

RESEARCH ARTICLE

# Genomic prediction with parallel computing for slaughter traits in Chinese Simmental beef cattle using high-density genotypes

Peng Guo<sup>1,2</sup>, Bo Zhu<sup>1</sup>, Lingyang Xu<sup>1\*</sup>, Hong Niu<sup>1</sup>, Zezhao Wang<sup>1</sup>, Long Guan<sup>1</sup>, Yonghu Liang<sup>1</sup>, Hemin Ni<sup>3</sup>, Yong Guo<sup>3</sup>, Yan Chen<sup>1</sup>, Lupei Zhang<sup>1</sup>, Xue Gao<sup>1</sup>, Huijiang Gao<sup>1</sup>, Junya Li<sup>1\*</sup>

**1** Laboratory of Molecular Biology and Bovine Breeding, Institute of Animal Science, Chinese Academy of Agricultural Sciences, Beijing, China, **2** College of Computer and Information Engineering, Tianjin Agricultural University, Tianjin, China, **3** Animal Science and Technology College, Beijing University of Agriculture, Beijing, China

\* [xulingyang@163.com](mailto:xulingyang@163.com) (LYX); [jl1@iascaas.net.cn](mailto:jl1@iascaas.net.cn) (JYL)



**OPEN ACCESS**

**Citation:** Guo P, Zhu B, Xu L, Niu H, Wang Z, Guan L, et al. (2017) Genomic prediction with parallel computing for slaughter traits in Chinese Simmental beef cattle using high-density genotypes. PLoS ONE 12(7): e0179885. <https://doi.org/10.1371/journal.pone.0179885>

**Editor:** Qin Zhang, China Agricultural University, CHINA

**Received:** January 16, 2017

**Accepted:** June 6, 2017

**Published:** July 19, 2017

**Copyright:** © 2017 Guo et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** We confirm that all data underlying our findings are publicly available without restriction. Data is available from the Dryad Digital Repository: <http://datadryad.org/resource/doi:10.5061/dryad.4qc06>

**Funding:** This work was supported by the National High Technology Research and Development Program of China (863 Program 2013AA102505-4), National Natural Science Foundation of China (31201782, 31672384 and 31372294), the Agricultural Science and Technology Innovation

## Abstract

Genomic selection has been widely used for complex quantitative trait in farm animals. Estimations of breeding values for slaughter traits are most important to beef cattle industry, and it is worthwhile to investigate prediction accuracies of genomic selection for these traits. In this study, we assessed genomic predictive abilities for average daily gain weight (ADG), live weight (LW), carcass weight (CW), dressing percentage (DP), lean meat percentage (LMP) and retail meat weight (RMW) using Illumina Bovine 770K SNP Beadchip in Chinese Simmental cattle. To evaluate the abilities of prediction, marker effects were estimated using genomic BLUP (GBLUP) and three parallel Bayesian models, including multiple chains parallel BayesA, BayesB and BayesCπ (PBayesA, PBayesB and PBayesCπ). Training set and validation set were divided by random allocation, and the predictive accuracies were evaluated using 5-fold cross validations. We found the accuracies of genomic predictions ranged from 0.195±0.084 (GBLUP for LMP) to 0.424±0.147 (PBayesB for CW). The average accuracies across traits were 0.327±0.085 (GBLUP), 0.335±0.063 (PBayesA), 0.347±0.093 (PBayesB) and 0.334±0.077 (PBayesCπ), respectively. Notably, parallel Bayesian models were more accurate than GBLUP across six traits. Our study suggested that genomic selections with multiple chains parallel Bayesian models are feasible for slaughter traits in Chinese Simmental cattle. The estimations of direct genomic breeding values using parallel Bayesian methods can offer important insights into improving prediction accuracy at young ages and may also help to identify superior candidates in breeding programs.

## Introduction

Genomic prediction has been widely used to predict breeding values of candidates with genome-wide SNP markers [1], this technology offers great promise to predict genetic merits

Program (ASTIP- IAS03 and ASTIP-IAS-TS-9), Cattle Breeding Innovative Research Team of Chinese Academy of Agricultural Sciences (cxgc-ias-03), Beijing Natural Science Foundation (6154032). L.Y.X was supported by the Elite Youth Program in Chinese Academy of Agricultural Sciences. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors declare that they have no competing interests.

of selection candidates for economic traits which are difficult or expensive to measure, for instance, traits which may only be measured by sacrificing potential breeding candidates, like carcass traits [2]. With the advance of genomic prediction, the genomic breeding values can be estimated at young ages, and help to promote the genetic progress of breeding in farm animals [3–6].

Carcass traits are important traits in beef cattle, many studies have been conducted to estimate the genomic breeding values of these traits including hot carcass weight, longissimus muscle area, carcass average backfat thickness, lean meat yield and carcass marbling score using BovineSNP50K Beadchip [2, 7–9].

Genomic prediction in beef cattle have been mainly carried out using lower-density SNP chip including Illumina BovineSNP50K Beadchip [2, 8–15], 15K SNP chip and 25K SNP chip [3]. In recent year, several studies have been conducted for genomic prediction using high-density SNP panels [4, 7, 16, 17], and they found genotyping with high density SNP chip can improve the accuracy of genomic prediction for Bayesian methods [18–20]. To obtain higher accuracies from low-density SNP panels, previous studies have attempted to impute lower-density SNPs into high density SNPs data [20–22], and these results suggested that predictive accuracies using imputation data outperformed those using low-density SNPs, while performance (both GBLUP and Bayesian methods) were also influenced by their imputation errors [18].

Many methods have been proposed for genomic prediction including Genomic Best Linear Unbiased Prediction (GBLUP) [23] and Bayesian methods [1, 24]. GBLUP is widely used for its merits of high estimation accuracies and short running time. Bayesian methods, implemented with Markov Chain Monte Carlo (MCMC), show high performances of predictive ability (easy implementation and robustness) in animals and plants breeding [25–27]. However, the iteration process in MCMC requires long computation time. Parallel computing using multiple processing units can shorten the running time of an intensive computational task [28]. Recently, Wu et al. used parallel MCMC to explore high-performance Bayesian computation in animal breeding, and their result suggested parallel MCMC could revolutionize computational tools for breeding programs for animals [15]. In this study, we further extended the parallel computing in genomic prediction by combining multiple chains parallel MCMC with Bayesian models. The objectives of this study were to 1) estimate prediction abilities of genomic selection for slaughter traits in Chinese Simmental beef cattle with GBLUP, parallel Bayesian methods. 2) evaluate the predictive accuracies of these methods. 3) provide valuable insights for application of genomic selection for slaughter traits in Chinese Simmental cattle.

## Methods

### Ethics statement

Animal experiments were approved by the Science Research Department of the Institute of Animal Science, Chinese Academy of Agricultural Sciences (CAAS) (Beijing, China).

### Simulation

We evaluated predictive accuracies and running time of our algorithm in simulation. Here, GPOPSIM software was used to generate simulation dataset including markers and QTLs [29]. Heritability was set to 0.5, the population included 1000 individuals, each chromosome included 10000 markers and the numbers of chromosomes per animal were set to 10. Mutation rate of marker and mutation rate of QTL were both set to  $1.25 \times 10^{-3}$  per locus per generation.

## Animals, phenotypes and SNP data

Analysis data were retrieved from the Dryad Digital Repository: <http://datadryad.org/resource/doi:10.5061/dryad.4qc06> which have been previously described in [30]. Average daily gain weight (ADG) was obtained with body weight gain divided by number of fatten day, the weight gain was the difference between the weight before slaughter and the weight entering in cattle farm. Live weights (LW) were measured before slaughter, and carcass weights (CW) were measured before carcasses being moved to chilling room. Then, carcasses were placed in chilling units for 48 hours before cuts. Retail meat weight (RMW) was estimated as  $RMW = \text{carcass weight} - \text{bone weight} - \text{weight of fat covering the carcass}$ . Dressing percentage (DP) were estimated as  $DP = \text{carcass weight} / \text{live weight}$ , and lean meat percentage (LMP) was  $LMP = (\text{carcass weight} - \text{bone weight}) / \text{live weight}$ . Summary statistics of these traits including number of animal, mean, standard deviation (SD), minimum and maximum of six traits were listed in Table 1.

To eliminate potential impact of environmental effects including farm, year of measurement and age for slaughter traits, we corrected phenotypes using the following equation as in [31],

$$y_{ijkm} = \mu + \text{Farm}_i + \text{Month}_j + \text{Year}_k + e_{ijkm}$$

where,  $y_{ijkm}$  is the vector of phenotype,  $\mu$  is the population mean,  $\text{Farm}_i$  is the category of the farm where the animal was raised,  $\text{Month}_j$  is the number of months after birth,  $\text{Year}_k$  is the year of slaughter,  $e_{ijkm}$  is the random residual. We processed SNP quality control using PLINK v1.07 [32] software and selected SNPs based on minor allele frequency ( $>0.05$ ), proportion of missing genotypes ( $<0.05$ ), Hardy-Weinberg equilibrium ( $p > 10^{-6}$ ). 1217 individuals remained after quality control (Table 2) and 671220 SNPs were included in autosomes.

## Statistical model

In this study, linear mixture model was used as following,

$$y_i = \mu + \sum_{j=1}^M Z_{ij} \alpha_j + e_i$$

where  $y_i$  is phenotype for individual  $i$ ,  $M$  is the number of SNPs,  $\mu$  is the overall mean,  $\alpha_j$  is the effect of locus  $j$ ,  $Z_{ij}$  is the SNP genotype code for individual  $i$  at locus  $j$  (coded as 0, 1, 2),  $e_i$  is the random residual effect for individual  $i$ .

**Table 1. Summary statistics of slaughter traits (number of animal, mean, and standard deviation (SD), maximum and minimum of each trait).**

Trait	Num.	Mean±SD	Maximum	Minimum
ADG	1294	0.97 ±0.22	2.41	0.38
LW	1302	505.26±70.73	776	318
CW	1302	271.35 ±45.63	486	162.6
DP	1301	53.56±2.91	68.98	41.03
LMP	1301	45.47 ±3.08	61.56	32.51
RMW	1299	169.94±29.80	280.87	84

ADG: Average daily gain weight (kg), LW: Live weight (kg), CW: Carcass weight (kg), DP: Dressing Percentage (%), LMP: Lean meat percentage (%), RMW: Retail meat weight (kg). Num: Number of animal, Mean±SD: phenotypic mean and standard deviation.

<https://doi.org/10.1371/journal.pone.0179885.t001>

**Table 2. Birth year distribution for genotyped animals.**

Birth year	Animals
2008	91
2009	193
2010	328
2011	276
2012	208
2013	121
Total	1217

<https://doi.org/10.1371/journal.pone.0179885.t002>

### BayesA

All loci are assumed to have effects for the trait of interest in BayesA. The prior distribution of effects  $\alpha_j$  is assumed to be a normal distribution with a mean 0 and a variance  $\sigma_{\alpha_j}^2$ , whereas the prior distribution of  $\sigma_{\alpha_j}^2$  belongs to scaled inverted chi-square distribution,  $\chi^{-2}(\nu, S)$ , where S is a scale parameter and  $\nu$  is the number of degrees of freedom.  $\nu = 4.012$  and  $S = 0.0020$  are used as the prior distribution of  $\sigma_{\alpha_j}^2$ . Gibbs sampling is used for the estimation of marker effects and variances [1].

### BayesB

BayesB assumes that some SNPs have zero effect, while other SNPs are assumed to have large effects. Therefore, parameter  $\pi$  is used in BayesB to control whether the locus has a nonzero effect or not.

$$\begin{cases} \sigma_{a_i}^2 = 0 & \text{probability } \pi \\ \sigma_{a_i}^2 \sim \chi^{-2}(\nu, S) & \text{probability } (1 - \pi) \end{cases}$$

Where  $\nu = 4.234$  and  $S = 0.0429$  are suggested to yield the mean and variance of  $\sigma_{a_i}^2$ . Metropolis Hasting algorithm is used to implement the sampling of variances [1]. In our study, we set  $\pi$  to 0.99.

### BayesCπ

BayesC $\pi$  modifies BayesB method by replacing the locus-specific variance components with a common effect variance, and this method assumes an unknown fraction  $\pi$  [with uniform (0, 1) prior] of SNP with a nonzero effect, the common variance has a scaled inverse chi-square prior with parameter  $\nu = 4.2$  and scale factor S, where S is derived as for BayesB [24]. The probability  $\pi$  is treated as an unknown with uniform (0,1) prior, and the effect of a SNP fitted with probability (1- $\pi$ ) comes from a mixture of multivariate student's t-distributions.

### GBLUP

GBLUP uses mixed model equations with a genomic relationship matrix, assuming a prior normal distribution for SNP markers. The relationship matrix (**A**) based on pedigree is substituted by the genomic relationship matrix (**G**) in GBLUP as defined by VanRaden [23], the G matrix is formulated as follows,

$$G = \frac{ZZ'}{2 \sum_{i=1}^n q_i(1 - q_i)},$$

where  $n$  is the number of loci,  $q_i$  is the frequency of an allele of the marker  $i$ , and  $\mathbf{Z}$  is a centered incidence matrix of SNP effects, corrected for allele frequencies [23].

### Implementation of multiple chains parallel Bayesian prediction

MCMC includes two steps, sampling in burn-in and sampling after burn-in. In multiple chains parallel MCMC, sampling in burn-in should be implemented sequentially and parallelization can only happen in sampling after burn-in [33]. Thus, sampling in burn-in in parallel MCMC requires the same number of iteration as that in sequential MCMC. In experiments, the number of chains used in parallel Bayesian models in simulation were set to 1 chain (sequential models) and 9 different multiple (2, 4, 6, 8, 10, 12, 14, 16, 18) chains, while the number of chains was set to 16 on real dataset. The maximal iteration of MCMC (both on simulation dataset and real dataset) was set to 50000 with 5000 burn-in.

In parallel computing, computing tasks are executed in process, and each process is dispatched to one computing core. In our study, computation in Bayesian model was divided into sequential part and parallel parts, sequential part was implemented by master process, and its tasks included loading data from files, initializing parameters, broadcasting data and parameters to parallel parts. Parallel parts were implemented by slave processes independently, slave processes computing tasks included random number seed setting, burn-in computing, estimating locus effect and calculating GEBVs (Fig 1).

### Multiple chains convergence diagnosis

Multiple chains convergence diagnosis followed Gelman and Rubin's method [34].  $\hat{R}$  is the shrink factor, if  $\sqrt{\hat{R}} > 1$ , the chains don't converge, if  $\sqrt{\hat{R}} \approx 1$ , the chains converge.

### GEBV calculation

GEBV is calculated as the sum of all SNP effects according to marker genotypes and genotype effects. Just as the following equation.

$$GEBV_i = \sum_j Z_{ij}g_j$$

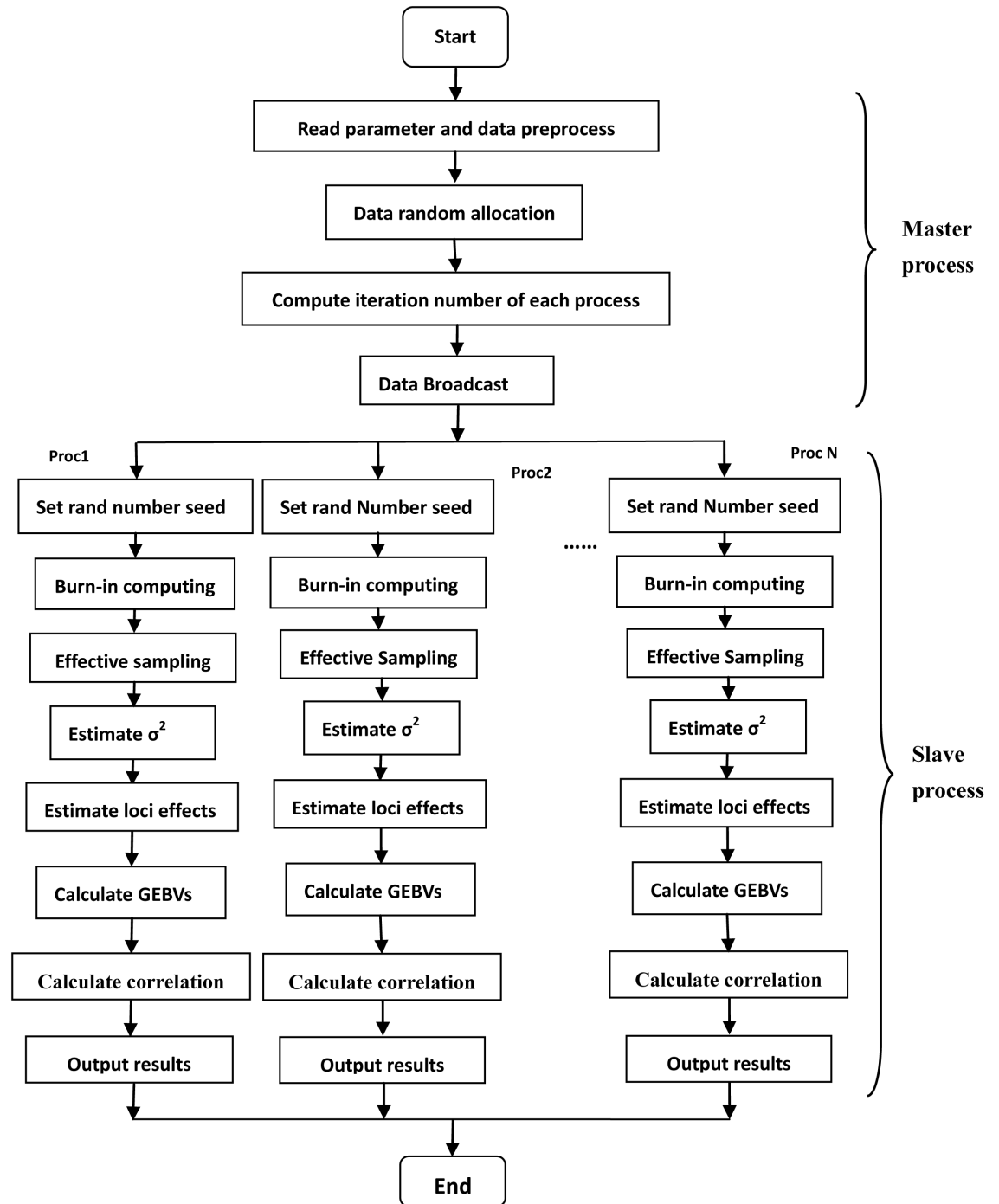
where  $GEBV_i$  is the genomic estimated breeding value of animal  $i$ ,  $Z_{ij}$  is a genotype for SNP  $j$  of animal  $i$ , and  $g_j$  is the estimated effect of the  $j$ th SNP locus.

### Cross-validation procedure

To evaluate the predictive accuracies, random masking cross-validation method was used in this study [13]. A total of 1217 Simmental cattle were divided into validation set and training set. Phenotypes of animals in the validation set were assumed unknown. Five-fold cross validation was used to assess the accuracies of prediction, and 1217 individuals were randomly partitioned into five groups. In each time, about one-fifth of 1217 Simmental were randomly picked out as the validation set and the remaining individuals were used as the training set. For each trait, the procedure was repeated 10 times and the average value was calculated as the GEBV.

### Predictive criterion

To remove the influence of the heritability for predictive ability, we used Pearson's correlation between GEBVs and corrected phenotypes divided by square root of heritability ( $r_{\hat{g},\hat{y}}/\sqrt{h^2}$ ),



**Fig 1. Workflow of multiple chains parallel Bayesian genomic prediction.** Proc1: Process 1, Proc2: Process 2, ProcN: Process N.  $\sigma^2$ : variance of normal distribution for estimated effects.

<https://doi.org/10.1371/journal.pone.0179885.g001>

here,  $r_{\hat{g}, \hat{y}}$  was the correlation between GEBVs and corrected phenotypes,  $\hat{y}$  was the vector of corrected phenotype in validation set and  $\hat{g}$  was the vector of GEBV calculated with SNP data in validation set and effects obtained in training set [8]. Moreover, we compared results using average values and standard deviations of predictive accuracies.

## Computer system

Our experiments were conducted on HP ProLiant DL585 G7 server, which was equipped with AMD Opteron 6344(2.6GHz) CPU, 272G Memory and L2 cache size 4M, L3 cache size 16M. We wrote programs in C language within Message Passing Interface (MPI) system, MPICH2 is an open source MPI implementation and a standard for message-passing in parallel computing, it is available freely (<http://www.mpich.org/downloads>). The Integrated Development Environment we used is Dev-C++ 5.1, which is published freely (<http://www.bloodshed.net/index.html>).

## Result

### Results using simulation dataset

Predictive accuracies using multiple chains were shown in Table 3. For PBayesA, PBayesB or PBayesCπ, there were tiny differences of predictive accuracies among different chains' results from the same parallel Bayesian method, the maximal difference was from PBayesCπ, where the largest accuracy (0.868763 using 10 chains) was 0.18% (percent point difference) higher than the smallest accuracy (0.86716 using 4 chains), and the minimal difference was from PBayesA, where the largest accuracy (0.836904 using 6 chains) was 0.04% higher than the smallest accuracy (0.836559 using 8 chains). In this study, the descending order of predictive accuracies for four methods were found (PBayesB > PBayesCπ > PBayesA > GBLUP) in simulation. We evaluated the running time across PBayesA, PBayesB and PBayesCπ in simulation, and we found the running time reduced obviously for the three parallel Bayesian methods with increase of chain number (Fig 2).

### Predictive accuracies

In this study, we calculated heritabilities of slaughter traits (Table 4) using restricted maximum likelihood (REML) based on animal model. Random masking cross-validation method was applied to assess the predictive accuracies of slaughter traits in Simmental cattle population. In general, the predictive accuracies for most traits were slightly different between parallel Bayesian models and GBLUP. Accuracies of genomic predictions were ranged from 0.195±0.084 (GBLUP for LMP) to 0.424±0.147 (PBayesB for CW). The average accuracies across traits were

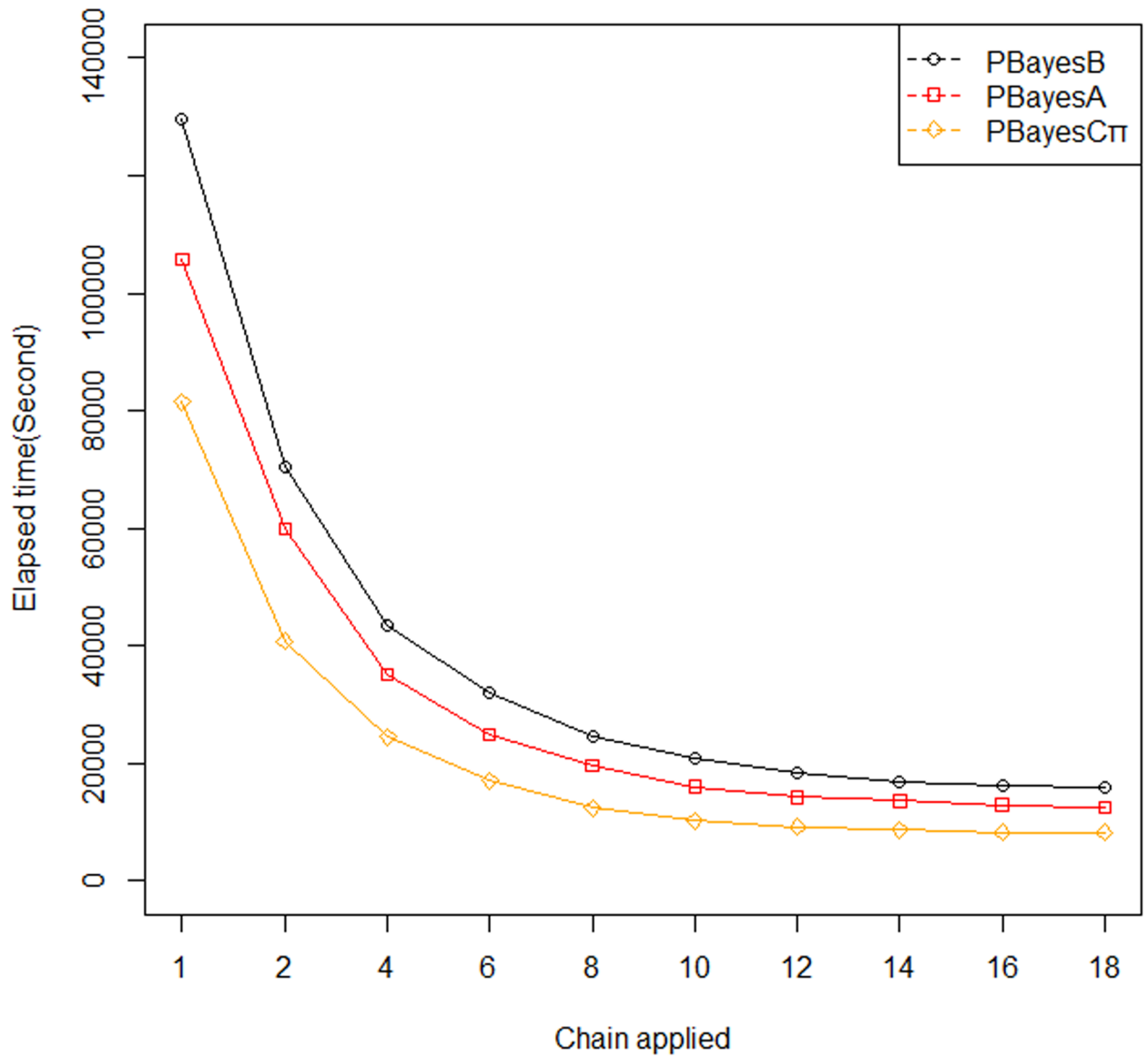
**Table 3. Predictive accuracies using four methods using PBayesA, PBayesB, PBayesCπ and GBLUP in simulation.**

	PBayesA	PBayesB	PBayesCπ	GBLUP
1 ch	0.836777	0.887837	0.867312	0.829173
2ch	0.836731	0.887741	0.867161	
4ch	0.836797	0.887449	0.867160	
6ch	0.836904	0.888563	0.868017	
8ch	0.836559	0.888305	0.868319	
10ch	0.836770	0.888232	0.868763	
12ch	0.836862	0.888997	0.868347	
14ch	0.836814	0.888535	0.868426	
16ch	0.836844	0.888852	0.868385	
18ch	0.836820	0.888823	0.868210	

ch means chains used in experiments, GBLUP: Genomic Best Linear Unbiased Prediction, PBayesA: multiple chains parallel BayesA, PBayesB: multiple chains parallel BayesB, PBayesCπ: multiple chains parallel BayesCπ.

<https://doi.org/10.1371/journal.pone.0179885.t003>

### Running Time in simulation



**Fig 2. Comparisons of running time in simulation.** Axis x indicates chains number used in parallelization and axis y indicates running time.

<https://doi.org/10.1371/journal.pone.0179885.g002>

0.327±0.085 for GBLUP, 0.335±0.063 for PBayesA, 0.347±0.093 for PBayesB and 0.334±0.077 for PBayesCπ (Table 4). Prediction accuracies among the four methods for six traits were presented in Fig 3.

For most traits, parallel Bayesian methods resulted in slightly higher accuracies than GBLUP. For LW, CW and DP, PBayesB performed best among these four methods, and the percentage point differences between PBayesB and GBLUP were 9.02% for LW, 6.80% for CW and 12.27% for DP respectively. For LMP, PBayesA showed higher predictive accuracy than GBLUP (10.77%). For RMW, we found PBayesCπ, PBayesB and PBayesA were superior to GBLUP, while GBLUP was superior over parallel Bayesian methods for ADG.



**Table 4. Heritabilities estimation and predictive accuracies of GEBVs for slaughter traits in Chinese Simmental cattle.**

Trait	h <sup>2</sup>	Average value (standard deviation) of correlations				Average value (standard deviation) of correlations divided by square root of heritability			
		GBLUP	PBayesA	PBayesB	PBayesCπ	GBLUP	PBayesA	PBayesB	PBayesCπ
LW	0.37	0.236 (0.060)	0.247 (0.032)	0.257 (0.040)	0.242 (0.044)	0.388 (0.099)	0.405 (0.053)	0.423 (0.066)	0.398 (0.072)
CW	0.45	0.266 (0.059)	0.271 (0.057)	0.285 (0.098)	0.268 (0.080)	0.397 (0.089)	0.404 (0.085)	0.424 (0.147)	0.399 (0.119)
DP	0.16	0.111 (0.031)	0.114 (0.021)	0.124(0.045)	0.114 (0.021)	0.277 (0.076)	0.285 (0.053)	0.311 (0.113)	0.285 (0.052)
LMP	0.14	0.073 (0.031)	0.081 (0.025)	0.080 (0.009)	0.074 (0.010)	0.195 (0.084)	0.216 (0.066)	0.214 (0.023)	0.198 (0.028)
RMW	0.43	0.258 (0.058)	0.265 (0.050)	0.265 (0.021)	0.271 (0.050)	0.393 (0.088)	0.404(0.076)	0.404 (0.032)	0.413 (0.076)
ADG	0.47	0.214 (0.052)	0.203 (0.029)	0.210 (0.120)	0.213 (0.079)	0.312 (0.076)	0.297 (0.042)	0.306 (0.175)	0.311 (0.115)
mean		0.193 (0.049)	0.197 (0.036)	0.204 (0.056)	0.197 (0.047)	0.327 (0.085)	0.335 (0.063)	0.347 (0.093)	0.334 (0.077)

LW: Live weight (kg), CW: Carcass weight (kg), DP: Dressing Percentage (%), LMP: Lean meat percentage (%), RMW: Retail meat weight (kg), ADG: Average daily gain weight (kg).

<https://doi.org/10.1371/journal.pone.0179885.t004>

### Posterior samples of residual variance

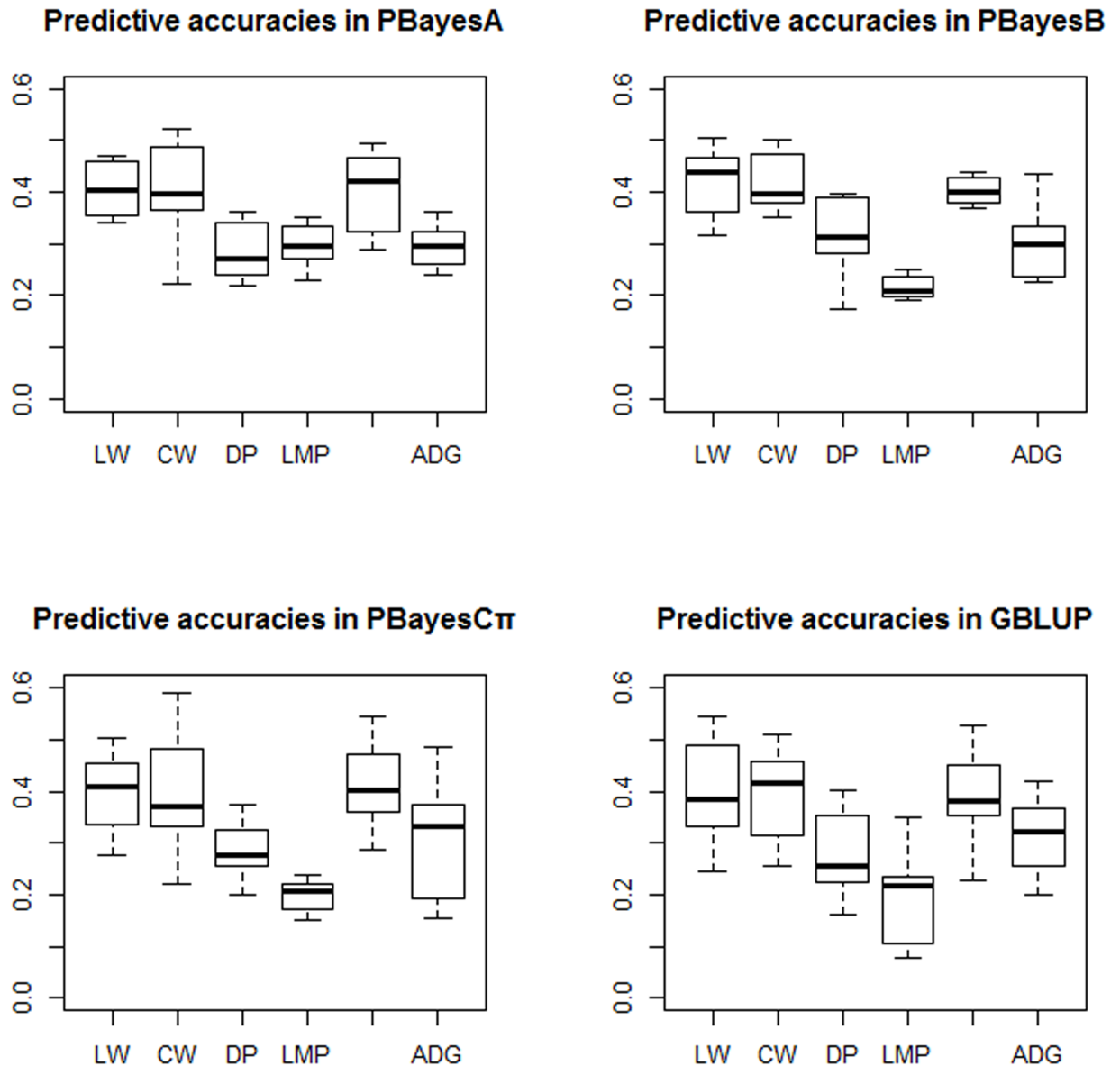
Posterior samples of residual variance were used in convergence diagnosis analysis as described in previous study [15]. The largest percent point difference among PBayesA, PBayesB and PBayesCπ was found for ADG, the difference happened between PBayesB and PBayesCπ (30.91%), posterior samples of residual variance approached 0.0061 (PBayesA), 0.0072 (PBayesB) and 0.0055 (PBayesCπ) which were shown in Fig 4P–Fig 4R. While the slightest percent point difference was found in RMW, the difference happened between PBayesA and PBayesCπ (1.33%), posterior samples of residual variance were 152 (PBayesA), 154 (PBayesB) and 150 (PBayesCπ) (Fig 4M–Fig 4O). For LW, CW, DP and LMP, we also observed slight differences for posterior samples of the residual variances using PBayesA, PBayesB and PBayesCπ (Fig 4A–4L).

### Convergence diagnose of multiple chains

In multiple Markov chains parallel Bayesian genomic prediction, convergence diagnose helps determine the equilibrium of MCMCs. With convergence diagnosis criterion proposed by Gelman and Rubin [34], we assessed the convergence of multiple chains for the genomic prediction of slaughter traits, and we observed the shrink factors of PBayesA, PBayesB and PBayesCπ quickly approached 1.00 for six traits (Fig 5), which indicated multiple chains converged in parallel Bayesian models.

### Discussion

In this study, we carried out genomic prediction for slaughter traits using GBLUP and Bayesian models in Chinese Simmental cattle. In the last decade, beef cattle have been selected for various economic traits such as growth [2, 7–9], carcass [10–14], meat [35] and reproduction [35, 36]. To maximize the economic benefits of beef cattle reproduction, selection for economically important traits is desirable. Therefore, slaughter traits (live weight, carcass weight,

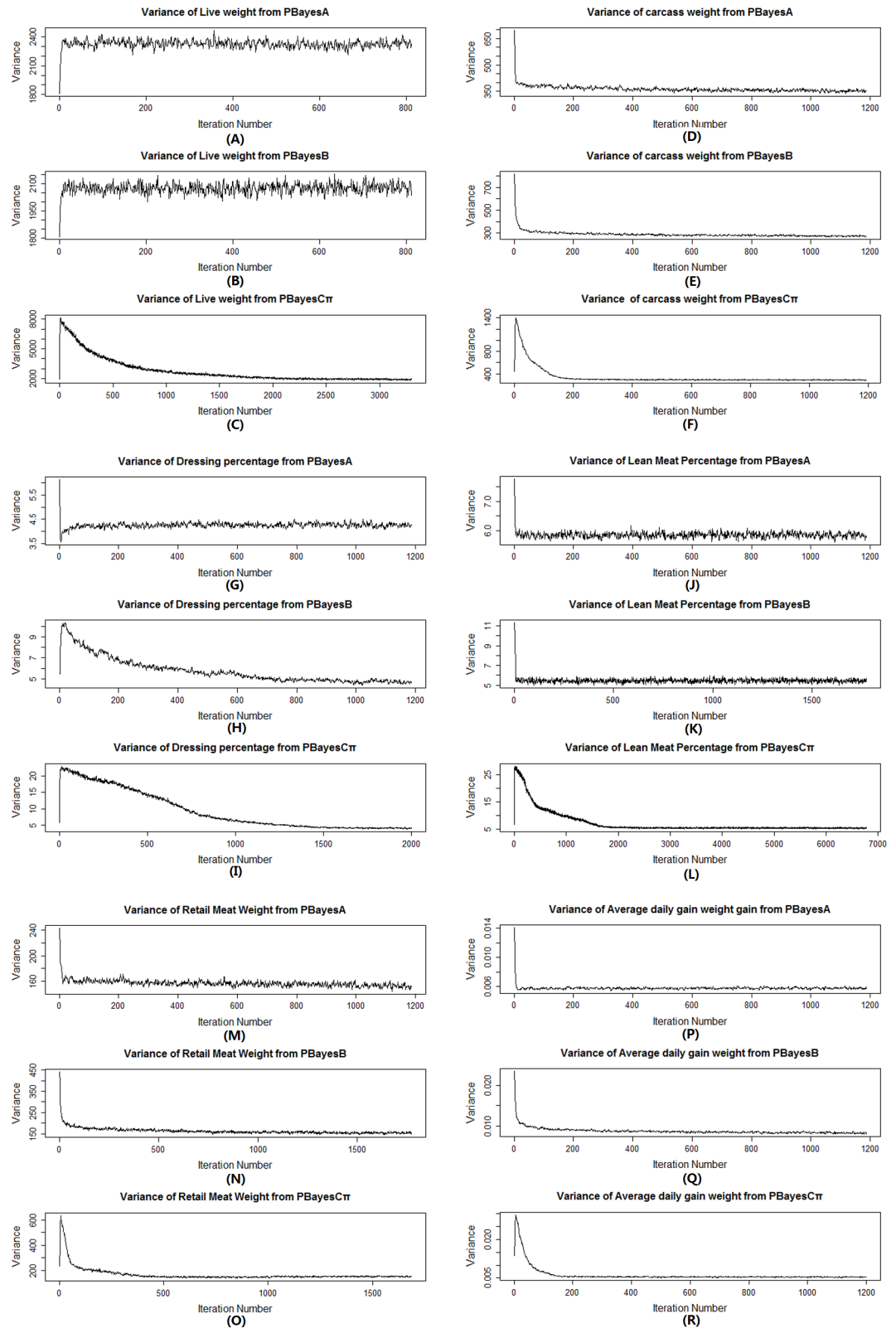


**Fig 3. Predictive accuracies using GBLUP, PBayesA, PBayesB and PBayesC $\pi$  for slaughter traits in Chinese Simmental cattle.** LW: Live weight (kg), CW: Carcass weight (kg), DP: Dressing Percentage (%), LMP: Lean meat percentage (%), RMW: Retail meat weight (kg), ADG: Average daily gain weight (kg). PBayesA: multiple chains parallel BayesA, PBayesB: multiple chains parallel BayesB, PBayesC $\pi$ : multiple chains parallel BayesC $\pi$ . GBLUP: Genomic Best Linear Unbiased Prediction.

<https://doi.org/10.1371/journal.pone.0179885.g003>

dressing percentage, lean meat percentage, retail meat weight and average daily gain weight) have been mostly focused by beef cattle industry.

Genomic predictions have aroused scientists' interests for merits of robustness, easy implementation and higher predictive capability. The intensive computing of Bayesian models may require days, weeks, or even months of computing time on personal computers or workstations [15] and this computational burden is the most obvious obstacle for its application in animals and plants breeding. Strandén et al. used parallel preconditioned conjugate gradient method to estimate breeding values in Finnish dairy cattle, running time using four processors was obviously reduced in contrast to that of sequential mode [37]. Using theoretical and



**Fig 4. Trace plots of posterior samples of residual variances in burn-in (from start to equilibrium) from multiple chains parallel Bayesian models for 6 traits.** (A) Trace plots for live weight, (B) Trace plots for carcass weight, (C) Trace plots for dressing percentage, (D) Trace plots for lean meat percentage (E) Trace plots for retail

meat weight, (F) Trace plots for average daily weight gain. PBayesA: multiple chains parallel BayesA, PBayesB: multiple chains parallel BayesB, PBayesCπ: multiple chains parallel BayesCπ.

<https://doi.org/10.1371/journal.pone.0179885.g004>

experiment analyses, Wu et al. found obvious reduction of running time for experimental results using parallel MCMC method in breeding estimation [15]. Running time reduction of PBayesA, PBayesB and PBayesCπ using simulation dataset (Fig 2) were consistent with previous studies [15, 37]. In current study, we used parallel BayesA, BayesB and BayesCπ to estimate genomic breeding values for slaughter traits by dividing the heavy computing task into several segments, and our results provided valuable insights for application of genomic selection using parallel MCMC for these traits in Chinese Simmental cattle.

## Model comparisons

GBLUP shows obvious superiority over Bayesian models on computing time, for instance, the time taken in GBLUP is less than one minute for each of the 5-fold cross validation, while 3 days were required in the genomic prediction using Bayesian models. The reason for obvious difference in computing time may be caused by model, population size and marker number. In GBLUP, genomic matrix calculation is a time consuming process, and for a population with certain number of individuals and genotyped data, genomic matrix calculation is implemented only once and the result can be reused in genomic prediction for other traits in the same population. While in Bayesian models, effect of each locus was estimated with MCMC method, the MCMC sampling procedure should be implemented thousands of times.

Bayesian methods can appropriately model the architecture of QTL effects within the genome, especially for traits that possess large effect QTLs [13]. It has previously been observed that the genomic predictive ability depends on attributes of genetic architecture of the trait, population size and particular model. We observed the predictive accuracies of Bayesian models were slightly different for 6 traits using 3 parallel Bayesian methods, and the performance of accuracy was PBayesB > PBayesA > PBayesCπ.

## Genomic prediction methods

In this study, we found parallel Bayesian models outperformed GBLUP for most traits. Previous studies have suggested that GBLUP outperformed Bayesian methods using low-density chip including 15K SNP chip [3] and 25K SNP chip [3, 38]. In contrast, Erbe et al. suggested Bayesian method (Bayes R) was superior over GBLUP after analyzing genomic selection in dairy cattle using imputed high-density panel, and their finding also implied Bayesian methods may take full advantage of the increased marker density [25]. Bayesian methods outperformed GBLUP for traits controlled by several SNPs with large effects, while GBLUP performed better for those traits which were not controlled by large effects SNPs. This could be explained that the genetic architecture of ADG was different from other traits. Our results also suggested that GBLUP was suitable for ADG, while PBayesA, PBayesB and PBayesCπ were suitable for other traits in Chinese Simmental cattle population.

## Accuracies of genomic predictions

To comprehensively evaluate the accuracies of estimated breeding values among PBayesA, PBayesB and PBayesCπ, we ran different multiple chains in simulation data set using the three methods. For the same Bayesian methods, we found that slight difference among predictive accuracies of sequential Bayesian method and multiple chains parallel Bayesian methods, this indicated that parallel Bayesian methods can generate equivalent accuracies comparing to that

of sequential Bayesian methods. In general, the descending order of predictive accuracies in simulation was  $P_{\text{BayesB}} > P_{\text{BayesC}\pi} > P_{\text{BayesA}} > \text{GBLUP}$ .

Accuracies of genomic prediction can be impacted by the model, heritability of the trait, the size of the reference population, the density of the SNP panel and level of LD [2]. Previous study revealed that traits with a larger number of genotyped animals and higher heritability generated the higher accuracy of GEBV [7]. For six studied traits, we found obvious differences among the estimated heritabilities, heritabilities of LW ( $h^2 = 0.37$ ), CW ( $h^2 = 0.45$ ), RMW ( $h^2 = 0.43$ ) were higher than those of DP ( $h^2 = 0.16$ ) and LMP ( $h^2 = 0.14$ ), while predictive accuracies for LW, CW and RMW were higher than those for DP and LMP, and our findings were consistent with previous studies [7, 39]. Notably, we found the heritability for ADG was 0.47, and the predictive accuracies were  $0.297 \pm 0.042$  (for  $P_{\text{BayesA}}$ ),  $0.306 \pm 0.175$  (for  $P_{\text{BayesB}}$ ),  $0.311 \pm 0.115$  (for  $P_{\text{BayesC}\pi}$ ) and  $0.312 \pm 0.076$  (for GBLUP), thus, our results suggested that density of the SNP panel, level of LD and the model may also have important impacts on predictive accuracies.

Compared to accuracies of CW in previous studies [7, 39], our results ( $0.397 \pm 0.089$  for GBLUP,  $0.404 \pm 0.085$  for BayesA,  $0.424 \pm 0.147$  for BayesB and  $0.399 \pm 0.119$  for BayesC $\pi$ ) was higher than those of Nellore ( $0.37 \pm 0.053$  for Bayesian ridge regression,  $0.36 \pm 0.058$  for BayesC and  $0.37 \pm 0.056$  for Bayesian Lasso) [39], Angus (0.16 for GBLUP), Shorthorn (0.19 for GBLUP), Brahman (0.28 for GBLUP) and Santa Gertrudis (0.29 for GBLUP) cattle, Hereford (0.32 for GBLUP), Belmont Red (0.33 for GBLUP), and was similar to that of Murray Grey (0.39 for GBLUP) cattle [7]. For ADG, accuracies of our results ( $0.312 \pm 0.076$  for GBLUP,  $0.297 \pm 0.042$  for  $P_{\text{BayesA}}$ ,  $0.306 \pm 0.175$  for  $P_{\text{BayesB}}$ ,  $0.311 \pm 0.115$  for  $P_{\text{BayesC}\pi}$ ) were higher than Angus (0.24 for GBLUP and for Bayes R), Belmont Red (0.24 for GBLUP and 0.18 for BayesR), Brahman crosses (0.13 for GBLUP and 0.27 for BayesR), Santa Gertrudis (0.21 for GBLUP and 0.23 for BayesR) [7]. This indicated that genomic selection using multiple chains parallel Bayesian models was suitable for genomic prediction for LW, CW, RMW, DP and LMP in Chinese Simmental beef cattle.

## Multiple chains convergence diagnosis

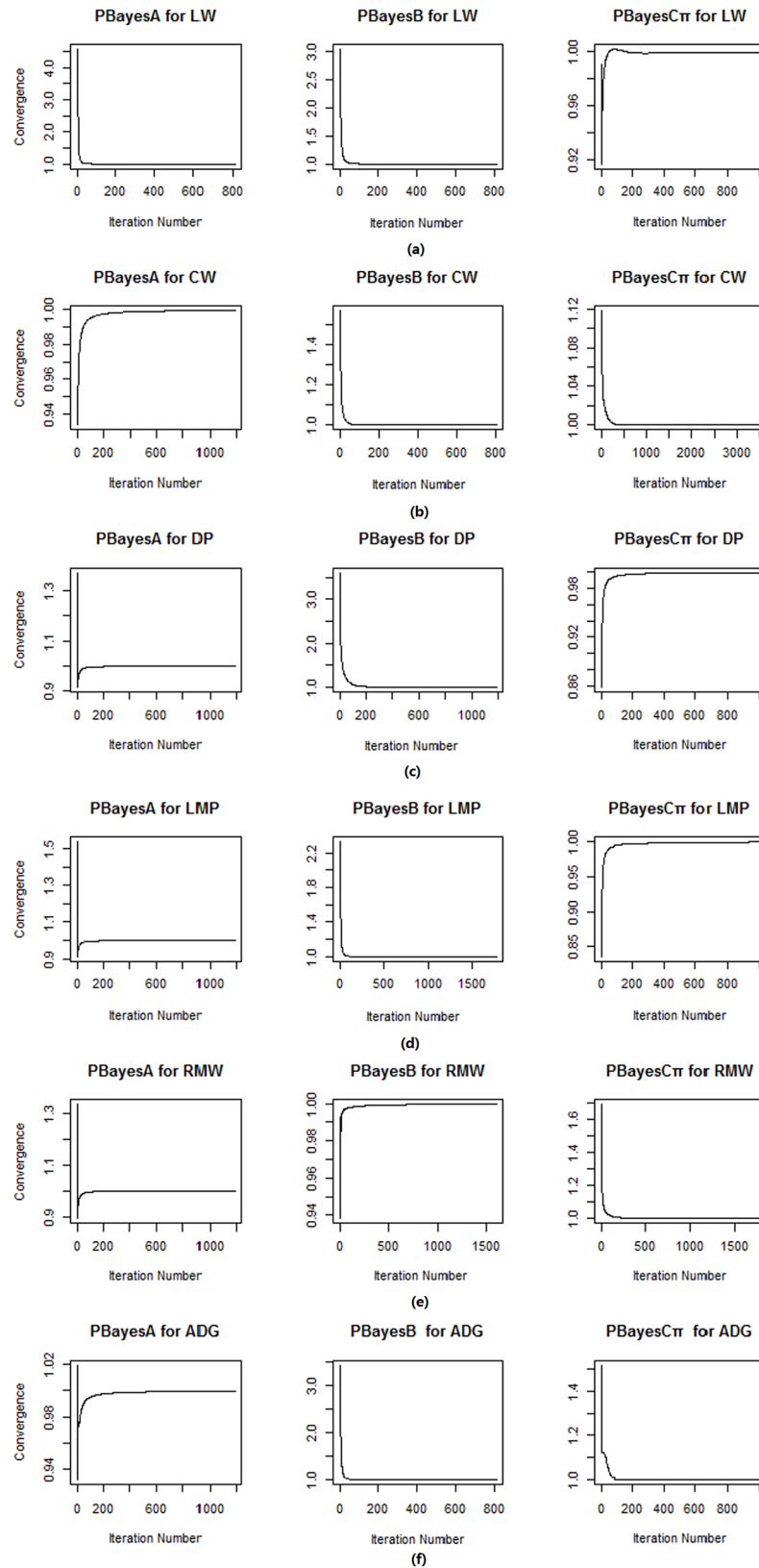
In multiple chains MCMC, effective sampling should happen when chains converges. During the evaluation of convergence of multiple chains, we observed sampling results from start to the point when chains being in equilibrium in burn-in step. Sampling results and shrink factors in equilibrium were stable and we omitted part of trace plots in equilibrium across traits.

For multiple chains MCMC, each one started with different initial value, and all chains should converge after a certain number of iteration. We used Gelman and Rubin's method [34] to evaluate multiple chains' convergence state. The convergence was examined using posterior samples of the residual variance collected from each chain. Posterior samples showed slight differences among parallel Bayesian models for the same traits and trace plots of posterior samples of the residual variance indicated that most chains tended to stabilize after 2000 iterations (Fig 4).

We assessed parallel Bayesian models for six traits, and all shrink factors approached 1 (Fig 5(A)–5(F)). Wu et al. suggested that a burn-in of 3000 iterations being more appropriate [15], our results showed that shrink factors approached 1 with less than 3000 iteration, this finding suggested that multiple chains MCMC converged obviously in Simmental beef cattle dataset.

## Conclusions

Our study demonstrated that it is feasible for the application of parallel genomic prediction for slaughter traits in Chinese Simmental beef cattle. Our results indicated that parallel BayesB outperformed GBLUP, parallel BayesA and parallel BayesC $\pi$ . Moreover, the predictive



**Fig 5. Trace plot of convergence of multiple chains (from start to equilibrium) for 6 traits.** (a) Parallel Bayesian models for LW, (b) Parallel Bayesian models for CW, (c) Parallel Bayesian models for DP, (d)

Parallel Bayesian models for LMP, (e) Parallel Bayesian models for RMW, (f) Parallel Bayesian models for ADG. PBayesA: multiple chains parallel BayesA, PBayesB: multiple chains parallel BayesB, PBayesCт: multiple chains parallel BayesCт.

<https://doi.org/10.1371/journal.pone.0179885.g005>

accuracies of parallel Bayesian models were more accurate than GBLUP for most traits and these methods are interest for the future application of genomic selection in farm animals.

## Author Contributions

**Conceptualization:** Huijiang Gao, Junya Li.

**Data curation:** Hong Niu, Yan Chen, Xue Gao.

**Formal analysis:** Peng Guo.

**Funding acquisition:** Hemin Ni, Yong Guo, Lupei Zhang, Junya Li.

**Investigation:** Zezhao Wang, Yonghu Liang.

**Methodology:** Peng Guo, Huijiang Gao.

**Project administration:** Huijiang Gao, Junya Li.

**Resources:** Bo Zhu, Long Guan.

**Software:** Peng Guo, Bo Zhu.

**Supervision:** Huijiang Gao, Junya Li.

**Validation:** Peng Guo, Bo Zhu.

**Visualization:** Peng Guo.

**Writing – original draft:** Peng Guo, Lingyang Xu.

**Writing – review & editing:** Peng Guo, Lingyang Xu.

## References

1. Meuwissen THE, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. *Genetics*. 2001; 157(4):1819–29. PMID: [11290733](https://pubmed.ncbi.nlm.nih.gov/11290733/)
2. Chen L, Vinsky M, Li C. Accuracy of predicting genomic breeding values for carcass merit traits in Angus and Charolais beef cattle. *Animal Genetics*. 2015; 46(1):55–9. <https://doi.org/10.1111/age.12238> PMID: [25393962](https://pubmed.ncbi.nlm.nih.gov/25393962/)
3. Moser G, Tier B, Crump RE, Khatkar MS, Raadsma HW. A comparison of five methods to predict genomic breeding values of dairy bulls from genome-wide SNP markers. *Genetics Selection Evolution*. 2009; 41. <https://doi.org/10.1186/1297-9686-41-56> PMID: [20043835](https://pubmed.ncbi.nlm.nih.gov/20043835/)
4. Neves HHR, Carvalheiro R, O'Brien AMP, Utsunomiya YT, do Carmo AS, Schenkel FS, et al. Accuracy of genomic predictions in *Bos indicus* (Nellore) cattle. *Genetics Selection Evolution*. 2014; 46. <https://doi.org/10.1186/1297-9686-46-17> PMID: [24575732](https://pubmed.ncbi.nlm.nih.gov/24575732/)
5. de Campos CF, Lopes MS, Silva FFE, Veroneze R, Knol EF, Lopes PS, et al. Genomic selection for boar taint compounds and carcass traits in a commercial pig population. *Livest Sci*. 2015; 174:10–7. <https://doi.org/10.1016/j.livsci.2015.01.018>
6. Duchemin SI, Colombani C, Legarra A, Baloché G, Larroque H, Astruc JM, et al. Genomic selection in the French Lacaune dairy sheep breed. *Journal of Dairy Science*. 2012; 95(5):2723–33. <https://doi.org/10.3168/jds.2011-4980> PMID: [22541502](https://pubmed.ncbi.nlm.nih.gov/22541502/)
7. Bolormaa S, Pryce JE, Kemper K, Savin K, Hayes BJ, Barendse W, et al. Accuracy of prediction of genomic breeding values for residual feed intake and carcass and meat quality traits in *Bos taurus*, *Bos indicus*, and composite beef cattle. *J Anim Sci*. 2013; 91(7):3088–104. Epub 2013/05/10. <https://doi.org/10.2527/jas.2012-5827> PMID: [23658330](https://pubmed.ncbi.nlm.nih.gov/23658330/).

8. Rolf MM, Garrick DJ, Fountain T, Ramey HR, Weaber RL, Decker JE, et al. Comparison of Bayesian models to estimate direct genomic values in multi-breed commercial beef cattle. *Genet Sel Evol.* 2015; 47:23. Epub 2015/04/18. <https://doi.org/10.1186/s12711-015-0106-8> PMID: 25884158; PubMed Central PMCID: PMC4433095.
9. Saatchi M, Schnabel RD, Rolf MM, Taylor JF, Garrick DJ. Accuracy of direct genomic breeding values for nationally evaluated traits in US Limousin and Simmental beef cattle. *Genetics Selection Evolution.* 2012; 44. <https://doi.org/10.1186/1297-9686-44-38> PMID: 23216608
10. Zeng ZY, Tang GQ, Ma JD, Plastow G, Moore S, Lai SJ, et al. Developing a genome-wide selection model for genetic improvement of residual feed intake and carcass merit in a beef cattle breeding program. *Chinese Sci Bull.* 2012; 57(21):2741–6. <https://doi.org/10.1007/s11434-012-5325-6>
11. Lu D, Akanno EC, Crowley JJ, Schenkel F, Li H, De Pauw M, et al. Accuracy of genomic predictions for feed efficiency traits of beef cattle using 50K and imputed HD genotypes. *Journal of Animal Science.* 2016; 94(4):1342–53. <https://doi.org/10.2527/jas.2015-0126> PMID: 27135994
12. Elzo MA, Thomas MG, Martinez CA, Lamb GC, Johnson DD, Rae DO, et al. Genomic-polygenic evaluation of multibreed Angus-Brahman cattle for postweaning feed efficiency and growth using actual and imputed Illumina50k SNP genotypes. *Livest Sci.* 2014; 159:1–10. <https://doi.org/10.1016/j.livsci.2013.11.005>
13. Saatchi M, McClure MC, McKay SD, Rolf MM, Kim J, Decker JE, et al. Accuracies of genomic breeding values in American Angus beef cattle using K-means clustering for cross-validation. *Genetics Selection Evolution.* 2011; 43. <https://doi.org/10.1186/1297-9686-43-40> PMID: 22122853
14. Saatchi M, Ward J, Garrick DJ. Accuracies of direct genomic breeding values in Hereford beef cattle using national or international training populations. *Journal of Animal Science.* 2013; 91(4):1538–51. <https://doi.org/10.2527/jas.2012-5593> PMID: 23345550
15. Wu XL, Sun C, Beissinger TM, Rosa GJ, Weigel KA, Gatti Nde L, et al. Parallel Markov chain Monte Carlo—bridging the gap to high-performance Bayesian computation in animal breeding and genetics. *Genet Sel Evol.* 2012; 44:29. Epub 2012/09/27. <https://doi.org/10.1186/1297-9686-44-29> PMID: 23009363; PubMed Central PMCID: PMC3517397.
16. Gunia M, Saintilan R, Venot E, Hoze C, Fouilloux MN, Phocas F. Genomic prediction in French Charolais beef cattle using high-density single nucleotide polymorphism markers. *Journal of Animal Science.* 2014; 92(8):3258–69. <https://doi.org/10.2527/jas.2013-7478> PMID: 24948648
17. Silva RMO, Fragomeni BO, Lourenco DAL, Magalhaes AFB, Irano N, Carvalheiro R, et al. Accuracies of genomic prediction of feed efficiency traits using different prediction and validation methods in an experimental Nelore cattle population. *Journal of Animal Science.* 2016; 94(9):3613–23. <https://doi.org/10.2527/jas.2016-0401> PMID: 27898889
18. Chen LH, Li CX, Sargolzaei M, Schenkel F. Impact of Genotype Imputation on the Performance of GBLUP and Bayesian Methods for Genomic Prediction. *PLoS One.* 2014; 9(7). <https://doi.org/10.1371/journal.pone.0101544> PMID: 25025158
19. Clark SA, Hickey JM, van der Werf JHJ. Different models of genetic variation and their effect on genomic evaluation. *Genetics Selection Evolution.* 2011;43. <https://doi.org/10.1186/1297-9686-43-18> PMID: 21575265
20. Khatkar MS, Moser G, Hayes BJ, Raadsma HW. Strategies and utility of imputed SNP genotypes for genomic analysis in dairy cattle. *BMC Genomics.* 2012; 13:1–12.
21. He S, Wang S, Fu W, Ding X, Zhang Q. Imputation of missing genotypes from low- to high-density SNP panel in different population designs. *Animal Genetics.* 2015; 46(1):1–7. <https://doi.org/10.1111/age.12236> PMID: 25431355
22. Pausch H, Aigner B, Emmerling R, Edel C, Gotz KU, Fries R. Imputation of high-density genotypes in the Fleckvieh cattle population. *Genetics Selection Evolution.* 2013; 45. <https://doi.org/10.1186/1297-9686-45-3> PMID: 23406470
23. VanRaden PM. Efficient methods to compute genomic predictions. *J Dairy Sci.* 2008; 91(11):4414–23. Epub 2008/10/24. <https://doi.org/10.3168/jds.2007-0980> PMID: 18946147.
24. Habier D, Fernando RL, Kizilkaya K, Garrick DJ. Extension of the bayesian alphabet for genomic selection. *BMC Bioinformatics.* 2011; 12:186. Epub 2011/05/25. <https://doi.org/10.1186/1471-2105-12-186> PMID: 21605355; PubMed Central PMCID: PMC3144464.
25. Erbe M, Hayes BJ, Matukumalli LK, Goswami S, Bowman PJ, Reich CM, et al. Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. *Journal of Dairy Science.* 2012; 95(7):4114–29. <https://doi.org/10.3168/jds.2011-5019> PMID: 22720968
26. Heffner EL, Sorrells ME, Jannink JL. Genomic Selection for Crop Improvement. *Crop Sci.* 2009; 49(1):1–12. <https://doi.org/10.2135/cropsci2008.08.0512>



27. Hayes BJ, Lewin HA, Goddard ME. The future of livestock breeding: genomic selection for efficiency, reduced emissions intensity, and adaptation. *Trends Genet.* 2013; 29(4):206–14. <https://doi.org/10.1016/j.tig.2012.11.009> PMID: 23261029
28. Karp RM. *A Survey of Parallel Algorithms for Shared-Memory Machines.* ACM Digital Library: University of California at Berkeley Berkeley, CA, USA 1988.
29. Zhang Z, Li X, Ding X, Li J, Zhang Q. GPOPSIM: a simulation tool for whole-genome genetic data. *BMC Genet.* 2015; 16:10. Epub 2015/02/06. <https://doi.org/10.1186/s12863-015-0173-4> PMID: 25652552; PubMed Central PMCID: PMC4328615.
30. Zhu B, Zhu M, Jiang J, Niu H, Wang Y, Wu Y, et al. The Impact of Variable Degrees of Freedom and Scale Parameters in Bayesian Methods for Genomic Prediction in Chinese Simmental Beef Cattle. *PLoS One.* 2016; 11(5):e0154118. Epub 2016/05/04. <https://doi.org/10.1371/journal.pone.0154118> PMID: 27139889; PubMed Central PMCID: PMC4854473.
31. Wu Y, Fan HZ, Wang YH, Zhang LP, Gao X, Chen Y, et al. Genome-Wide Association Studies Using Haplotypes and Individual SNPs in Simmental Cattle. *PLoS One.* 2014; 9(10). <https://doi.org/10.1371/journal.pone.0109330> PMID: 25330174
32. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics.* 2007; 81(3):559–75. <https://doi.org/10.1086/519795> PMID: 17701901
33. Wu XL, Beissinger TM, Bauck S, Woodward B, Rosa GJ, Weigel KA, et al. A primer on high-throughput computing for genomic selection. *Front Genet.* 2011; 2:4. Epub 2012/02/04. <https://doi.org/10.3389/fgene.2011.00004> PMID: 22303303; PubMed Central PMCID: PMC3268564.
34. Gelman A, Rubin D. Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science.* 1992; 7(4):457–511.
35. Gregory KE, Cundiff LV, Koch RM, Dikeman ME, Koohmaraie M. Breed effects and retained heterosis for growth, carcass, and meat traits in advanced generations of composite populations of beef cattle. *J Anim Sci.* 1994; 72(4):833–50. Epub 1994/04/01. PMID: 8014148.
36. Orange JSK, Ilatsia ED, Kosgey IS, Kahi AK. Genetic and phenotypic parameters and annual trends for growth and fertility traits of Charolais and Hereford beef cattle breeds in Kenya. *Trop Anim Health Pro.* 2009; 41(5):767–74. <https://doi.org/10.1007/s11250-008-9250-2> PMID: 18975120
37. Strandén I, Lidauer M. Parallel computing applied to breeding value estimation in dairy cattle. *Journal of Dairy Science.* 2001; 84(1):276–85. [https://doi.org/10.3168/jds.S0022-0302\(01\)74477-3](https://doi.org/10.3168/jds.S0022-0302(01)74477-3) PMID: 11210042
38. Luan T, Woolliams JA, Lien S, Kent M, Svendsen M, Meuwissen TH. The accuracy of Genomic Selection in Norwegian red cattle assessed by cross-validation. *Genetics.* 2009; 183(3):1119–26. Epub 2009/08/26. <https://doi.org/10.1534/genetics.109.107391> PMID: 19704013; PubMed Central PMCID: PMC2778964.
39. Fernandes GA, Rosa GJM, Valente BD, Carvalheiro R, Baldi F, Garcia DA, et al. Genomic prediction of breeding values for carcass traits in Nelore cattle. *Genetics Selection Evolution.* 2016; 48. <https://doi.org/10.1186/S12711-016-0188-Y> PMID: 26830208