $\lambda_k$ , which can allow markers with large effects to escape from the shrinkage, we developed two novel Bayesian models, named BayesHP and BayesHE. The BayesHP model uses Horseshoe+ prior, whereas the BayesHE model assumes local parameter  $\lambda_k$ , after a half-t distribution with an unknown degree of freedom. The performances of BayesHP and BayesHE models were compared with three classical prediction models, including GBLUP, BayesA, and BayesB, and BayesU, which also applied global–local prior (Horseshoe prior). To assess model performances for traits with various genetic architectures, simulated data and real data in cattle (milk production, health, and type traits) and mice (type and growth traits) were analyzed. The results of simulation data analysis indicated that models based on global–local priors, including BayesU, BayesHP, and BayesHE, performed better in traits with higher heritability and fewer quantitative trait locus. The results of real data analysis showed that BayesHE was optimal or suboptimal for all traits, whereas BayesHP was not superior to other classical models. For BayesHE, its flexibility to estimate hyperparameter automatically allows the model to be more adaptable to a wider range of traits. The BayesHP model, however, tended to be suitable for traits having major/large quantitative trait locus, given its nature of the “U” type-like shrinkage pattern. Our results suggested that auto-estimate the degree of freedom (e.g., BayesHE) would be a better choice other than increasing the local parameter layers (e.g., BayesHP). In this study, we introduced the global–local prior with unknown hyperparameter to Bayesian regression models for genomic prediction, which can trigger further investigations on model development.

**Keywords:** half-Cauchy, half-t distribution, Horseshoe+ prior, hyperparameter estimating, Horseshoe

## INTRODUCTION

The genomic prediction has been widely applied in animal and plant breeding. The statistical method being used is one of the critical factors for the accuracy of genomic estimated breeding value and consequently impacting the genetic gain achieved by genomic prediction. Models commonly used for predicting genomic estimated breeding value can be divided into two categories: (i) methods based on the framework of best linear unbiased prediction (BLUP), such as GBLUP (Habier et al., 2007; VanRaden, 2008); (ii) a set of Bayesian regression models – such as BayesA and BayesB (Meuwissen et al., 2001), which are also called as “Bayesian alphabet” models (Gianola et al., 2009). GBLUP assumes that the effects of all genetic markers are normally distributed and share the same variance, thus fitting well for traits with polygenic inheritance. In Bayesian methods, the marker effects are relevant to the choice of the prior probability distribution. By giving different priors, the Bayesian models can fit well for traits with different genetic architectures. For example, the widely used BayesA and BayesB models allow effects of genetic markers to follow a heavy-tailed distribution and therefore in line with the real distribution of marker effects for traits with large quantitative trait locus (QTL).

Bayesian regression models can be further classified into a one-group model and a two-group model from the perspective of the number of groups of genetic markers being used when estimating marker effects. The one-group model is generally a variable shrinkage model that shrinks the effects of some markers toward zero, such as BayesA (Meuwissen et al., 2001). The two-group model (or spike-and-slab model), can also be a multigroup model, is generally a variable selection model that only selects a subset of markers to be included in the model and assumes the remaining markers to have zero effects, such as BayesB (Meuwissen et al., 2001) and BayesC (Habier et al., 2011). The shrinkage process and the selection process can also be combined in a Bayesian regression model. In such a model, a subset of markers was selected to be included in the model, and then, the effects of some selected markers were further shrunk toward zero. BayesC $\pi$  (Habier et al., 2011) is an example of this type of model. Compared with BayesB and BayesC, BayesC $\pi$  can estimate the proportion of genetic markers with a non-zero effect based on the data. However, BayesC $\pi$  could be challenged by problems such as poor convergence and mixing in some situations (Wolc et al., 2011).

A Bayesian regression model with “global-local” shrinkage prior could be a good alternative for genomic prediction. With the global-local prior, the variances of marker effects can be shaped by global and local parameters simultaneously. The global parameter,  $\tau$ , can shrink the marker effects to approach zero, whereas the local parameter,  $\lambda_k$ , allows a marker to escape from the shrinkage when the marker has a big effect (Piironen and Vehtari, 2017). Horseshoe prior (Carvalho et al., 2010) is one of the most popular estimators of global-local prior. In Horseshoe prior, the local parameter follows a positive half-Cauchy distribution, which is a special case of half-t distribution where the degree of freedom is one. The Horseshoe prior has a similar form as the one-group model, but its prior can lead

to a “pseudo-posterior,” which shows the same pattern as the “two-group” model (Bhadra et al., 2017).

Horseshoe prior has already been applied to many scenarios, such as genomic prediction (Pong-Wong and Woolliams, 2014), genome-wide association study (Johndrow et al., 2017), and eQTL mapping (Li et al., 2019). Until now, BayesU (Pong-Wong and Woolliams, 2014) is the only model that uses the Horseshoe prior, and BayesU had similar performance with BayesA and BayesB tested with simulation data. However, the performance of BayesU has not yet been tested with real data. To better separate signals and noise, an extension of the Horseshoe estimator, named Horseshoe+ prior (Bhadra et al., 2017), was proposed. Horseshoe+ introduces one more local parameter with a positive half-Cauchy distribution, which leads a heavier tail than using standard Horseshoe prior. The investigation of using Horseshoe+ prior in Bayesian regression models for genomic prediction could be interesting but has not yet been explored previously. Besides, in variable shrinkage and selection models, hyperparameters is a challenge, and many previous studies have tried to estimate hyperparameter to improve prediction accuracy (Habier et al., 2011; Zhu et al., 2016).

This study’s objectives were to (i) develop two Bayesian methods for genomic prediction based on the global-local prior, which have the flexibility in estimating hyperparameters, and (ii) to test the model performance with simulated and real data for traits with various genetic architectures.

## MATERIALS AND METHODS

### Statistical Models

In Bayesian regression models, the differences in different models were the prior assumptions on the effects of single-nucleotide polymorphisms (SNPs). All Bayesian multiple regression model can be described as follows:

$$y = \mu + \sum_{k=1}^m x_k \beta_k + e,$$

where  $y$  is the vector of pre-corrected phenotypes,  $\mu$  is the overall mean,  $x_k$  is the vector of genotypes for the  $k$ th SNP,  $m$  is the number of SNPs,  $\beta_k$  is the effect of the  $k$ th SNP, and  $e$  is a vector of random residuals. The assumptions of the residuals are  $e \sim N(0, D\sigma_e^2)$ , where  $\sigma_e^2$  is the random residual variance. The  $D$  is an identity matrix when using pre-corrected phenotypes other than de-regressed proofs (DRPs). DRPs were derived from an official estimated breeding value (EBV) with the method that Jairath et al. (1998) suggested. When using DRP as  $y$ ,  $D$  is a diagonal matrix with diagonal elements calculated as  $d_{ii} = \frac{1-r_i^2}{r_i^2}$ , to account for heterogeneities in  $\sigma_e^2$  due to differences in reliability ( $r_i^2$ ) of DRP.

In Bayesian inference, a total of 50,000 Markov chain Monte Carlo samples were generated, with the first 20,000 samples discarded as burn-in and every 50th sample of the remaining 30,000 samples saved for inferring posterior statistics. All analyses with Bayesian regression models were conducted using in-house scripts written in Fortran 95 by the first author.

### BayesU

The BayesU (Pong-Wong and Woolliams, 2014) was developed based on Horseshoe prior (HS). To make it comparable with other methods, the global prior  $\tau$  was set as a flat prior. A detailed description is given as:

$$\beta_k \sim N(0, \lambda_k^2 \tau^2), \lambda_k \sim C^+(0, 1), \text{ and } \tau \sim \text{flat}$$

where  $\lambda_k$  and  $\tau$  are local and global parameters, respectively. The local parameter  $\lambda_k$  follows a positive half-Cauchy distribution, which is a special case of student-t distribution where the degree of freedom equals to 1.

### BayesHP

Compared with HS, its modified version Horseshoe+ (HS+) can form a heavier tailed prior distribution by introducing an additional local parameter with a positive half-Cauchy distribution. Regarding the performance, HS+ can better distinguish the signals and noise than the standard HS (Bhadra et al., 2017). However, Horseshoe+ prior (HS+) (Bhadra et al., 2017) has not yet been applied in Bayesian regression models for genomic prediction. In this study, we proposed a novel BayesHP model based on HS+:

$$\beta_k \sim N(0, \lambda_k^2 \tau^2), \lambda_k \sim C^+(0, \eta_k), \eta_k \sim C^+(0, 1), \\ \text{and } \tau \sim C^+(0, N^{-1})$$

where  $\lambda_k$  and  $\eta_k$  are local parameters, and  $\tau$  is a global parameter following a positive half-Cauchy distribution with scale parameter equals to  $N^{-1}$ . The  $N$  is the size of training data, as Bhadra et al. (2017) suggested. Compared with HS, HS+ introduces one more layer of local parameter  $\eta_k$ . As described by Makalic and Schmidt (2016) and Makalic et al. (2016) (Appendix), the half-Cauchy distribution can be modeled as a scale mixture of inverse gamma distributions: if  $x^2 \sim IG(\frac{1}{2}, \frac{1}{a})$  and  $a \sim IG(\frac{1}{2}, \frac{1}{A^2})$ , then  $x \sim C^+(0, A)$ . Finally, the distribution of parameters for the revised Horseshoe+ hierarchy is as follows:

$$\beta_k \sim N(0, \lambda_k^2 \tau^2), \lambda_k^2 \sim IG\left(\frac{1}{2}, \frac{1}{\theta_k}\right), \theta_k \sim IG\left(\frac{1}{2}, \frac{1}{\eta_k^2}\right), \eta_k^2 \\ \sim IG\left(\frac{1}{2}, \frac{1}{\nu_k}\right), \nu_k \sim IG\left(\frac{1}{2}, 1\right), \tau^2 \sim IG\left(\frac{1}{2}, \frac{1}{\xi}\right), \\ \text{and } \xi \sim IG\left(\frac{1}{2}, N^2\right).$$

The conditional posterior distributions of  $\lambda_k$  and  $\eta_k$  parameters are inverse-gamma distributions, which makes Gibbs sampling straightforward.

### BayesHE

In BayesU and BayesHP, the prior of local parameter  $\lambda_k$  followed a positive half-Cauchy distribution. Both of these two models used a fixed value as the degree of freedom. To increase the flexibility and the suitability of the prediction model, we

proposed a new model, named BayesHE, which assumed the local parameter  $\lambda_k$  to follow a half-t distribution with an unknown degree of freedom  $\nu$ :

$$\beta_k \sim N(0, \lambda_k^2 \tau^2), \lambda_k \sim \text{half-t}^+(\nu, 1), \tau \sim C^+(0, N^{-1}), \\ \text{and } \nu \sim G(a, b)$$

By introducing auxiliary variables (Wand et al., 2011), the revised hierarchy is as follows:

$$\beta_k \sim N(0, \lambda_k^2 \tau^2), \lambda_k^2 \sim IG\left(\frac{\nu}{2}, \frac{\nu}{\theta_k}\right), \theta_k \sim IG\left(\frac{1}{2}, 1\right), \\ \tau^2 \sim IG\left(\frac{1}{2}, \frac{1}{\xi}\right), \xi \sim IG\left(\frac{1}{2}, N^2\right), \text{ and } \nu \sim G(a, b).$$

All parameters, including  $\beta_k, \lambda_k^2, \theta_k, \tau^2$ , and  $\xi$ , had a standard form, except  $\nu$ . The full conditional distribution of the hyperparameter  $\nu$  is described as follows:

$$f(\nu | \cdot) \propto f(\lambda_k^2 | \nu) * f(\nu) \propto \prod_{k=1}^m \frac{\left(\frac{\nu}{\theta_k}\right)^{\left(\frac{\nu}{2}\right)}}{\Gamma\left(\frac{\nu}{2}\right)} \lambda_k^2 - \left(\frac{\nu}{2} + 1\right) \exp \\ \left(-\frac{\nu}{\lambda_k^2}\right) * \nu^{(a-1)} \exp(-b\nu) \propto \nu^{\left(\frac{\nu * m}{2} + a - 1\right)} * \Gamma\left(\frac{\nu}{2}\right)^{-m} \\ * \exp\left(-\nu \left(\frac{1}{2} \sum_{k=1}^m \ln(\theta_k \lambda_k^2) + \sum_{k=1}^m \frac{1}{\theta_k \lambda_k^2} + b\right)\right)$$

where  $m$  is the number of SNPs,  $\Gamma(\cdot)$  is the gamma function,  $a$  is the shape parameter in gamma distribution, and  $b$  is the scale parameter of the gamma distribution for  $\nu$ . In this study, we compared two models with the same  $b$  ( $b$  equals to 1) but with different  $a$ , including BayesHE1 with  $a$  equals to 4 and BayesHE2 with  $a$  equals to 5. The hyperparameter was inferred by applying a univariate Metropolis-Hastings sampling (DFMH) process (Yang et al., 2015). The random walk M-H step worked with  $\zeta = \log(\nu)$  because  $\nu$  was inherently positive. The corresponding full conditional distribution of  $\zeta$  is as follows:

$$f(\zeta | ELSE) \propto \exp(\zeta)^{\left(\frac{\exp(\zeta) * m}{2} + a - 1\right)} * \Gamma\left(\frac{\exp(\zeta)}{2}\right)^{-m} \\ * \exp\left(-\exp(\zeta) \left[\frac{1}{2} \sum_{k=1}^m \ln(\theta_k \lambda_k^2) + \sum_{k=1}^m \frac{1}{\theta_k \lambda_k^2} + b\right]\right) * \exp(\zeta) \\ \propto \exp(\zeta)^{\frac{\exp(\zeta) * m + 2a}{2}} * \Gamma\left(\frac{\exp(\zeta)}{2}\right)^{-m} * \\ \exp\left(-\exp(\zeta) \left[\frac{1}{2} \sum_{k=1}^m \ln(\theta_k \lambda_k^2) + \sum_{k=1}^m \frac{1}{\theta_k \lambda_k^2} + b\right]\right)$$

where  $\exp(\zeta)$  is the Jacobian from  $\nu$  to  $\zeta$ .

The performance of three Bayesian regression models applying global-local priors, including BayesU, BayesHP, and BayesHE, were further compared with three widely used genomic prediction models, including GBLUP, BayesA, and BayesB.

## BayesA/BayesB

In BayesB, the prior distribution of  $\beta_k$  is as follows:

$$\beta_k | S_{\beta}^2, \upsilon, \pi \sim \text{IID} \begin{cases} 0 & \text{with probability } \pi \\ t(0, S_{\beta}^2, \upsilon \pi) & \text{with probability } 1 - \pi \end{cases}$$

The BayesA model can be considered as a specific case of BayesB, where  $\pi = 0$ . In this study, we set  $\pi = 0.95$  for BayesB.

## GBLUP

The GBLUP model is described as follows:

$$y = \mu + Zg + e$$

where  $y$  is the vector of pre-corrected phenotypes,  $\mu$  is the overall mean,  $Z$  is the design matrix linking genetic value ( $g$ ) to  $y$ , and  $e$  is a vector of random residuals. It was assumed that,

$$g \sim N(0, G\sigma_g^2) \text{ and } e \sim N(0, D\sigma_e^2)$$

where  $\sigma_g^2$  is the additive genetic variance and  $\sigma_e^2$  is the random residual variance. The genomic relationship matrix ( $G$ ) (VanRaden, 2008) was calculated with SNPs:

$$G = \frac{(M-P)(M-P)'}{2 \sum_{k=1}^m p_k(1-p_k)}$$

where  $M$  is a  $n \times m$  matrix with  $n$  for the number of individuals and  $m$  for the number of SNPs,  $p_k$  is the MAF of  $i$ th SNP, and  $P$  is the matrix in which the  $k$ th column elements are  $2p_k$ . In this study, GBLUP was implemented using the DMU software (Madsen and Jensen, 2012).

## Datasets

### Quantitative Trait Locus-Marker-Assisted Selection Data

We used the simulated data from the 15th QTL-marker-assisted selection (MAS) workshop (Elsen et al., 2012) to test model performances. The founder animals consist of 20 sires and 200 dams. For each generation, one sire was mated to 10 dams, and each dam produced 15 offspring. Eight QTLs were simulated across the five chromosomes, with one QTL being quadri-allelic, two linked in phase, two linked in repulsion, one imprinting, and two epistatic. Random residual effects were added to achieve a realized heritability of 0.3. After removing loci without polymorphisms, 7,121 SNPs were retained for analysis. Details on the simulated dataset are in Li et al. (2018).

In each full-sib family, 10 individuals had marker genotypes and phenotype, and the remaining five individuals only had marker genotypes. In total, 2,000 individuals had both genotype and phenotype information, and 1,000 individuals only had genotype information. In this study, only 2,000 individuals with genotypes and phenotypes were used for cross-validation.

### Cattle Data

For real data analysis in dairy cattle, we collected phenotypic and genomic data from Chinese Holsteins. In total, 7,052 individuals were available for analyses on three milk production traits,

including milk yield (MY), fat yield (FY), and protein yield (PY), and on one health traits (somatic cell score, SCS), and 3,530 individuals were available for three type traits including conformation (CONF), feet lag (FL), and mammary system (MS). DRP derived from the official EBV were used as pseudo-phenotypes for genomic prediction. The reliability of DRP for each individual was estimated as  $r_{DRP}^2 = ERC_i / (ERC_i + \lambda)$ , with  $\lambda = \frac{1-h^2}{h^2}$ , where  $ERC_i$  refers to the effective record contribution and  $h^2$  refers to the estimated heritability of the trait. On note, effective record contribution ( $ERC_i$ ) was relevant to the reliability ( $REL_i$ ) of the EBV of animal  $i$  (Přibyl et al., 2013),  $ERC_i = \lambda * REL / (1 - REL)$ . Animals were genotyped by the Illumina 50K chip. Missing genotypes were imputed with Beagle version 3 (Browning and Browning, 2011). We further removed the SNPs with a minor allele frequency below 0.01 and significantly deviated from Hardy-Weinberg equilibrium ( $p < 10^{-6}$ ) and the individuals with call rates lower than 0.90. After quality control, 43,447 SNPs remained for subsequent analyses.

### Mice Data

For real data analysis in mice, we used the heterogeneous stock mice dataset generated by the Wellcome Trust Centre for Human Genetics<sup>1</sup>. As described by Legarra et al. (2008), the extent of linkage disequilibrium in this population is strong, with an average  $r_{LD}^2$  among adjacent SNPs being 0.62. To compare the performance of different methods, we selected three traits: growth rate between 6 and 10 weeks of age (GSL), body mass index (BMI), and body length (BL). There were 1,821, 1,814, and 1,901 individuals available for analysis on BL, BMI, and GSL, respectively. In total, 9,098 SNPs were available. A detailed description of the population can be found in Li et al. (2018).

## Cross-Validation and Prediction Accuracy

To assess the prediction accuracy, a  $5 \times 6$  cross-validation (six-fold cross-validation repeated five times) procedure was used, and the results are shown as the mean and standard error for replicates. The performances of all methods were evaluated by examining the accuracy of direct genomic value (DGV) in test data. For QTL-MAS and mice data, Pearson correlation of DGV and phenotype/pre-corrected phenotype was used; for cattle data, the prediction accuracy was further corrected by the average accuracy (square root of reliability) of DRP in test data:

$$acc = \frac{cor(DRP, DGV)}{\bar{r}}$$

where  $cor(DRP, DGV)$  is the Pearson correlation of DRP and DGV of the validation data, and  $\bar{r}$  is the average of the square root of the testing data DRP reliabilities.

In addition, the regression of DRP on phenotype,  $y$ , was used to evaluate the unbiasedness of prediction for all three datasets. The closer the regression coefficient to one, the more unbiased the prediction result.

<sup>1</sup><http://gscan.well.ox.ac.uk/>



## RESULTS

In this study, the performance of our newly proposed Bayesian models with global-local priors was compared with GBLUP, BayesA, BayesB ( $\pi = 0.95$ ), and BayesU, using the simulated data generated by QTL-MAS, and real data in cattle and mice.

To assess the convergence of Markov chain Monte Carlo, trace plots of the overall mean ( $\mu$ ) and additive variance [ $V_g = \text{var}(\sum_{k=1}^m x_k \beta_k)$ ] are shown in **Figure 1**. Also, the trace plots suggested that parameters mixed well. However, additive variance from BayesHE (**Figures 1E,F**) converges faster than BayesHP (**Figure 1D**).

### Quantitative Trait Locus-Marker-Assisted Selection Data

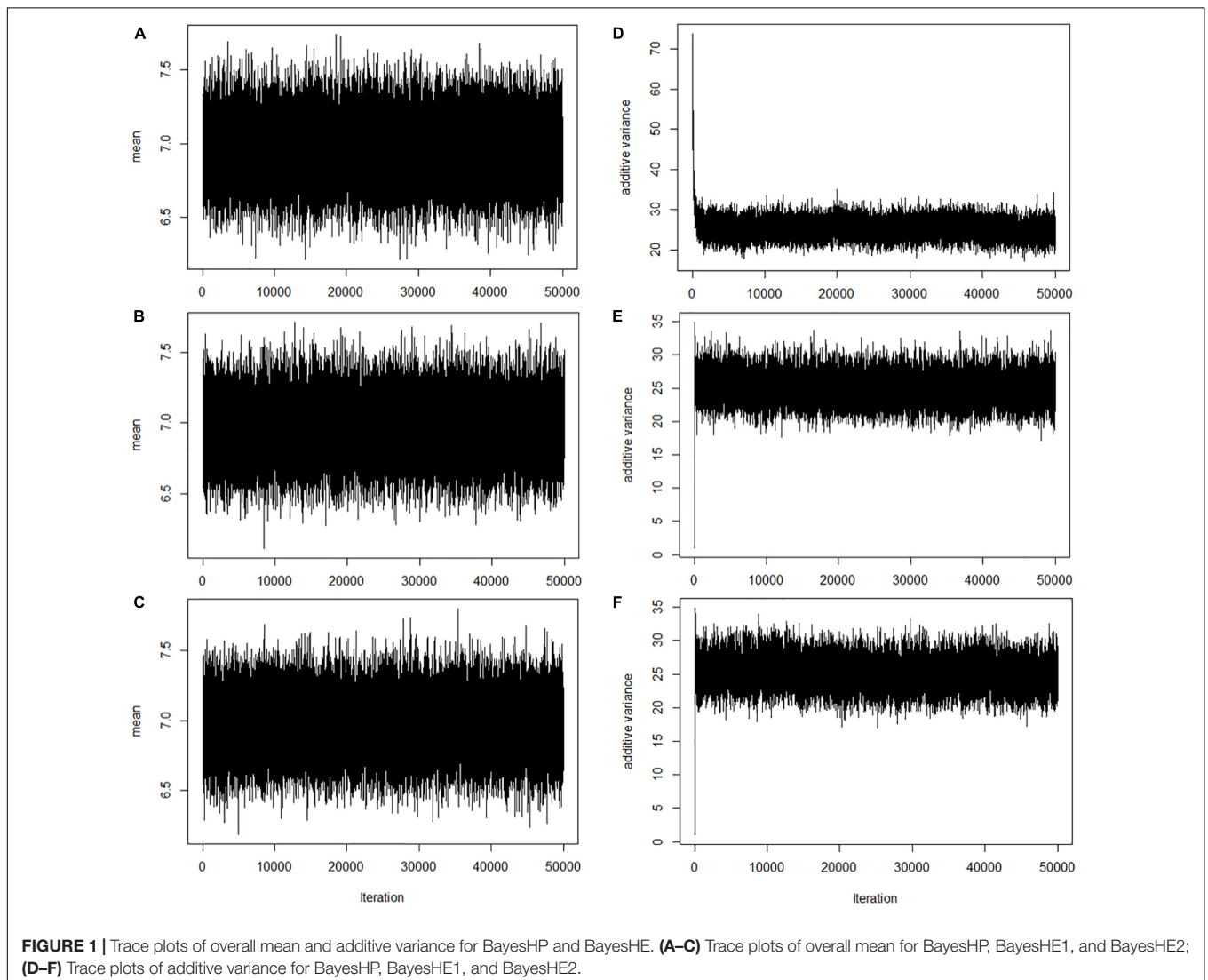
**Table 1** shows the prediction accuracies and bias for all models based on the 15th QTL-MAS workshop dataset. Regarding the

prediction accuracy, Bayesian regression models with global-local priors, such as BayesU (0.506), BayesHP (0.505), BayesHE1 (0.505), and BayesHE2 (0.505), outperformed all other methods. The prediction biases of the seven methods were similar and close to one.

### Cattle Data

The prediction accuracies of seven traits in the Chinese Holstein population that the mean  $r_{LD}^2$  of adjacent SNP pairs ranged from 0.16 to 0.24 (Zhou et al., 2013) are shown in **Table 2**. Generally, BayesHE with two modalities (e.g., BayesHE1 and BayesHE2) on hyperparameters achieved optimal or suboptimal prediction accuracy for all of the seven traits.

For milk production traits, Bayesian regression models with global-local priors had a better performance compared with GBLUP, BayesA, or BayesB, especially for MY. For example, the prediction accuracy of BayesHE1 was 0.473, which increased approximately 2.2% than GBLUP. Also, BayesHE1 had similar prediction accuracy than BayesHE2. However,



**TABLE 1** | Prediction accuracies and biases of DGVs of test dataset from 15th QTL-MAS data using six-fold cross-validation with five replications.

	Accuracy	Bias
GBLUP	0.456±0.002	1.010±0.004
BayesA	0.474±0.009	0.924±0.008
BayesB	0.475±0.013	0.925±0.004
BayesU	0.506±0.002	0.996±0.004
BayesHP	0.505±0.002	1.000±0.005
BayesHE1	0.505±0.002	1.005±0.006
BayesHE2	0.505±0.002	1.003±0.005

DGVs, direct genomic value; and the mean and standard errors are shown in the table.

for SCS, BayesHP achieved the lowest prediction accuracy (0.341), and BayesHE and BayesA had similar prediction accuracy (0.365).

For type traits, BayesB, BayesU, and BayesHP did not perform well, and GBLUP, BayesA, and BayesHE had similar prediction accuracy. Notably, BayesHE2 performed similarly to BayesHE1. Although BayesHE did not perform the best, the prediction accuracy was very close to that of the best model. For example, the prediction accuracy for MS from BayesHE2 was 0.427, which was only slightly lower than the highest prediction accuracy achieved by GBLUP (0.428).

The biases of prediction for the seven traits are shown in **Table 3**. For MY, FY, PY, and SCS, BayesHE1 achieved the least bias of prediction. The performance of BayesHE2, regarding bias,

was very close to that of BayesHE1. For FL, BayesA was the most unbiased model, and BayesHE2 was the second best.

## Mice Data

The prediction accuracies of three mice traits with different methods are shown in **Table 4**. In the analysis of mice data, there were two kinds of traits: one growth trait (GSL) and two type traits (BL and BMI). For all three traits, BayesHP performed the worst. For example, the prediction accuracy of BL from BayesHP was 0.253, but accuracies from other methods were greater than 0.260. However, GBLUP, BayesA, and BayesHE had similar prediction accuracies.

**Table 5** shows the prediction bias. The regression coefficients were close to the unity for all traits using all models, which indicated unbiasedness of the predictions. Nevertheless, there were still some slight differences. For example, the unbiasedness of BayesB was slightly lower than other models for all traits.

## DISCUSSION

In Bayesian regression models, the differences among methods are the assumptions on the genetic marker effects. Because of the flexibility of the Bayesian method, it has attracted increasing attention. In classical one-group models, both signals and noises were assumed to follow one single continuous prior distribution, where the effects of some markers were shrunk toward zero, relying on the posterior distribution. For the two-group model or the spike-and-slab model, the

**TABLE 2** | Prediction accuracies of seven traits of dairy cattle using six-fold cross-validation with five replications.

	MY	FY	PY	SCS	CONF	FL	MS
GBLUP	0.451±0.002	0.410±0.002	0.435±0.001	0.356±0.002	0.480±0.003	0.676±0.004	0.428±0.006
BayesA	0.467±0.002	0.425±0.002	0.433±0.001	0.365±0.002	0.478±0.003	0.677±0.004	0.425±0.006
BayesB	0.455±0.003	0.401±0.002	0.421±0.002	0.345±0.002	0.380±0.037	0.656±0.005	0.399±0.006
BayesU	0.463±0.003	0.415±0.002	0.420±0.002	0.346±0.003	0.447±0.007	0.664±0.007	0.404±0.008
BayesHP	0.459±0.003	0.410±0.002	0.414±0.003	0.341±0.003	0.440±0.009	0.660±0.007	0.401±0.008
BayesHE1	0.473±0.003	0.427±0.002	0.435±0.001	0.365±0.002	0.478±0.003	0.674±0.004	0.426±0.006
BayesHE2	0.473±0.003	0.427±0.002	0.434±0.001	0.365±0.002	0.478±0.003	0.674±0.004	0.427±0.005

MY, milk yield; FY, fat yield; PY, protein yield; SCS, somatic cell score, CONF, conformation; FL, feet lag; MS, mammary system; and the mean and standard errors are shown in the table.

**TABLE 3** | Prediction biases of seven traits of dairy cattle data using six-fold cross-validation with five replications.

	MY	FY	PY	SCS	CONF	FL	MS
GBLUP	0.865±0.004	0.817±0.003	0.814±0.002	0.807±0.006	0.803±0.006	0.829±0.007	0.826±0.014
BayesA	0.877±0.005	0.807±0.003	0.807±0.002	0.812±0.005	0.798±0.006	0.837±0.007	0.809±0.014
BayesB	0.824±0.006	0.755±0.003	0.750±0.004	0.756±0.006	0.763±0.021	0.775±0.006	0.740±0.014
BayesU	0.887±0.007	0.828±0.004	0.816±0.007	0.812±0.008	0.749±0.019	0.812±0.011	0.786±0.020
BayesHP	0.889±0.007	0.829±0.005	0.816±0.007	0.815±0.007	0.735±0.022	0.805±0.011	0.772±0.019
BayesHE1	0.906±0.007	0.835±0.003	0.821±0.002	0.821±0.006	0.805±0.007	0.833±0.007	0.831±0.015
BayesHE2	0.905±0.008	0.834±0.003	0.819±0.002	0.820±0.006	0.807±0.007	0.834±0.007	0.833±0.014

MY, milk yield; FY, fat yield; PY, protein yield; SCS, somatic cell score, CONF, conformation; FL, feet lag; MS, mammary system; and the mean and standard errors are shown in the table.

**TABLE 4** | Prediction accuracies of mice data using six-fold cross-validation with five replications.

	BL	BMI	GSL
GBLUP	0.272±0.002	0.226±0.002	0.386±0.003
BayesA	0.275±0.002	0.227±0.002	0.385±0.003
BayesB	0.268±0.001	0.217±0.002	0.374±0.003
BayesU	0.261±0.002	0.220±0.003	0.374±0.003
BayesHP	0.253±0.003	0.214±0.003	0.368±0.004
BayesHE1	0.274±0.002	0.229±0.002	0.386±0.003
BayesHE2	0.272±0.001	0.227±0.002	0.386±0.003

BL, body length; BMI, body mass index; GSL, growth slope; and the mean and standard errors are shown in the table.

**TABLE 5** | Prediction biases of mice data using six-fold cross-validation with five replications.

	BL	BMI	GSL
GBLUP	0.988±0.012	1.023±0.022	1.004±0.010
BayesA	0.995±0.009	0.988±0.015	0.988±0.008
BayesB	0.948±0.004	0.929±0.017	0.959±0.007
BayesU	0.972±0.007	0.999±0.036	0.996±0.012
BayesHP	0.981±0.011	1.032±0.039	1.002±0.015
BayesHE1	1.006±0.011	1.024±0.021	1.004±0.009
BayesHE2	1.003±0.009	1.016±0.021	1.005±0.008

BL, body length; BMI, body mass index; GSL, growth slope; and the mean and standard errors are shown in the table.

prior regarding the proportion of genetic markers being signal usually impact its performance in genomic prediction. Some two-group models, such as BayesC $\pi$  (Habier et al., 2011), have been developed to estimate the proportion of non-zero effect markers based on both prior and the analyzed data. However, there is a poor convergence and mixing in some situations. The global-local prior, which can shrink signals and noises through local and global parameters, seems to be a good alternative, theoretically. Global-local priors is a kind of continuous shrinkage prior, which can adaptively shrink noise to zero while leaving the large data-supported signal unshrunk (Ge et al., 2019).

The model's performance depended on the genetic architecture of the trait. The results of simulation indicated that models based on global-local priors, e.g., BayesU, BayesHP, and BayesHE, performed better in traits with higher heritability (i.e., in this study, heritability is 0.3) and fewer QTL. In real data, BayesHE can achieve optimal or suboptimal performance; however, BayesHP performed better only for production traits. Our results suggested that auto-estimate the degree of freedom (e.g., BayesHE) would be a better choice other than increasing the layers of the local parameter (e.g., BayesHP).

The Bayesian models with the assumption more in line with the real distribution of marker effects will result in more accurate predictions. The Bayesian model shrinks the effect of noise markers toward zero and thus increases the prediction accuracy. However, in genomic prediction, markers are not simply signal or noise due to the existence of linkage

disequilibrium. It is reasonable that for some traits, GBLUP will achieve better prediction accuracy. For example, GBLUP performed better for type traits (e.g., CONF, FL, and MS) than BayesB, BayesU, and BayesHP, as shown in our results. Notably, in genome-wide association study, regardless of dairy cattle (Wu et al., 2013) or beef cattle (Vallée et al., 2016), there are few significant signals for type traits, suggesting that most genetic variants have similar medium or small effects on the traits. Therefore, it is reasonable why GBLUP had a better performance for type traits.

Many previous studies have suggested that using hyperparameters is likely to improve classical methods (Habier et al., 2011; Yang and Tempelman, 2012; Zhu et al., 2016). In our study, we assumed that the local parameter,  $\lambda_k$ , followed a half-t distribution (BayesHE) with an unknown degree of freedom instead of half-Cauchy (BayesU). By introducing auxiliary variables (Wand et al., 2011), half-t distribution was translated into a scale mixture of the inverse gamma distribution. In BayesHE,  $\lambda_k^2$  was assumed to follow an inverse gamma distribution  $IG\left(\frac{\nu}{2}, \frac{\nu}{\theta_k}\right)$ , which led to an assumption of student-t distribution for marker effects (Wand et al., 2011). The use of unknown shape parameter is similar to the study of Zhu et al. (2016), but the difference is that there is a global parameter  $\tau^2$  in BayesHE model. Besides, in studies of Habier et al. (2011) and Zhu et al. (2016), they set a gamma distribution  $G(1, 1)$  for the scale parameter. In our study, the scale parameter  $\theta_k$  was assumed to follow an inverse gamma distribution,  $\theta_k \sim IG\left(\frac{1}{2}, 1\right)$ . This inspired the authors that the shape parameter of inverse gamma distribution that  $\theta_k$  followed can also be set as a variable other than a constant.

Horseshoe-like prior with "U" type shrinkage pattern means strong distinguishment of single and noise. According to the results of QTL detection (Wu et al., 2013; Vallée et al., 2016), Horseshoe-like prior with "U" type may be suitable for genomic prediction of the traits affected by many QTLs with large effect. In our study, we assumed an unknown hyperparameter for the distribution of local parameters, which increased the model flexibility and, therefore, more adaptable to traits with different genetic architectures. The possibility to fit a suitable hyperparameter for the global parameter has been proposed by Armagan et al. (2011), where they assumed global parameter followed a gamma distribution with different shape parameter or just set as a constant value. Their study suggested that the changes of hyperparameters of distributions that local parameters followed and the value of global parameter led to different shrinkage patterns on covariates. In the study of Piironen and Vehtari (2016), the global parameter  $\tau$  was set as a constant value or followed a normal or half-Cauchy distribution, and they recommended  $\tau$  half-Cauchy distribution,  $\tau \sim C^+(0, \tau_0^2)$ , where the scale parameter  $\tau_0^2$  is relevant to the effective number of variables with non-zero effects. In the global-local prior method, the marker variances were shaped by global and local parameters simultaneously. The global parameter,  $\tau$ , usually causes the marker effect to approach zero, whereas the local parameter,  $\lambda_k$ , allows marker variance to escape the shrinkage

when that marker has a large effect. In future research, more investigation on choosing the type of distribution for global parameters could be interesting.

The limitation of our study is that it mostly focused on statistical perspectives and lack of consideration of the biological information. With time, an increasing amount of biological information affecting complex traits will be detected. It is reasonable to integrate these genomic features into the prediction model, and then, how to effectively utilize these genomic features is worth exploring. The BayesRC model proposed by MacLeod et al. (2016) divides the genome into three major categories: trait-associated genes, regular regions, and other variations. However, there are some challenges in utilizing biological information because of the dynamics in biological processes.

## CONCLUSION

Our results showed that BayesHE could achieve optimal or suboptimal performance. Compared with other methods, such as GBLUP and BayesA, BayesHP did not perform better. With the automatic estimation of hyperparameters, BayesHE was more flexible than BayesU and BayesHP for the adaptation to a wider range of traits. This suggested that auto-estimate the degree of freedom (e.g., BayesHE) would be a better choice other than increasing the layers of a local parameter (e.g., BayesHP).

## DATA AVAILABILITY STATEMENT

The data analyzed in this study is subject to the following licenses/restrictions: All information supporting the results is included in the text and tables. The dataset of cattle is not publicly available due to commercial restrictions. Requests to access these datasets should be directed to the corresponding author (SZ).

## REFERENCES

- Armagan, A., Dunson, D. B., and Clyde, M. (2011). Generalized beta mixtures of gaussians. *Adv. Neural. Inf. Process. Syst.* 24, 523–531.
- Bhadra, A., Datta, J., Polson, N. G., and Willard, B. (2017). The horseshoe+ estimator of ultra-sparse signals. *Bayesian Anal.* 12, 1105–1131. doi: 10.1214/16-ba1028
- Browning, B. L., and Browning, S. R. (2011). A fast, powerful method for detecting identity by descent. *Am. J. Hum. Genet.* 88, 173–182. doi: 10.1016/j.ajhg.2011.01.010
- Carvalho, C. M., Polson, N. G., and Scott, J. G. (2010). The horseshoe estimator for sparse signals. *Biometrika* 97, 465–480. doi: 10.1093/biomet/asq017
- Elsen, J.-M., Tesseydre, S., Filangi, O., Le Roy, P., and Demeure, O. (2012). XVth QTLMAS: simulated dataset. *BMC Proc.* 6:S1. doi: 10.1186/1753-6561-6-S2-S1
- Ge, T., Chen, C. Y., Ni, Y., Feng, Y. A., and Smoller, J. W. (2019). Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat. Commun.* 10:1776. doi: 10.1038/s41467-019-09718-5
- Gianola, D., de los Campos, G., Hill, W. G., Manfredi, E., and Fernando, R. (2009). Additive genetic variability and the Bayesian alphabet. *Genetics* 183, 347–363. doi: 10.1534/genetics.109.103952
- Habier, D., Fernando, R. L., and Dekkers, J. C. M. (2007). The impact of genetic relationship information on genome-assisted breeding values. *Genetics* 177, 2389–2397. doi: 10.1534/genetics.107.081190
- Habier, D., Fernando, R. L., Kizilkaya, K., and Garrick, D. J. (2011). Extension of the bayesian alphabet for genomic selection. *BMC Bioinformatics* 12:186. doi: 10.1186/1471-2105-12-186
- Jairath, L., Dekkers, J. C. M., Schaeffer, L. R., Liu, Z., Burnside, E. B., and Kolstad, B. (1998). Genetic evaluation for herd life in Canada. *J. Dairy Sci.* 81, 550–562. doi: 10.3168/jds.S0022-0302(98)75607-3
- Johndrow, J. E., Orenstein, P., and Bhattacharya, A. (2017). Bayes shrinkage at GWAS scale: convergence and approximation theory of a scalable MCMC algorithm for the horseshoe prior. *arXiv [Preprint]*. arXiv: 170500841.
- Legarra, A., Robert-Granie, C., Manfredi, E., and Elsen, J. M. (2008). Performance of genomic selection in mice. *Genetics* 180, 611–618. doi: 10.1534/genetics.108.088575
- Li, X., Liu, X., and Chen, Y. (2018). The influence of a first-order antedependence model and hyperparameters in BayesC $\pi$  for genomic prediction. *Asian-Australas. J. Anim. Sci.* 31, 1863–1870. doi: 10.5713/ajas.18.0102
- Li, Y., Datta, J., Craig, B. A., and Bhadra, A. (2019). Joint mean-covariance estimation via the horseshoe with an application in genomic data analysis. *arXiv [Preprint]*. arXiv:190306768.

## ETHICS STATEMENT

The animal study was reviewed and approved by the Institutional of Animal Care and Use Committee (IACUC), China Agricultural University. Written informed consent was obtained from the owners for the participation of their animals in this study.

## AUTHOR CONTRIBUTIONS

SS conceived the study, wrote the program, analyzed the data, and drafted the manuscript. SZ and XD conceived the study and supervised the project. XL and LF participated in the design and helped to draft the manuscript. AL helped to analyze the data and revised the manuscript. GS revised the manuscript. YZ and BL helped to draft the manuscript. All authors read and approved the final manuscript.

## FUNDING

This work was supported by the China Agriculture Research System (CARS-36 and CARS-35), Ningxia Province High Quality and High Yield Breeding Project [(2019)21-1], Tianjin Science and Technology Project (19ZXZYSN00130), the National Key Research and Development Project (2019YFE0106800), and the Pearl River S&T Nova Program of Guangzhou (201906010040).

## ACKNOWLEDGMENTS

We thank all contributors to the present study. We also thank Dr. Emre Karaman and Dr. Ganjun Tu for their assistance in learning professional knowledge.



- MacLeod, I. M., Bowman, P. J., Vander Jagt, C. J., Haile-Mariam, M., Kemper, K. E., Chamberlain, A. J., et al. (2016). Exploiting biological priors and sequence variants enhances QTL discovery and genomic prediction of complex traits. *BMC Genomics* 17:144. doi: 10.1186/s12864-016-2443-6
- Madsen, P., and Jensen, J. (2012). *A User's Guide to DMU. A Package for Analysing Multivariate Mixed Models. Version 6, Release 5.1. Tjele, Denmark; 2012.* Available online at: [http://dmu.agrsci.dk/DMU/Doc/Previous/dmuv6\\_guide.5.1.pdf](http://dmu.agrsci.dk/DMU/Doc/Previous/dmuv6_guide.5.1.pdf) (accessed December 1, 2012).
- Makalic, E., and Schmidt, D. F. (2016). A simple sampler for the horseshoe estimator. *IEEE Signal Process. Lett.* 23, 179–182. doi: 10.1109/LSP.2015.2503725
- Makalic, E., Schmidt, D. F., and Hopper, J. L. (2016). “Bayesian robust regression with the horseshoe+ estimator,” in *Proceedings of the 29th Australasian Joint Conference Hobart, TAS, Australia, December 5–8, 2016: AI 2016: Advances in Artificial Intelligence*, eds B. H. Kang and Q. Bai (Cham: Springer International Publishing), 429–440.
- Meuwissen, T. H., Hayes, B. J., and Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819–1829.
- Piironen, J., and Vehtari, A. (2016). *On the Hyperprior Choice for the Global Shrinkage Parameter in the Horseshoe Prior.* Available online at: <https://ui.adsabs.harvard.edu/abs/2016arXiv161005559P> (accessed October 01, 2016).
- Piironen, J., and Vehtari, A. (2017). Sparsity information and regularization in the horseshoe and other shrinkage priors. *Electr. J. Statist.* 11, 5018–5051. doi: 10.1214/17-ejs1337si
- Pong-Wong, R., and Woolliams, J. (2014). “Bayes U: a genomic prediction method based on the horseshoe prior,” in *Proceedings of the 10th World Congress of Genetics Applied to Livestock Production*, Vancouver, BC, 679.
- Příbyl, J., Madsen, P., Bauer, J., Příbylová, J., Šimečková, M., Vostrý, L., et al. (2013). Contribution of domestic production records, Interbull estimated breeding values, and single nucleotide polymorphism genetic markers to the single-step genomic evaluation of milk production. *J. Dairy Sci.* 96, 1865–1873. doi: 10.3168/jds.2012-6157
- Vallée, A., Daures, J., van Arendonk, J. A. M., and Bovenhuis, H. (2016). Genome-wide association study for behavior, type traits, and muscular development in Charolais beef cattle. *J. Anim. Sci.* 94, 2307–2316. doi: 10.2527/jas.2016-0319
- VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91, 4414–4423. doi: 10.3168/jds.2007-0980
- Wand, M. P., Ormerod, J. T., Padoan, S. A., and Frühwirth, R. (2011). Mean field variational bayes for elaborate distributions. *Bayesian Anal.* 6, 847–900. doi: 10.1214/11-ba631
- Wolc, A., Stricker, C., Arango, J., Settar, P., Fulton, J. E., O'Sullivan, N. P., et al. (2011). Breeding value prediction for production traits in layer chickens using pedigree or genomic relationships in a reduced animal model. *Genet. Sel. Evol.* 43:5. doi: 10.1186/1297-9686-43-5
- Wu, X., Fang, M., Liu, L., Wang, S., Liu, J., Ding, X., et al. (2013). Genome wide association studies for body conformation traits in the Chinese Holstein cattle population. *BMC Genomics* 14:897. doi: 10.1186/1471-2164-14-897
- Yang, W., Chen, C., and Tempelman, R. J. (2015). Improving the computational efficiency of fully Bayes inference and assessing the effect of misspecification of hyperparameters in whole-genome prediction models. *Genet. Sel. Evol.* 47:13. doi: 10.1186/s12711-015-0092-x
- Yang, W., and Tempelman, R. J. (2012). A Bayesian antedependence model for whole genome prediction. *Genetics* 190, 1491–1501. doi: 10.1534/genetics.111.131540
- Zhou, L., Ding, X., Zhang, Q., Wang, Y., Lund, M. S., and Su, G. (2013). Consistency of linkage disequilibrium between Chinese and Nordic Holsteins and genomic prediction for Chinese Holsteins using a joint reference population. *Genet. Sel. Evol.* 45:7. doi: 10.1186/1297-9686-45-7
- Zhu, B., Zhu, M., Jiang, J., Niu, H., Wang, Y., Wu, Y., et al. (2016). The impact of variable degrees of freedom and scale parameters in Bayesian methods for genomic prediction in Chinese Simmental beef cattle. *PLoS One* 11:e0154118. doi: 10.1371/journal.pone.0154118

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Shi, Li, Fang, Liu, Su, Zhang, Luobu, Ding and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

## APPENDIX

### Appendix A: Gibbs Sampler for SNP Effect $\beta_k$

The full conditional distribution of  $\beta_k$  for BayesHP and BayesHE can be written as,

$$f(\beta_k|ELSE) \propto f(y|\mu, \beta, \sigma_e^2) * f(\beta_k|\tau^2, \lambda_k^2) \propto \exp\left[-\frac{(y - \mu - \sum_{k=1}^m x_k \beta_k)' R^{-1} (y - \mu - \sum_{k=1}^m x_k \beta_k)}{2\sigma_e^2}\right] * \exp\left[-\frac{\beta_k^2}{2\lambda_k^2 \tau^2}\right]$$

$$\propto \exp\left[-\frac{1}{2} \frac{(\beta_k - \hat{\beta}_k)^2}{\sigma_e^2/c_k}\right] \propto N(\hat{\beta}_k, \sigma_e^2/c_k)$$

where  $\hat{\beta}_k = \frac{x_k' R^{-1} w}{c_k}$ ,  $w = y - \mu - \sum_{j \neq k} x_j \beta_j$  and  $c_k = (x_k' R^{-1} x_k + \frac{\sigma_e^2}{\lambda_k^2 \tau^2})$ .

### Appendix B: Gibbs Sampler for $\lambda_k^2$

The full conditional distribution of  $\lambda_k^2$  can be written as Makalic et al. (2016),

$$f(\lambda_k^2|ELSE) \propto f(\beta_k|\tau^2, \lambda_k^2) * f(\lambda_k^2|\theta_k, \nu) \propto (2\pi\lambda_k^2\tau^2)^{-\frac{1}{2}} \exp\left[-\frac{\beta_k^2}{2\lambda_k^2\tau^2}\right] * \lambda_k^{2-(\frac{\nu}{2}+1)} \exp\left(-\frac{\nu}{\lambda_k^2}\right) \propto \lambda_k^{2-(\frac{\nu}{2}+1)}$$

$$\exp\left[-\frac{1}{\lambda_k^2} \left(\frac{\beta_k^2}{2\tau^2} + \frac{\nu}{\theta_k}\right)\right] \propto IG\left(\frac{\nu}{2} + \frac{1}{2}, \frac{\beta_k^2}{2\tau^2} + \frac{\nu}{\theta_k}\right)$$

Notably, in BayesHP,  $\nu$  equals to one.

### Appendix C: Gibbs Sampler for $\theta_k$

The full conditional distribution of  $\theta_k$  can be written as Makalic et al. (2016),

$$f(\theta_k|ELSE) \propto f(\lambda_k^2|\theta_k, \nu) * f(\theta_k) \propto \left(\frac{p}{\theta_k}\right)^{\left(\frac{p}{2}\right)} * \exp\left(-\frac{\nu}{\lambda_k^2}\right) * \theta_k^{-(\frac{1}{2}+1)}$$

$$\exp\left(-\frac{1}{\theta_k}\right) \propto \theta_k^{-(\frac{\nu}{2}+\frac{1}{2}+1)} \exp\left(-\frac{1}{\theta_k} \left(1 + \frac{\nu}{\lambda_k^2}\right)\right) \propto IG\left(\frac{\nu}{2} + \frac{1}{2}, 1 + \frac{\nu}{\lambda_k^2}\right)$$

Similarly, in BayesHP,  $\nu$  equals to one.

### Appendix D: Gibbs Sampler for $\eta_k^2$

The full conditional distribution of  $\eta_k^2$  can be written as Makalic et al. (2016),

$$f(\eta_k^2|ELSE) \propto f(\theta_k|\eta_k^2) * f(\eta_k^2|\nu_k) \propto (\eta_k^2)^{-\frac{1}{2}} \exp\left(-\frac{(\eta_k^2)^{-1}}{\theta_k}\right) * (\eta_k^2)^{-(\frac{1}{2}+1)} \exp\left(-\frac{\nu_k^{-1}}{\eta_k^2}\right) \propto \eta_k^{2-(1+1)}$$

$$* \exp\left(-\frac{1}{\eta_k^2} \left(\frac{1}{\nu_k} + \frac{1}{\theta_k}\right)\right) \propto IG\left(1, \frac{1}{\nu_k} + \frac{1}{\theta_k}\right)$$

### Appendix E: Gibbs Sampler for $\nu_k$

The full conditional distribution of  $\nu_k$  can be written as Makalic et al. (2016),

$$f(\nu_k|ELSE) \propto f(\eta_k^2|\nu_k) * f(\nu_k) \propto \nu_k^{-\frac{1}{2}} \exp\left(-\frac{\nu_k^{-1}}{\eta_k^2}\right) * \nu_k^{-(\frac{1}{2}+1)} \exp\left(-\frac{1}{\nu_k}\right) \propto \nu_k^{-(1+1)} \exp\left(-\frac{1}{\nu_k} \left(1 + \frac{1}{\eta_k^2}\right)\right) \propto IG\left(1, 1 + \frac{1}{\eta_k^2}\right)$$

(1)

## Appendix F: Gibbs Sampler for $\tau^2$

The full conditional distribution of  $\tau^2$  can be written as Makalic et al. (2016),

$$f(\tau^2 | ELSE) \propto f(\tau^2 | \zeta) * \prod_{k=1}^m f(\beta_k | \tau^2, \lambda_k^2) \propto (\tau^2)^{-\left(\frac{1}{2}+1\right)} \exp\left(-\frac{\zeta-1}{\tau^2}\right) * \prod_{k=1}^m (2\pi\lambda_k^2\tau^2)^{-\frac{1}{2}} \exp\left(-\frac{\beta_k^2}{2\lambda_k^2\tau^2}\right) \propto \tau^{2-\left(\frac{k+1}{2}+1\right)}$$

$$\exp\left[-\frac{1}{\tau^2}\left(\frac{1}{\zeta} + \sum \frac{\beta_k^2}{2\lambda_k^2}\right)\right] \propto IG\left(\frac{k+1}{2}, \frac{1}{\zeta} + \sum \frac{\beta_k^2}{2\lambda_k^2}\right)$$

## Appendix G: Gibbs Sampler for $\xi$

The full conditional distribution of  $\xi$  can be written as Makalic et al. (2016),

$$f(\xi | ELSE) \propto f(\tau^2 | \xi) * f(\xi) \propto \xi^{-\frac{1}{2}} \exp\left(-\frac{\xi-1}{\tau^2}\right) * \xi^{-\left(\frac{1}{2}+1\right)} \exp\left(-\frac{N^2}{\xi}\right) \propto \xi^{-(1+1)} * \exp\left[-\left(\frac{1}{\tau^2} + N^2\right)\frac{1}{\xi}\right] \propto IG\left(1, \frac{1}{\tau^2} + N^2\right)$$