Genomic prediction has been widely utilized to estimate genomic breeding values (GEBVs) in farm animals. In this study, we conducted genomic prediction for 20 economically important traits including growth, carcass and meat quality traits in Chinese Simmental beef cattle. Five approaches (GBLUP, BayesA, BayesB, BayesC\(p\) and BayesR) were used to estimate the genomic breeding values. The predictive accuracies ranged from 0.159 (lean meat percentage estimated by BayesC\(p\)) to 0.518 (striploin weight estimated by BayesR). Moreover, we found that the average predictive accuracies across 20 traits were 0.361, 0.361, 0.367, 0.367 and 0.378, and the averaged regression coefficients were 0.89, 0.86, 0.89, 0.94 and 0.95 for GBLUP, BayesA, BayesB, BayesC\(p\) and BayesR respectively. The genomic prediction accuracies were mostly moderate and high for growth and carcass traits, whereas meat quality traits showed relatively low accuracies. We concluded that Bayesian regression approaches, especially for BayesR and BayesC\(p\), were slightly superior to GBLUP for most traits. Increasing with the sizes of reference population, these two approaches are feasible for future application of genomic selection in Chinese beef cattle.

Keywords accuracy, Bayesian methods, cross-validation, economic traits, prediction

Introduction
Genomic prediction has been widely utilized to estimate genomic breeding values (GEBVs) for quantitative traits in breeding program of farm animals (Hayes et al. 2009; Goddard et al. 2010). The application of genomic selection is considered an important revolution for the theory of animal breeding over the past two decades (Hayes et al. 2009; Heffner et al. 2009; de Los Campos et al. 2013; Spelman et al. 2013). With the advances of genomic selection technologies, this strategy has led to dramatic increases in genetic progress in farm animals (Goddard et al. 2016; Kumar & Hedges 2016). For instance, rates of genetic gain per year for US Holstein increased from about 50% to 100% for yield traits and from threefold to fourfold for lowly heritable traits (Garcia-Ruiz et al. 2016).

In beef cattle, genomic prediction offers great promise to predict genetic merits of selection candidates, especially for traits that are difficult or expensive to measure, such as carcass merit traits. The success of genomic selection depends on the accuracies of GEBVs, which are largely affected by the predictive approaches, the size of reference population, trait heritability and the extent of the linkage disequilibrium between SNPs and QTL (Hayes et al. 2009; VanRaden et al. 2009). Many studies have assessed the predictive accuracies of GEBVs for economically important traits in different beef cattle populations using the BovineSNP50 BeadChip (Saatchi et al. 2011, 2012; Todd et al. 2014; Chen et al. 2015), and their results show varying degrees of accuracy for GEBVs. For instance, genomic prediction for growth, meat quality and reproduction traits in US Limousin and Simmental beef cattle revealed accuracies of GEBVs ranging from 0.39 to 0.76 and 0.29 to 0.65 respectively (Saatchi et al. 2012). Accuracies of GEBVs for US Angus ranged from 0.22 to 0.69...
Genomic prediction for twenty traits of Simmental cattle

Materials and methods

Ethics statement

All animals were treated following the guidelines for the experimental animals established by the Council of China.

Animals and phenotypes

Animals originated from five farms in Ulgai Grassland, Xilingole League, Inner Mongolia of China. All animals were born between 2008 and 2013 and were weaned at approximately six months of age. After weaning, animals were moved to Beijing Jinweifuren farm for fattening and raised under the same feeding conditions. Animal were measured for growth traits every six or 12 months until slaughter. Live weight was measured after 24 h of fasting. Slaughter age ranged from 18 to 24 months. After slaughter, carcass traits and meat quality traits were measured according to the Institutional Meat Purchase Specification for fresh beef guidelines and GB/T 27643-2011. In this study, a total of 20 traits were measured and analyzed (Table 1): (i) growth traits: average daily gain (ADG; kg) was calculated by subtracting the entering farm weight from the live weight and dividing by the number of days spent in the farm, and live weight (LW; kg) was measured before slaughter with fasting 24 h; (ii) carcass traits: hot carcass weight (CW; kg), dressing percentage (DP; %), lean meat percentage (LMP;%), back fat thickness (BFT; mm) and retail beef cuts including striploin (ST; kg), spencer roll (SR; kg), chuck roll (CR; kg), tenderloin (TP; kg), fore shank (FS; kg), conical muscle (CM; kg), outside (OU; kg), Silverside (ST; kg), knuckle (KN; kg), inside cap off (ICO; kg), hind shank (HS; kg) and retail meat weight (RMW; kg); and (iii) meat quality traits: potential of hydrogen (pH) and shear force (SF; kg).

Genotype and quality control

A total of 1302 Simmental beef cattle were genotyped with the Illumina BovineHD SNP array. Missing SNPs were imputed using BEAGLE v3.3.1 software (Browning & Browning 2007). Prior to statistical analysis, genotypes were edited using PLINK v1.07 software (Purcell et al. 2007) for the following: (i) minor allele frequency (>0.05), (ii) proportion of missing genotypes (<0.05) and (iii) Hardy-Weinberg equilibrium (P > 10^{-6}). SNPs satisfying these criteria were used to interrogate the linkage disequilibrium with syntenic SNPs located with a window of 100 neighboring markers, which resulted in only one SNP form each pair of highly correlated SNPs (r² > 0.995) remaining in the SNP dataset. After these quality controls, the final data consisted of 1217 individuals and 459 268 filtered SNPs in the autosomes. Genotype data are available from the Dryad Digital Repository (https://doi.org/10.5061/dryad.4qc06).

Statistical model and genetic analyses

Heritabilities were calculated using a restricted maximum likelihood method with an animal model in ASREML v3.0 software (Gilmour et al. 2009). Relationships between animals were estimated using a G matrix, where G was the genomic relationship matrix and inferred from SNP markers, as suggested by VanRaden (2008). The animal model included random additive polygenic effects, fixed effects and residual for all traits. The additive polygenic effects were treated as random and assumed to be mutually independent. Fixed effects in the model included gender, farm and year of measurement. In addition, animals’ age at

© 2019 The Authors. Animal Genetics published by John Wiley & Sons Ltd on behalf of Stichting International Foundation for Animal Genetics, 50, 634–643
slaughter were considered covariates in the model except for average daily gain traits.

To estimate GEBVs, we used the following five statistical methods: (i) GBLUP (VanRaden 2008) and (ii) Bayesian regression using mixture models, including BayesA and BayesB (Meuwissen et al. 2001). BayesCq (Habier et al. 2011) and a variable selection method BayesR (Erbe et al. 2012). These methods were compiled using C language (Zhu et al. 2016).

In all cases, the statistical model used was the following:

$$y = Xb + Zg + e$$

**GBLUP**

For GBLUP, y is an N × 1 (N = number of observations) vector of phenotype in Equation 1. X is an incidence matrix of the fixed effects, Z is the incidence matrix allocating records to GEBVs, g is the vector of GEBVs and e is a vector of residuals. It is assumed that g follows a normal distribution N(0, \(G\sigma^2_g\)). Given b and g, y is conditionally independent and distributed as:

$$(y | b, g, \sigma^2) \sim N(Xb + Zg, \sigma^2_e)$$

**BayesA, BayesB and BayesCq**

The Bayesian model used the same equation as in (1), where y, X, b and e were defined as before, but g represented an M × 1 vector of SNP marker effects; Z is an N × M matrix of SNPs, coded with values 0, 1 or 2 for genotypes 11, 12 and 22 respectively; and \(Z_{ij}\) denotes marker j of individuals i.

The prior for \(g_j\) depends on the variance \(\sigma^2_j\) and the prior probability \(\pi\). In BayesA, all SNPs have effects, i.e. \(\sigma^2_j \sim x^{-2}(v, \nu \sigma^2_1)\), and \(\pi\) is equal to 0. \(\sigma^2_j\) denotes that SNP j has its own variance, with the parameters of \(v\) and \(\sigma^2_1\). In BayesB, the two-component mixture, with one component being \((0, v, \nu \sigma^2_1)\) and the other component being a spike at 0, are provided as:

$$g_j | \sigma^2_j \sim (\text{idd}) \left\{ \begin{array}{ll} \text{0} & \text{x0026; with probability } \pi \\ \text{N}(0, \sigma^2_j) & \text{x0026; with probability } (1 - \pi) \end{array} \right.$$  

where \(j = 1, \ldots, P\).

Here, \(\pi\) represents the proportion of SNPs with no genetic effects on the trait of interest. \(\sigma^2_b\) is derived using following equation, \(\sigma^2_b = \frac{(v - 2) + 2 \sigma_1^2}{v}\), where \(v\) is 4.2, as reported by Meuwissen et al. (2001). In BayesCq, the SNP effects have a common variance, \(\sigma^2_q = \sigma^2_j (j = 1, \ldots, P)\), and \(\sigma^2_q\) follows a scaled inverse \(\chi^2\) prior with parameters \(v\) and \(\nu \sigma^2_1\). As a result, the SNP effects with probability \((1 - \pi)\) follows a mixture of multivariate Student’s t-distributions \(t(0, v, IS^2_q)\). The parameter \(\pi\) is treated as unknown with a uniform (0,1) prior distribution.

**BayesR**

BayesR is an extension of BayesCq, where SNP effects are assumed to be sampled from a mixture of normal distributions (Erbe et al. 2012; Bolormaa et al. 2013). The variance of each component of the mixture is fixed (0, 0.01%, 0.1% or 1% of the genetic variance). The number of SNPs belonging to each component of the mixture is assumed to come from a multinomial distribution with proportions \(p_i\) (i = 1, 2, 3 or 4) in which the

© 2019 The Authors. *Animal Genetics* published by John Wiley & Sons Ltd on behalf of Stichting International Foundation for Animal Genetics, 50, 634–643
is drawn from a Dirichlet distribution (a multivariate generalization of a beta distribution) with pseudo-counts of 1 for each component of the mixture. Thus, the prior assumes that the four components of the mixture are equally probable but with minimal prior knowledge of these probabilities.

SNP effects estimation

SNP effects were estimated using the Markov chain Monte Carlo sampling algorithm in BayesA, BayesB, BayesC\(\alpha\) and BayesR. Markov chains were run for 50,000 cycles of Gibbs sampling. The first 10,000 were discarded as burn-in. Then GEBVs were calculated as \(\text{GEBV}_i = \sum Z_g \alpha_j\).

Validation of the models

To assess the predictive accuracies for 20 economically important traits, we used a five-fold cross-validation method (Luan et al. 2009). Overall, 1,217 individuals were divided into reference and validation populations. The genotyped individuals were randomly divided into five groups, whereas phenotypes of animals in validation set were masked to be unknown. Thus, 974 individuals were randomly sampled as the reference set, and the remaining 243 individuals as the validation set. The whole procedure was repeated 10 times.

Comparison criteria

Three methods were utilized to evaluate the predictive ability based on comparison of GEBVs with corrected phenotypes of animals in the validation population:

The correlations between GEBV and corrected phenotype were calculated to evaluate the predictive ability \(r_{\text{GEBV}, \hat{y}}\). To remove the influence of heritability on the predictive ability, Pearson’s correlation between GEBV and corrected phenotype was divided by the square root of heritability \(\frac{r_{\text{GEBV}, \hat{y}}}{\sqrt{h^2}}\), where \(\hat{y}\) was the corrected phenotype. This value is approximated to the correlation between the true breeding value and GEBVs (Pryce et al. 2012).

The slope of the regression of \(\hat{y}\) on GEBV for animals in the validation population \((b_\hat{y}, \text{GEBV})\) was calculate to measure the degree of inflation or deflation of genomic prediction. Estimates of \(b_\hat{y}, \text{GEBV}\) close to 1 are indicative of predictions that are similar to that of corrected phenotype on scale.

The mean squared error of prediction (MSE) between \(\hat{y}\) and GEBV in the validation population was used to measure the overall fit of model, and the computation equation was \(\text{MSE} = \frac{1}{N} \sum (\text{GEBV} - \hat{y})^2\), where \(N\) is the number of individuals. A large estimated value of predictive accuracy is indicative of reliable prediction, and a low MSE value means a better overall fit.

Results

Heritability estimates

Heritability estimates for 20 traits ranged from 0.04 to 0.62. We found that 10 traits showed relatively high heritabilities: FS (0.4), LW (0.43), RMW (0.43), ADG (0.47), BI (0.47), ICO (0.51), ER (0.52), OU (0.6), HS (0.61) and KN (0.62). Eight traits showed moderate heritabilities—BFI (0.21), ST (0.24), SR (0.26), CR (0.27), DP (0.31), CW (0.38), LMP (0.39) and TD (0.39)—whereas three traits had low heritabilities—pH (0.04), BFT (0.1) and SF (0.15). The standard errors for all heritability estimates were less than 0.04 (Table 2).

Predictive ability of five methods

We evaluated the predictive abilities for these 20 traits using different methods based on 5-fold cross-validation. The summary of predictive results is presented in Table 3. Overall, we observed that the predictive abilities \(r_{\text{GEBV}, \hat{y}}\) ranged from 0.059 (for LMP estimated by BayesC\(\alpha\)) to 0.376 (for HS estimated by BayesB; Table 3). The average predictive abilities across 20 traits were 0.216, 0.216, 0.221, 0.220 and 0.225 for GBLUP, BayesA, BayesB, BayesC\(\alpha\) and BayesR respectively.

Table 2 Estimates of variance components and heritability with standard errors for twenty traits

| Trait   | \(\sigma_a^2\) | \(\sigma_e^2\) | \(h^2\) |
|---------|----------------|----------------|---------|
| ADG     | 0.012          | 0.013          | 0.47 ± 0.03 |
| LW      | 663.380        | 1139.880       | 0.37 ± 0.02 |
| CW      | 299.650        | 361.820        | 0.45 ± 0.02 |
| DP      | 0.950          | 4.960          | 0.16 ± 0.02 |
| LMP     | 0.850          | 5.110          | 0.14 ± 0.01 |
| ST      | 0.380          | 1.160          | 0.24 ± 0.02 |
| SR      | 0.490          | 1.390          | 0.26 ± 0.01 |
| CR      | 1.080          | 2.880          | 0.27 ± 0.02 |
| TD      | 0.078          | 0.120          | 0.39 ± 0.03 |
| FS      | 0.150          | 0.220          | 0.40 ± 0.02 |
| CM      | 0.008          | 0.009          | 0.47 ± 0.02 |
| OU      | 1.320          | 0.880          | 0.60 ± 0.04 |
| ER      | 1.200          | 1.130          | 0.52 ± 0.03 |
| KN      | 0.590          | 0.360          | 0.62 ± 0.04 |
| ICO     | 0.910          | 0.880          | 0.51 ± 0.03 |
| HS      | 0.370          | 0.230          | 0.61 ± 0.02 |
| RMW     | 126.130        | 169.570        | 0.43 ± 0.02 |
| BFT     | 0.330          | 2.940          | 0.10 ± 0.01 |
| pH      | 0.005          | 0.112          | 0.04 ± 0.01 |
| SF      | 0.340          | 2.000          | 0.15 ± 0.01 |

\(\sigma_a^2\), additive genetic variance; \(\sigma_e^2\), environmental variance; \(h^2\), heritability ± standard error; ADG, BFT, back fat thickness; CM, conical muscle; CR, chuck roll; CW, carcass weight; DP, dressing percentage; FS, fore shank; ICO, inside cap off; HS, hind shank; KN, knuckle; LMP, lean meat percentage; LW, live weight; OU, outside; pH, potential of hydrogen; RMW, retail meat weight; SF, shear force; SI, silverside; SR, Spencer roll; ST, striploin; TD, tenderloin.
To investigate the relationship between predictive ability and trait heritability, we next estimated the regression coefficient of predictive ability on heritability for these methods. Our results showed that the predictive abilities were linearly correlated with trait heritability. The strongest correlation (0.5) was observed across 20 traits using GBLUP compared with Bayesian methods (Fig. 1). Among Bayesian methods, we found the highest regression coefficient of predictive ability on heritability was 0.47 in BayesR.

**Predictive accuracy of five methods**

After removing the influence of heritability from predictive ability, we found obvious difference in predictive accuracies ranging from 0.159 (LMP) to 0.518 (ST). The average predictive accuracies across traits were 0.361 for GBLUP, 0.361 for BayesA, 0.367 for BayesB, 0.367 for BayesCp and 0.378 for BayesR respectively (Table 3). In general, we observed the predictive accuracies of BayesA were consistent with those of GBLUP for most traits, whereas other Bayesian regression methods (BayesB, BayesCp and BayesR) showed slightly higher accuracies than did GBLUP. The advantage of Bayesian regression methods over GBLUP was obtained for HS (6.5%), CW (1.5%) and ST (3.1%) using BayesB, BayesCp and BayesR. Conversely, GBLUP was slightly superior to BayesA, BayesB, BayesCp and BayesR for KN (1.5%), RMW (0.9%), KN (0.4%) and KN (0.5%) in the present study (Table 3).

**Scale of genomic predictions and mean squared prediction error**

The regression coefficient of corrected phenotype on GEBV was calculated as a measurement of the bias for the prediction. In this study, we observed that predictions of GEBV using both GBLUP and Bayesian regression methods were inflated for most traits, whereas for traits ST, TD and pH, the predictions from GBLUP and Bayesian regression methods tended to be slightly deflated (Table 4). We found that predictions of GEBV using GBLUP, BayesA and BayesB were inflated for LW and RMW, whereas those from BayesCp and BayesR were slightly deflated. The average regression coefficients across traits were 0.89, 0.86, 0.89, 0.94 and 0.95 for GBLUP, BayesA, BayesB, BayesCp and BayesR respectively.

For most traits, we found GBLUP generally outperformed Bayesian regression methods based on the MSE (Table 4). However, for ADG, LMP, TD, FS, BI, KN, HS and RMW, lower estimates of MSE were obtained for Bayesian methods.

| Trait  | $r_{\text{GEBV}}$ | $r_{\text{GEBV}}/\sqrt{h^2}$ |
|--------|-----------------|-----------------|
|        | GBLUP | BayesA | BayesB | BayesCp | BayesR | GBLUP | BayesA | BayesB | BayesCp | BayesR |
| ADG    | 0.194 | 0.197 | 0.197 | 0.204 | 0.214 | 0.283 | 0.287 | 0.288 | 0.298 | 0.312 |
| LW     | 0.231 | 0.232 | 0.230 | 0.239 | 0.242 | 0.379 | 0.381 | 0.378 | 0.393 | 0.398 |
| CW     | 0.251 | 0.253 | 0.252 | 0.261 | 0.268 | 0.374 | 0.377 | 0.376 | 0.389 | 0.400 |
| DP     | 0.111 | 0.111 | 0.112 | 0.109 | 0.119 | 0.277 | 0.276 | 0.279 | 0.273 | 0.298 |
| LMP    | 0.061 | 0.061 | 0.059 | 0.069 | 0.119 | 0.162 | 0.164 | 0.162 | 0.159 | 0.184 |
| ST     | 0.239 | 0.238 | 0.241 | 0.239 | 0.254 | 0.487 | 0.486 | 0.492 | 0.487 | 0.518 |
| SR     | 0.178 | 0.179 | 0.176 | 0.177 | 0.184 | 0.349 | 0.352 | 0.345 | 0.348 | 0.361 |
| CR     | 0.168 | 0.169 | 0.169 | 0.169 | 0.176 | 0.322 | 0.326 | 0.324 | 0.325 | 0.339 |
| TD     | 0.277 | 0.278 | 0.278 | 0.283 | 0.291 | 0.444 | 0.446 | 0.446 | 0.453 | 0.466 |
| FS     | 0.249 | 0.250 | 0.254 | 0.250 | 0.252 | 0.394 | 0.395 | 0.402 | 0.395 | 0.398 |
| CM     | 0.240 | 0.240 | 0.250 | 0.247 | 0.251 | 0.35 | 0.351 | 0.365 | 0.360 | 0.366 |
| OU     | 0.346 | 0.347 | 0.347 | 0.352 | 0.358 | 0.447 | 0.448 | 0.447 | 0.455 | 0.462 |
| ER     | 0.354 | 0.353 | 0.355 | 0.358 | 0.361 | 0.491 | 0.49 | 0.492 | 0.497 | 0.501 |
| KN     | 0.315 | 0.304 | 0.339 | 0.312 | 0.311 | 0.400 | 0.385 | 0.431 | 0.396 | 0.395 |
| ICO    | 0.258 | 0.259 | 0.259 | 0.269 | 0.268 | 0.362 | 0.363 | 0.363 | 0.377 | 0.375 |
| HS     | 0.325 | 0.326 | 0.376 | 0.332 | 0.331 | 0.416 | 0.417 | 0.481 | 0.425 | 0.424 |
| RMW    | 0.267 | 0.268 | 0.262 | 0.271 | 0.278 | 0.408 | 0.409 | 0.399 | 0.414 | 0.424 |
| BFT    | 0.074 | 0.072 | 0.074 | 0.076 | 0.077 | 0.235 | 0.227 | 0.233 | 0.241 | 0.243 |
| pH     | 0.073 | 0.073 | 0.074 | 0.074 | 0.074 | 0.366 | 0.365 | 0.367 | 0.370 | 0.390 |
| SF     | 0.109 | 0.107 | 0.107 | 0.114 | 0.119 | 0.280 | 0.277 | 0.277 | 0.294 | 0.307 |

Note: Predictive ability ($r_{\text{GEBV}}$) was calculated by the correlation between GEBV and corrected phenotype. Predictive accuracy ($r_{\text{GEBV}}/\sqrt{h^2}$) was computed as Pearson’s correlation between GEBV and corrected phenotype divided by square root of heritability.

ADG, average daily gain; BFT, back fat thickness; CM, conical muscle; CR, chuck roll; CW, carcass weight; DP, dressing percentage; FS, fore shank; HS, hind shank; ICO, inside cap off; KN, knuckle; LMP, lean meat percentage; LW, live weight; OU, outside; pH, potential of hydrogen; RMW, retail meat weight; SF, shear force; SI, silverside; SR, Spencer roll; ST, striploin; TD, tenderloin.
Discussion

Previous studies have been conducted for genomic prediction in multiple breeds including Angus, Limousin, Simmental, Charolais, Hereford, Japanese Black, Nellore and other crossbreds using BovineSNP50 and BovineHD SNP arrays (Saatchi et al. 2011, 2012; Bolormaa et al. 2013; Akanno et al. 2014; Gunia et al. 2014; Hulsman Hanna et al. 2014; Onogi et al. 2014; Todd et al. 2014; Rolf et al. 2015; Fernandes Jr. et al. 2016). Su et al. (2012) have investigated the difference of predictive accuracies between the BovineHD array and BovineSNP50 using the GBLUP method, and they found that the reliability of GEBV for protein, fertility and udder health traits using the BovineHD array was higher (0.5–1%) than that of 54K array in Holstein.

We previously investigated the pattern of linkage disequilibrium using the BovineHD SNP array in Chinese Simmental cattle, and our findings suggested that the high density SNP array was sufficient to achieve high accuracy for genomic prediction in Chinese Simmental population (Niu et al. 2016). To our knowledge, this study is the first attempt to investigate the performance of genomic prediction for 20 economically important traits using BovineHD SNP arrays in Chinese Simmental beef cattle.

Comparisons of genomic prediction methods

In the current study, we found that the Bayesian regression approaches performed better than did GBLUP for most traits. Based on the estimations of predictive accuracies and regression coefficients, Bayesian regression methods were superior to GBLUP, whereas GBLUP had smaller MSE compared to Bayesian regression approaches.

Previous studies suggested the superiority of Bayesian regression approaches over GBLUP when the number of SNPs is larger than the genotyped animals, i.e. several simulation studies revealed that Bayesian regression
approaches have higher accuracies than does GBLUP (Meuwissen et al. 2001; Habier et al. 2007; Solberg et al. 2008; Clark et al. 2011). These findings were also consistent with many previous studies using real data (Erbe et al. 2012; Pryce et al. 2012; Gunia et al. 2014; Neves et al. 2014; Fernandes Jr. et al. 2016). In this study, we found that genomic predictions using Bayesian regression approaches were superior to that of GBLUP (Table 3). This can be explained by the fact that the assumption of Bayesian approaches is more suitable for fitting the genetic architecture of quantitative trait (Rolf et al. 2015). However, most of the Bayesian regression approaches, except for BayesA, showed high prediction accuracies for these 20 traits. A large number of SNPs with small effects in BayesA are likely to cause noise for the estimation of the GEBVs (Habier et al. 2011). In contrast, Bayesian approaches like BayesB, BayesCr and BayesR assume that only a small proportion of markers have effects, which may avoid the potential bias caused by linkage disequilibrium (Erbe et al. 2012).

Moreover, we observed higher predictive accuracies using BayesR for most of the traits, which implies the segregation of genes with larger effects for them. For traits with mutations of moderate effect segregating and a high number of significant SNPs, a recent study showed that the accuracy of GEBVs with BayesR was higher than with GBLUP (Bolormaa et al. 2013). In our study, the average predictive accuracies of 20 traits for BayesR increased ~1.7% compared to GBLUP. Among 20 traits, the highest increase in accuracy of BayesR over GBLUP was observed for ADG (2.9%) and ST (3.1%). We also identified several SNPs with large effects for ADG using BayesR. These SNPs had been previously identified as significant associated SNPs in the gene NCAPG and can explain ~4.01% of the phenotypic variances (Zhang et al. 2016).

### Predictive abilities and accuracies

The accuracy of GEBVs can be affected by trait heritabilities, size of training population, and breed and statistical method (Bolormaa et al. 2013). For example, traits with high heritability (\( h^2 \)) and a large training population (\( n = T \)) give higher accuracies, which can be expected from the theory that \( Th^2 \) is a critical parameter (de Roos et al. 2008). In our study, we found that Bayesian regression approaches outperformed GBLUP in predictive accuracies for most traits. Moreover, the accuracy of GEBVs for several important traits varied in other cattle populations. For instance, Neves et al. (2014) presented the results of implementation of genomic prediction for weight and carcass traits, gestation length and scrotal circumference traits in 685 Nellore cattle. They found that the average accuracy was 0.39 for GBLUP and 0.44 for both BayesC and Bayesian LASSO methods, which was higher than in our study. Under data-splitting strategies by birth year, Chen et al. (2015) accessed the predictive accuracies for hot carcass weight via PBLUP (0.33 for Angus, 0.42 for Charolais), GBLUP (0.34 for Angus, 0.18 for Charolais) and BayesB (0.34 for Angus, 0.21 for Charolais) using the BovineSNP50 Beadchip. Using

### Table 4

| Trait | b(ŷ, GEBV) | MSE |
|-------|------------|-----|
|       | GBLUP | BayesA | BayesB | BayesCr | BayesR | GBLUP | BayesA | BayesB | BayesCr | BayesR |
| ADG   | 0.63   | 0.67   | 0.70   | 0.93    | 0.96    | 0.16   | 0.17   | 0.17   | 0.16    | 0.15    |
| LW    | 0.84   | 0.85   | 0.83   | 1.08    | 1.04    | 45.48  | 47.16  | 46.21  | 46.51   | 46.54   |
| CW    | 0.72   | 0.74   | 0.76   | 0.94    | 0.95    | 27.26  | 29.99  | 29.79  | 28.73   | 28.94   |
| DP    | 0.87   | 0.88   | 0.91   | 0.65    | 0.88    | 2.34   | 2.40   | 2.41   | 2.54    | 2.48    |
| LMP   | 0.28   | 0.31   | 0.27   | 0.47    | 0.43    | 2.44   | 2.49   | 2.47   | 2.42    | 2.46    |
| ST    | 1.56   | 1.58   | 1.59   | 1.76    | 1.79    | 19.00  | 19.41  | 19.44  | 22.08   | 20.12   |
| SR    | 0.82   | 0.82   | 0.86   | 0.85    | 0.88    | 1.42   | 1.45   | 1.45   | 1.45    | 1.45    |
| CR    | 0.59   | 0.61   | 0.62   | 0.70    | 0.76    | 1.55   | 1.55   | 1.55   | 1.55    | 1.56    |
| TD    | 1.01   | 1.04   | 1.09   | 1.19    | 1.21    | 2.36   | 2.42   | 2.37   | 2.36    | 2.35    |
| FS    | 0.94   | 0.95   | 0.97   | 0.98    | 0.99    | 0.49   | 0.49   | 0.53   | 0.49    | 0.48    |
| BI    | 0.96   | 1.00   | 1.07   | 1.27    | 1.25    | 0.88   | 0.88   | 0.89   | 0.88    | 0.87    |
| OU    | 0.83   | 0.84   | 0.87   | 0.92    | 0.94    | 0.18   | 0.20   | 0.19   | 0.19    | 0.18    |
| ER    | 0.78   | 0.77   | 0.78   | 0.81    | 0.87    | 1.65   | 1.84   | 1.92   | 1.91    | 1.89    |
| KN    | 0.67   | 0.67   | 0.77   | 0.78    | 0.79    | 1.65   | 1.64   | 2.12   | 1.75    | 1.96    |
| ICO   | 0.70   | 0.72   | 0.74   | 0.90    | 0.85    | 1.10   | 1.11   | 1.45   | 1.12    | 1.35    |
| HS    | 0.75   | 0.76   | 0.86   | 0.87    | 0.89    | 1.41   | 1.41   | 1.55   | 1.40    | 1.48    |
| RMW   | 0.94   | 0.95   | 0.97   | 1.15    | 1.02    | 0.86   | 0.87   | 0.84   | 0.92    | 0.89    |
| BFT   | 0.45   | 0.37   | 0.45   | 0.50    | 0.48    | 0.96   | 0.96   | 0.96   | 0.96    | 0.97    |
| pH    | 1.98   | 1.95   | 2.08   | 1.33    | 1.32    | 1.71   | 1.71   | 1.75   | 1.72    | 1.74    |
| SF    | 1.39   | 0.64   | 0.63   | 0.79    | 0.76    | 1.30   | 1.30   | 1.31   | 1.31    | 1.31    |

ADG, average daily gain; BFT, back fat thickness; CM, conical muscle; CR, chuck roll; CW, carcass weight; DP, dressing percentage; FS, fore shank; HS, hind shank; ICO, inside cap off; KN, knuckle; LMP, lean meat percentage; LW, live weight; OU, outside; pH, potential of hydrogen; RMW, retail meat weight; SF, shear force; SI, silverside; SR, Spencer roll; ST, striploin; TD, tenderloin.
five-fold cross-validation, we found that the predictive accuracy of carcass weight with Bovine HD SNP array was higher than those in Chen et al.’s (2015) study. Also, Bolormaa et al. (2013) found that the accuracy of genomic prediction for ADG and CW using the GBLUP approach in multiple beef cattle populations were 0.21 and 0.27. Their results also suggested relatively low accuracy compared to the results in our study (ADG, 0.312 and CW, 0.40). Using K-means clustering validation, the predictive abilities of CW, BFT and SF were 0.59, 0.29 and 0.53 in 2703 registered Simmental beef cattle (Saatchi et al. 2012). Their results revealed higher predictive ability for CW, BFT and SF compared with our study, and this finding is likely to be explained by the large population utilized in their study. In addition, other studies have performed genomic prediction for the ADG, CW, SF and BFT traits in diverse populations (Akanno et al. 2014; Rolf et al. 2015; Fernandes Jr. et al. 2016). However, no study has been reported for genomic prediction of DP, LMP and primal cuts in beef cattle.

Scale of genomic predictions

The scale of predictions is an important factor in determining whether GEBVs can be used for genetic evaluation. For instance, the results of one previous study suggested that overestimation of the genetic merit may cause potential exaggeration of GEBVs compared with traditional EBVs when both progeny-tested and genomic selection were used for selecting candidates (Vitezica et al. 2011). In our study, the average regression coefficients across traits were 0.89, 0.86, 0.89, 0.94 and 0.95 for GBLUP, BayesA, BayesB, BayesC and BayesR respectively (Table 4). This finding indicates that the BayesR and BayesC approaches generate more reliable predictions for these traits in Chinese Simmental population.

The estimation of scale of genomic predictions may vary with population, genetic inherent of the studied trait and the statistical approaches (Neves et al. 2014). Among our studied traits in Chinese Simmental beef cattle, 13 traits generated inflated prediction using GBLUP and Bayesian approaches. Many studies showed a similar inflation trend of genomic prediction in Nellore, Nellore-Angus crossbred, Holstein and Jersey populations using Bayesian approaches (Duchemin et al. 2012; Erbe et al. 2012; Hulsman Hanna et al. 2014), whereas other studies revealed opposite trends in American Angus using GBLUP and in French Holstein and Montbeliarde populations using Bayesian approaches (Saatchi et al. 2011; Colombani et al. 2013).

Conclusions

Using multiple methods (GBLUP, BayesA, BayesB, BayesC and BayesR), we conducted genomic prediction for economically important traits including growth, carcass (especially on retail beef cuts) and meat quality traits in Chinese Simmental cattle. Bayesian regression approaches, especially BayesR and BayesC, were superior to GBLUP for most traits. Thus, it may be feasible to apply these approaches for genomic prediction of these economically important traits in Chinese Simmental beef cattle. Further improvements are required to enlarge population size and reduce inflation of predictions. In addition, our findings provide valuable insights for further implementation of genomic selection in the commercial beef industry.

Acknowledgements

This work was supported by National Natural Science Foundation of China (31802049, 31372294 and 31201782), Chinese Academy of Agricultural Sciences of Technology Innovation Project (CAAS-XTCX2016010, CAAS-ZDXT2018006, ASTIP-IAS-TS-16 and ASTIP-IAS03), Cattle Breeding Innovative Research Team of Chinese Academy of Agricultural Sciences (cxgc-ias-03 and Y2016PT17), Beijing Natural Science Foundation (6154032) and Project of College Innovation Improvement under Beijing Municipality (PXM2016_014207_000012).

Conflict of interest

The authors have no conflict of interest to declare.

Data availability

Datasets are available from the Dryad Digital Repository (https://doi.org/10.5061/dryad.4qc06).

References

Akanno E.C., Plastow G., Woodward B.W. et al. (2014) Reliability of molecular breeding values for Warner-Bratzler shear force and carcass traits of beef cattle – an independent validation study. Journal of Animal Science 92, 2896–904.

Bolormaa S., Pryce J.E., Kemper K. et al. (2013) 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. Journal of Animal Science 91, 1084–104.

Browning S.R. & Browning B.L. (2007) Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. American Journal of Human Genetics 81, 1084–97.

Chen L., Vinsky M. & Li C. (2015) Accuracy of predicting genomic breeding values for carcass merit traits in Angus and Charolais beef cattle. Animal Genetics 46, 55–9.

Clark S.A., Hickey J.M. & van der Werf J.H. (2011) Different models of genetic variation and their effect on genomic evaluation. Genetics Selection Evolution 43, 18.

Colombani C., Legarra A., Fritz S., Guillaume F., Croiseau P., Ducrocq V. & Robert-Granie C. (2013) Application of Bayesian least absolute shrinkage and selection operator (LASSO) and BayesCpi methods for genomic selection in French Holstein and Montbeliarde breeds. Journal of Dairy Science 96, 575–91.
Duchemin S.I., Colombani C., Legarra A., Baloche G., Larroque H., Astruc J.M., Barillet F., Robert-Granier C. & Manfredi E. (2012) Genomic selection in the French Lacaune sheep breed. *Journal of Dairy Science* 95, 2723–31.

Erbe M., Hayes B.J., Matukumalli L.K., Goswami S., Bowman P.J., Reich C.M., Mason B.A. & Goddard M.E. (2012) Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. *Journal of Dairy Science* 95, 4114–29.

Fernandes Jr., G.A., Rosa G.J., Valente B.D. *et al.* (2016) Genomic prediction of breeding values for carcass traits in Nellore cattle. *Genetics Selection Evolution* 48, 7.

Garcia-Ruiz A., Cole J.B., VanRaden P.M., Wiggans G.R., Ruiz-Lopez F.J. & Van Tassell C.P. (2016) Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection. *Proceedings of the National Academy of Sciences of the United States of America* 113, E3995–4004.

Gilmour A.R., Gogel B.J., Cullis B.R. & Thompson R. (2009) *ASREML User Guide Release 3.0.* VSN International Ltd, Hemel Hempstead.

Goddard M.E., Hayes B.J. & Meuwissen T.H. (2009) Genomic selection in livestock populations. *Genetics Research* 92, 413–21.

Goddard M.E., Kemper K.E., MacLeod I.M., Chamberlain A.J. & Hayes B.J. (2016) Genetics of complex traits: prediction of phenotype, identification of causal polymorphisms and genetic architecture. *Proceedings of the Royal Society B: Biological Sciences* 283, 20160569.

Gunia M., Saintilan R., Venot E., Hoze C., Fouilloux M.N. & Phocas F. (2014) Genomic prediction in French Charolais beef cattle using high-density single nucleotide polymorphism markers. *Journal of Animal Science* 92, 3258–69.

Habier D., Fernando R.L. & Dekkers J.C. (2007) The impact of genetic relationship information on genome-assisted breeding values. *Genetics* 177, 2389–97.

Habier D., Fernando R.L., Kadiyala K. & Garrick D.J. (2011) Extension of the Bayesian alphabet for genomic selection. *BMC Bioinformatics* 12, 186.

Hayes B.J., Bowman P.J., Chamberlain A.J. & Goddard M.E. (2009) Invited review: genomic selection in dairy cattle: progress and challenges. *Journal of Dairy Science* 92, 433–43.

Hoffner E.L., Sorrells M.E. & Jannink J.L. (2009) Genomic selection for crop improvement. *Crop Science* 49, 1.

Hulsman Hanna L.L., Garrick D.J., GB C.A., Herring A.D., Sanders J.O. & Riley D.G. (2014) Comparison of breeding value prediction for two traits in a Nellore- Angus crossbred population using different Bayesian modeling methodologies. *Genetics and Molecular Biology* 37, 631–7.

Kumar S. & Hedges S.B. (2016) Advances in time estimation methods for molecular data. *Molecular Biology and Evolution* 33, 863–9.

de Los Campos G., Hickey J.M., Pong-Wong R., Daetwyler H.D. & Calus M.P. (2013) Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics* 193, 327–45.

Luan T., Woolliams J.A., Lien S., Kent M., Svendsen M. & Meuwissen T.H. (2009) The accuracy of genomic selection in Norwegian red cattle assessed by cross-validation. *Genetics* 183, 1119–26.

Meuwissen T.H., Hayes B.J. & Goddard M.E. (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819–29.

Neves H.H., Carvalheiro R., O’Brien A.M. *et al.* (2014) Accuracy of genomic predictions in Bos indicus (Nellore) cattle. *Genetics Selection Evolution* 46, 17.

Niu H., Zhu B., Guo P. *et al.* (2016) Estimation of linkage disequilibrium levels and haplotype block structure in Chinese Simmental and Wagyu beef cattle using high-density genotypes. *Livestock Science* 190, 1–9.

Onogi A., Ogino A., Komatsu T., Shoji N., Simizu K., Kurogi K., Yasumori T., Togashi K. & Iwata H. (2014) Genomic prediction in Japanese Black cattle: application of a single-step approach to beef cattle. *Journal of Animal Science* 92, 1931–8.

Pryce J.E., Arias J., Bowman P.J., Davis S.R., Macdonald K.A., Wughorn G.C., Wales W.J., Williams Y.J., Spelman R.J. & Hayes B.J. (2012) Accuracy of genomic predictions of residual feed intake and 250-day body weight in growing heifers using 625,000 single nucleotide polymorphism markers. *Journal of Dairy Science* 95, 2108–19.

Purcell S., Neale B., Todd-Brown K. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics* 81, 559–75.

Rolf M.M., Garrick D.J., Fountain T., Ramey H.R., Weaber R.L., Decker J.E., Pollak E.J., Schnabel R.D. & Taylor J.F. (2015) Comparison of Bayesian models to estimate direct genomic values in multi-breed commercial beef cattle. *Genetics Selection Evolution* 47, 23.

de Roos A.P., Hayes B.J., Spelman R.J. & Goddard M.E. (2008) Linkage disequilibrium and persistence of phase in Holstein-Friesian, Jersey and Angus cattle. *Genetics* 179, 1503–12.

Saatchi M., McClure M.C., McKay S.D. *et al.* (2011) Accuracies of genomic breeding values in American Angus beef cattle using K-means clustering for cross-validation. *Genetics Selection Evolution* 43, 40.

Saatchi M., Schnabel R.D., Rolf M.M., Taylor J.F. & Garrick D.J. (2012) Accuracy of direct genomic breeding values for nationally evaluated traits in US Limousin and Simmental beef cattle. *Genetics Selection Evolution* 44, 38.

Solberg T.R., Sonesson A.K., Woolliams J.A. & Meuwissen T.H. (2008) Genomic selection using different marker types and densities. *Journal of Animal Science* 86, 2447–54.

Spelman R.J., Hayes B.J. & Berry D.P. (2013) Use of molecular technologies for the advancement of animal breeding: genomic selection in dairy cattle populations in Australia, Ireland and New Zealand. *Animal Production Science* 53, 869–75.

Su G., Brondon R.F., Ma P., Guldbrandt B., Aamand G.P. & Lund M.S. (2012) Comparison of genomic predictions using medium-density (approximately 54,000) and high-density (approximately 777,000) single nucleotide polymorphism marker panels in Nordic Holstein and Red Dairy Cattle populations. *Journal of Dairy Science* 95, 4657–65.

Todd D.L., Roughsedge T. & Woolliams J.A. (2014) Genomic selection using beef commercial carcase phenotypes. *Animal* 8, 388–94.

VanRaden P.M. (2008) Efficient methods to compute genomic predictions. *Journal of Dairy Science* 91, 4414–23.

VanRaden P.M., Van Tassell C.P., Wiggans G.R., Sonstegard T.S., Schnabel R.D., Taylor J.F. & Schenkel F.S. (2009) Invited review: reliability of genomic predictions for North American Holstein bulls. *Journal of Dairy Science* 92, 16–24.

Vitezica Z.G., Aguilar I., Misztal I. & Legarra A. (2011) Bias in genomic predictions for populations under selection. *Genetics Research* 93, 357–66.
Wu Y., Fan H., Wang Y., Zhang L., Gao X., Chen Y., Li J., Ren H. & Gao H. (2014) Genome-wide association studies using haplotypes and individual SNPs in Simmental cattle. *PLoS ONE* **9**, e109330.

Xia J., Qi X., Wu Y., Zhu B., Xu L., Zhang L., Gao X., Chen Y., Li J. & Gao H. (2016) Genome-wide association study identifies loci and candidate genes for meat quality traits in Simmental beef cattle. *Mammalian Genome* **27**, 246–55.

Zhang W., Li J., Guo Y., Zhang L., Xu L., Gao X., Zhu B., Gao H., Ni H. & Chen Y. (2016) Multi-strategy genome-wide association studies identify the DCAF16-NCAPG region as a susceptibility locus for average daily gain in cattle. *Scientific Reports* **6**, 38073.

Zhu B., Zhu M., Jiang J. 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.