



# Accuracy of Genomic Selection for Important Economic Traits of Cashmere and Meat Goats Assessed by Simulation Study

Xiaochun Yan<sup>1</sup>, Tao Zhang<sup>1,2</sup>, Lichun Liu<sup>3</sup>, Yongsheng Yu<sup>1</sup>, Guang Yang<sup>1</sup>, Yaqian Han<sup>1</sup>, Gao Gong<sup>1</sup>, Fenghong Wang<sup>1</sup>, Lei Zhang<sup>1</sup>, Hongfu Liu<sup>1</sup>, Wenzhe Li<sup>1</sup>, Xiaomin Yan<sup>1</sup>, Haoyu Mao<sup>1</sup>, Yaming Li<sup>1</sup>, Chen Du<sup>4</sup>, Jinquan Li<sup>5,6,7</sup>, Yanjun Zhang<sup>1</sup>, Ruijun Wang<sup>1</sup>, Qi Lv<sup>1</sup>, Zhixin Wang<sup>1</sup>, Jiaxin Zhang<sup>1</sup>, Zhihong Liu<sup>1</sup>, Zhiying Wang<sup>1\*</sup> and Rui Su<sup>1\*</sup>

## OPEN ACCESS

### Edited by:

Mario Barbato,  
Catholic University of the Sacred  
Heart, Italy

### Reviewed by:

Stefano Biffani,  
Institute of Agricultural Biology and  
Biotechnology (IBBA-CNR), Italy  
Shogo Tsuruta,  
University of Georgia, United States  
Filippo Biscarini,  
National Research Council (CNR), Italy

### \*Correspondence:

Zhiying Wang  
wzhy0321@126.com  
Rui Su  
suruiyu@126.com

### Specialty section:

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

**Received:** 04 September 2021

**Accepted:** 24 January 2022

**Published:** 16 March 2022

### Citation:

Yan X, Zhang T, Liu L, Yu Y, Yang G, Han Y, Gong G, Wang F, Zhang L, Liu H, Li W, Yan X, Mao H, Li Y, Du C, Li J, Zhang Y, Wang R, Lv Q, Wang Z, Zhang J, Liu Z, Wang Z and Su R (2022) Accuracy of Genomic Selection for Important Economic Traits of Cashmere and Meat Goats Assessed by Simulation Study. *Front. Vet. Sci.* 9:770539. doi: 10.3389/fvets.2022.770539

<sup>1</sup> College of Animal Science, Inner Mongolia Agricultural University, Hohhot, China, <sup>2</sup> Inner Mongolia Bigvet Co., Ltd., Hohhot, China, <sup>3</sup> College of Veterinary Medicine, Inner Mongolia Agricultural University, Hohhot, China, <sup>4</sup> Department of Obstetrics and Gynaecology, Inner Mongolia Medical University, Hohhot, China, <sup>5</sup> Key Laboratory of Mutton Sheep Genetics and Breeding, Ministry of Agriculture, Hohhot, China, <sup>6</sup> Key Laboratory of Animal Genetics, Breeding and Reproduction in Inner Mongolia Autonomous Region, Hohhot, China, <sup>7</sup> Engineering Research Centre for Goat Genetics and Breeding, Inner Mongolia Autonomous Region, Hohhot, China

Genomic selection in plants and animals has become a standard tool for breeding because of the advantages of high accuracy and short generation intervals. Implementation of this technology is hindered by the high cost of genotyping and other factors. The aim of this study was to determine an optional marker density panel and reference population size for using genomic selection of goats, with speculation on the number of QTLs that affect the important economic traits of goats. In addition, the effect of buck population size in the reference population on the accuracy of genomic estimated breeding value (GEBV) was discussed. Based on the previous genetic evaluation results of Inner Mongolia White Cashmere Goats, live body weight (LBW,  $h^2 = 0.11$ ) and fiber diameter (FD,  $h^2 = 0.34$ ) were chosen to perform genomic selection in this study. Reasonable genome parameters and generation transmission processes were set, and phenotypic and genotype data of the two traits were simulated. Then, different sizes of the reference population and validation population were selected from progeny. The GEBVs were obtained by six methods, including GBLUP (Genomic Best Linear Unbiased Prediction), ssGBLUP (Single Step Genomic Best Linear Unbiased Prediction), BayesA, BayesB, Bayesian ridge regression, and Bayesian LASSO. The correlation coefficient between the predicted and realized phenotypes from simulation was calculated and used as a measure of the accuracy of GEBV in each trait. The results showed that the medium marker density Panel (45 K) could be used for genomic selection in goats, which can ensure the accuracy of the GEBV. The reference population size of 1,500 can achieve greater genetic progress in genomic selection for fiber diameter and live body weight in goats by comparing with the population size below this level. The accuracy of the GEBV for live body weight and fiber diameter was better when the number of QTLs was 100 and 50, respectively. Additionally, the accuracy of GEBV was discovered to be good

when the buck population size was up to 200. Meanwhile, the accuracy of the GEBV for medium heritability traits (FDs) was found to be higher than the accuracy of the GEBV for low heritability traits (LBWs). These findings will provide theoretical guidance for genomic selection in goats by using real data.

**Keywords:** genomic selection, marker density panel, reference population, number of QTLs, goats

## INTRODUCTION

As one of the earliest domesticated species, goats are distributed mainly in remote areas of some countries, including China, Mongolia, Australia, India, Iran, Pakistan, and New Zealand. Because of advantages of cashmere yield, fiber diameter, and body weight, Chinese cashmere goats are well-known worldwide. Inner Mongolia cashmere goats and Liaoning cashmere goats are widely used as paternal lines of other breeds of goats in China. Both breeds are prohibited from being exported abroad. Breeders in China have begun to perform selection for cashmere goats in 1980s. At first, it was selected according to cashmere color, then it was selected by phenotype records, and the estimated breeding value selection was implemented in 1998. Up to now, the selection of superior goats was based on estimated breeding value for liveweight and fleece traits in a large number of cashmere goat breeds (1–3). The increasing trend observed in the Chinese goat cashmere yield from 2001 to 2018 years reflects an increasing economic importance.

The Chinese goat population size is gradually decreasing, from 15.2 million in 2004 to 13.8 million in 2019. However, the cashmere yield increased first and then was kept stable from 2004 to 2019. Therefore, it is necessary to use more advanced breeding methods to improve the production performance of goats.

The idea of genomic selection was proposed and published by (4), enabling selection decisions to be made early in the life of animals. This approach is beneficial for traits that are difficult to measure and traits with low heritability. This method has been successfully applied to other livestock species, such as dairy cattle, beef, pigs, chickens, and sheep (5–9). Due to the limited marginal economic value of a goat breeding system, to the substantial number of markers required for genomic selection and to the high cost of sequencing, genomic selection in goats is still limited. In recent years, breeding programs based on genomic selection have been developed in dairy goats in France and the UK (10, 11), but the reference population size is relatively small. To date, genomic selection of cashmere goats has not been reported yet.

Meuwissen and Goddard used whole genome sequence data for the prediction of the genetic values of individuals for complex traits and obtained a prediction accuracy higher than 0.80 (12). The accuracy of genomic predictions is affected by many factors, including marker density, the level of linkage disequilibrium (LD) between the markers and QTLs, reference population size, heritability of the trait, and distribution of QTLs and GEBV methods (13–15). Muir illustrated that the increase of marker density and phenotype information can improve accuracy of genomic selection (16). The average variance proportion of each QTL decreases with the increase in number of QTLs.

When the number of QTLs is greater, it is more difficult to accurately estimate the effect of markers around each QTL, which can lead to an increase in estimation error rate. Ma et al. showed that QTL markers improved the reliability of genomic prediction. Additionally, this study illustrated that the reference population including bulls that have more progeny can increase GEBV predicted accuracy (17). Lillehammer et al. used simulated data to perform genomic selection of maternal traits in pigs, which illustrated that the genetic progress obtained by the population size of 1,000 was found to be 75% of the genetic progress of 5,000 (18). Anna Wolc et al. used simulated data to perform genomic selection of laying chickens, and found that the generation interval was shortened by half (19). Villumsen et al. used simulated data to perform genomic selection, which demonstrated that accuracy evaluation of genomic breeding value improved nearly 17% when the heritability increases from 0.02 to 0.30 (20). Clark et al. compared the accuracy of GEBV by using BLUP, GBLUP, and BayesB methods, which illustrated that the prediction will be more accurate by using BayesB if some important QTLs existed, no significant difference between GBLUP and BayesB was observed when the QTL effect was small (21).

Although the genomic selection has been well-applied to other breeding animals, including cattle, pigs, and chicken, it has been proved to obtain better selection accuracy. But the genomic selection of Chinese goats has not been reported yet. The aim of this study was to evaluate the potential effect of the density of marker panels, reference population sizes, number of QTLs, prediction methods, and buck population size in a reference population on the accuracy of GEBV for the important economic traits of goats.

## MATERIALS AND METHODS

Records of 7,102 animals collected from 1988 to 2000 at the Inner Mongolia White Cashmere goat breeding farm were used by Zhou (1) to estimate genetic parameters of Cashmere fiber diameter and live body weight. The results showed that their heritability ranged from medium (0.34 for fiber diameter) to low (0.11 for body weight). Based on these findings, both traits were used as example traits in a successive simulation analysis. Combining the genomic sequence information in goats published in NCBI (<https://www.ncbi.nlm.nih.gov/genome/?term=goats>), QMSim software was used to produce phenotype and genotype data by simulation. Then, GEBV for both traits was obtained with BGLR and HIBLUP packages in R (22). Then the related factors affected accuracy of GEBV was evaluated by correlation

coefficient between the predicted and realized phenotypes from simulation.

## Data Simulation

Using QMSim software (23), the populations were simulated based on a forward in-time process (24). In the first simulation step, 5,000 generations with a constant size of 1,000 (500 males and 500 females) were simulated, followed by 500 generations with a gradual increase in population size from 1,000 to 3,000 (400 males and 2,600 females) to create initial LD and establish mutation-drift equilibrium in historical generations. In the second step, an expansion of the population was created by initially randomly selecting 40 founder males and 400 founder females from the last generation of the historical population. To enlarge the population, 10 generations were simulated with 5 offspring per dam. The mating system was based on the random union of gametes with no selection. Subsequently, 40 males and 400 females from the last generation of the expanded population were randomly mated to generate another 10 generations with 15, 30, 45, and 60 K SNP markers, respectively. The parameters used in recent generations mimicked a real production system with one or two progeny per dam per year, 50% of male progeny, selection for high values of EBV in live body weight (*LBW*,  $h^2 = 0.11$ ) and low values of EBV in fiber diameter (*FD*,  $h^2 = 0.34$ ), then culling for individuals with a replacement rate of 80% for sires and 30% for dams. Sires and dams were randomly mated.

The simulated genome consisted of 29 pairs of autosomes with lengths identical to the real *Capra hircus* genome based on *de novo* assembly (<https://www.ncbi.nlm.nih.gov/genome/?term=goats>) (25) totaling 2,922 cM. In most reported simulation studies, only one chromosome was simulated because of the limitation of computing time and memory requirements. The advantage of simulating a real number of autosomes with lengths identical to the goat genome is to create a more realistic scenario. The SNP markers were randomly distributed, and the initial number of markers was chosen, such as 15, 30, 45, or 60 K. A total of 50, 100, and 150 QTLs were randomly distributed among the markers. The effects of QTLs were sampled from a gamma distribution with shape parameters of 0.40. The mutation rate of the markers and QTLs was assumed to be  $9.4 \times 10^{-6}$  per locus per generation. The crossover interference was set to 5.0 by referring to studies on other ruminants (26). The parameters used for simulating population structure and genomes are given in **Table 1**.

## Reference and Validation Sets

Some reports have shown that reference population size has a significant effect on the accuracy of GEBV (18, 27). Five reference population sizes (500, 1,000, 1,500, 2,000, 3,000) and one validation population size 1,000 were selected to perform genomic selection in this study. The individuals in the reference population were obtained by random sampling from the 2nd to 7th generations. The individuals in the validation population were obtained by random sampling from the 8–10th generations (**Figure 1**). The individuals in the reference population and the validation population were selected by the random sampling method, and each population size was repeated 3 times. Reference

**TABLE 1** | Parameters of the simulation process.

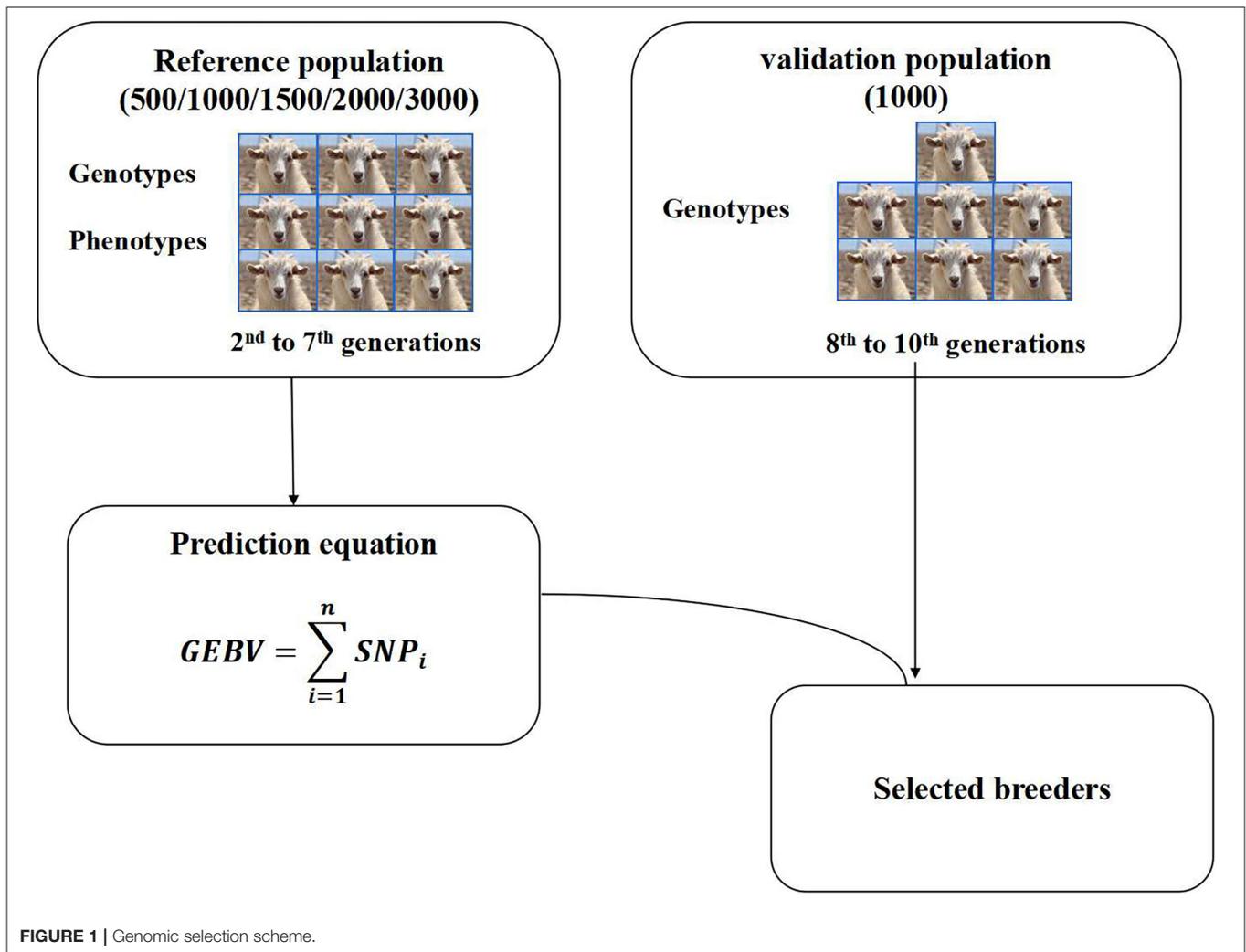
Population structure	Populations
<b>Step1: Historical generations (HGs)</b>	
Number of generations (size)-phase 1	5,000 (1,000)
Number of generations (size)-phase 2	500 (3,000)
<b>Step 2: Expanded generations (EGs)</b>	
Number of founder males from HG	400
Number of founder females from HG	2,600
Number of generations	10
Number of offspring per dam	5
<b>Step3: Recent generations</b>	
Number of founder males from EG	40
Number of founder females from EG	400
Number of generations	10
Number of offspring per dam	1, 2
Ratio of male	50%
Mating system	Random
Replacement ratio of males	80%
Replacement ratio of females	30%
Selection	EBV
Culling	Age
BV estimation method	BLUP animal model
Heritability of the traits	0.11, 0.34
Phenotypic variance	1.0
<b>Genome</b>	
Number of chromosomes	29
Total of genome length	2922 cM
Number of markers	15,000/30,000/45,000/60,000
Marker distribution	Random
Number of marker alleles	2
Number of QTLs	50/100/150
QTL distribution	Random
Number of QTL alleles	2
QTL allele effect	Gamma distribution (shape = 0.40)
Rate of recurrent mutation	$9.4 \times 10^{-6}$
Crossover interference	5.0

EBV, estimated breeding value; BLUP, best linear unbiased prediction; QTL, quantitative trait loci; the whole simulation was repeated 10 times.

datasets with phenotypes and genotypes were used to predict marker effects. The accuracy of genomic selection was evaluated based on the selected reference group and validation group. After qualifying the reference population size, different groups were set according to different male content in the reference population, and then the effects of different male content on the accuracy of genomic selection were studied.

## Methods of Estimating Genomic Breeding Value

Many methods have been applied to estimate genomic breeding value. However, the accuracy of GEBV among the different traits varies among the methods. Using the genotype and phenotype data produced from simulation, GEBV was obtained with



GBLUP (Genomic Best Linear Unbiased Prediction), ssGBLUP (Single Step Genomic Best Linear Unbiased Prediction), and Bayesian series methods such as BayesA, BayesB, Bayesian ridge regression (BRR), and Bayesian LASSO (BL). All models were fitted using the BGLR (22) and HIBLUP (<https://www.hiblup.com/>) R packages.

GBLUP is a method that utilizes genomic relationships to estimate the genetic merit of an individual (28, 29). The genomic relationship matrix defines the covariance between individuals based on observed similarity at the genomic level, rather than on expected similarity based on pedigree, so that more accurate predictions of merit can be obtained. The GBLUP method assigns the same variance to all loci and essentially treats them all as equally important. However, a series of Bayesian methods give more emphasis to some genomic regions by allowing the variance to differ between SNP loci. The BayesA method assumes that the effects of all SNPs on phenotype obey the normal distribution  $g_i \sim N(0, \alpha_{g_i}^2)$ , and the variance  $\sigma_{g_i}^2$  is subject to the inverse chi-square distribution  $\sigma_{g_i}^2 \sim \chi^{-2}(\nu, S)$ ,  $\nu$  is the degree of freedom, and  $S$  is the scale parameter (30). Most of the markers

have very small effects on phenotype, while only a few have large effects. The distribution of genetic variances across loci is that some have no genetic variance, and a few have genetic variance. However, the prior density of BayesA does not have a density peak at  $\sigma_{g_i}^2 = 0$ . In fact, its probability of  $\sigma_{g_i}^2 = 0$  is infinitesimal. The BayesB method uses a prior that has a high density,  $\pi$ , at  $\sigma_{g_i}^2 = 0$  and has an inverted chi-square distribution for  $\sigma_{g_i}^2 > 0$ . In the Bayesian Lasso (31), the prior assigned to marker effects is a Laplace (double exponential, DE) distribution. All marker effects are assumed to be independently and identically distributed as a DE. These priors assign the same variance or prior uncertainty to all marker effects, but they possess thicker tails than the normal or Gaussian prior. No fixed effects were considered in this study, and only additive genetic effects and standard deviation effects were included in the model. Therefore, the statistical methods of Bayes-Alphabet involved in this study can be written as:

$$y = \sum_{i=1}^n Z_i a_i + e$$

**TABLE 2** | Variance analysis of different factors on the accuracy of GEBV for fiber diameter and live body weight.

	Source	DF	SS	MS	F	P-value
Fiber diameter	Marker density panel	3	0.0111	0.0037	3.72	<0.05
	Number of QTLs	2	0.0418	0.0209	21.05	<0.01
	Reference population size	4	0.1693	0.0423	42.62	<0.01
	Methods	5	0.2295	0.0459	46.21	<0.01
	Error	201	0.1996	0.0010		
	Corrected Total	215	0.6512			
Live body weight	Marker density panel	3	0.1867	0.0621	36.42	<0.01
	Number of QTLs	2	0.0518	0.0259	15.20	<0.01
	Reference population size	4	0.5037	0.1259	73.92	<0.01
	Methods	5	0.6424	0.1285	75.41	<0.01
	Error	201	0.3424	0.0017		
	Corrected Total	215	1.5424			

DF, degree of freedom; SS, St dev square; MF, mean square.

$y$  is the phenotypic value vector of animals,  $Z_i$  is the design matrix of genotype at the  $i^{th}$  site,  $a_i$  is the effect value of the  $i^{th}$  marker,  $n$  is the number of markers.  $\sum_{i=1}^n Z_i a_i$  is the breeding value of animals, and  $e$  is the vector of residual effects. The hypothetical distribution of all markers' effects in different Bayes methods and the formula of effect distribution are various.

The method of GBLUP involved in the current study was as follow:

$$y = 1_n \mu + Za + e$$

$y$  is the phenotype vector of animals,  $1_n$  is a vector of ones,  $\mu$  is overall mean,  $Z$  is a design matrix corresponding to the additive effect value, and  $a$  is the vector of the breeding value for an individual. The covariance matrix of additive effects is represented by  $Var(a) = G\sigma_a^2$ , where  $G$  is the matrix of relationships between individuals obtained from genomic information,  $\sigma_a^2$  is the variance of additive genetic.  $e$  is a vector of random normal deviates.

The single-step genomic BLUP (ssGBLUP) was provided by Legarra et al. (32). The core idea of the ssGBLUP method is to combine a pedigree relationship matrix ( $A$ ) and a genomic relationship matrix ( $G$ ) to reconstruct a new relationship matrix ( $H$ ) (33–36). Excepting the relationship matrix, the theory and method of ssGBLUP had no difference from the GBLUP method.

## Accuracy of Genomic Estimated Breeding Value

Each marker effect was estimated by using phenotype and genotype information in the reference population with the above model. Then, the GEBV for the validation population was obtained by summing the effects of all the markers carried by individuals. The phenotype for the validation populations was computed by adding GEBV and residual error effects. The correlation coefficient between the predicted and realized phenotypes from simulation was calculated and used as a

measure of the accuracy of GEBV.

$$r = \frac{Cov(\hat{P}, P)}{\sigma_{\hat{P}}\sigma_P}$$

$Cov(\hat{P}, P)$  is the covariance of the predicted and realized phenotypes in the validation population,  $\sigma_{\hat{P}}$  is the standard deviation of the predicted phenotype, and  $\sigma_P$  is the standard deviation of the realized phenotype.

Finally, a generalized linear model was used to evaluate the effect of marker density panel (15, 30, 45, and 60 K), reference population size (500, 1,000, 1,500, 2,000, 3,000), number of QTLs (50, 100, 150), and the number of males in the reference population (100 M + 1,400 F, 200 M + 1,300 F, 400 M + 1,100 F, 800 M + 700 F, 120 M + 1,400 F, 1,500 M) on the accuracy of GEBV. Duncan's test was used to compare differences between various levels in each factor. The general linear model (GLM) was established to perform a significance test with SAS software (37). Duncan's test was used to compare differences between various levels in each factor.

## RESULTS

In this study, the effect of marker density panel, reference population size, number of QTLs, and buck population size in the reference population on accuracy of GEBV were discussed, which can guide the scientific and effective realization of genomic selection in cashmere and meat goats. Based on the previous genetic evaluation results, two important economic traits, fiber diameter (FD) and live body weight (LBW), were selected for analysis. Fiber diameter is a medium heritability trait, and live body weight is a low heritability trait.

## Variance Analysis of Factors on Accuracy of GEBV

The results of variance analysis of marker density panel, reference population size, number of QTLs, and methods is presented in **Table 2**. For fiber diameter, marker density panel had significant

**TABLE 3 |** Accuracy of GEBV in four marker density panels with heritability of fiber diameter and live body weight under different models.

Methods	Marker density panel ( $h^2 = 0.34$ )				Marker density panel ( $h^2 = 0.11$ )			
	15 K	30 K	45 K	60 K	15 K	30 K	45 K	60 K
BA	0.5754 ± 0.0069 <sup>b</sup>	0.6233 ± 0.0134 <sup>ab</sup>	0.6539 ± 0.0184 <sup>a</sup>	0.5703 ± 0.0328 <sup>b</sup>	0.4208 ± 0.0128 <sup>bc</sup>	0.4331 ± 0.0123 <sup>b</sup>	0.5269 ± 0.0340 <sup>a</sup>	0.3594 ± 0.0297 <sup>c</sup>
BB	0.5760 ± 0.0026 <sup>c</sup>	0.6261 ± 0.0100 <sup>b</sup>	0.6667 ± 0.0139 <sup>a</sup>	0.6493 ± 0.0175 <sup>ab</sup>	0.4135 ± 0.0127 <sup>b</sup>	0.4307 ± 0.0130 <sup>b</sup>	0.5270 ± 0.0249 <sup>a</sup>	0.3625 ± 0.0094 <sup>c</sup>
BL	0.5729 ± 0.0054 <sup>a</sup>	0.5510 ± 0.0329 <sup>a</sup>	0.5804 ± 0.0125 <sup>a</sup>	0.5767 ± 0.0266 <sup>a</sup>	0.4269 ± 0.0179 <sup>b</sup>	0.4235 ± 0.0152 <sup>b</sup>	0.4919 ± 0.0307 <sup>a</sup>	0.3731 ± 0.0218 <sup>c</sup>
BRR	0.5705 ± 0.0070 <sup>a</sup>	0.5489 ± 0.0327 <sup>a</sup>	0.5684 ± 0.0179 <sup>a</sup>	0.5757 ± 0.0272 <sup>a</sup>	0.4195 ± 0.0380 <sup>ab</sup>	0.4249 ± 0.0124 <sup>ab</sup>	0.4696 ± 0.0304 <sup>a</sup>	0.3703 ± 0.0232 <sup>b</sup>
GBLUP	0.5643 ± 0.0030 <sup>a</sup>	0.5366 ± 0.0356 <sup>a</sup>	0.5581 ± 0.0201 <sup>a</sup>	0.5718 ± 0.0221 <sup>a</sup>	0.4025 ± 0.0458 <sup>ab</sup>	0.4171 ± 0.0178 <sup>ab</sup>	0.4658 ± 0.0206 <sup>a</sup>	0.3498 ± 0.0273 <sup>b</sup>
ssGBLUP	0.7189 ± 0.0102 <sup>a</sup>	0.6535 ± 0.0246 <sup>b</sup>	0.6689 ± 0.0180 <sup>ab</sup>	0.6623 ± 0.0309 <sup>ab</sup>	0.6151 ± 0.0323 <sup>a</sup>	0.6299 ± 0.0050 <sup>a</sup>	0.5909 ± 0.0071 <sup>a</sup>	0.5957 ± 0.0677 <sup>a</sup>

BA, BayesA; BB, BayesB; BL, Bayesian Lasso; BRR, Bayesian Ridge Regression; GBLUP, Genomic Best Linear Unbiased Prediction; ssGBLUP, Single Step Genomic Best Linear Unbiased Prediction. <sup>a,b,c</sup> represent significant differences. The difference is significant with different letters.

effect on accuracy of GEBV ( $P < 0.05$ ), other factors in this study had highly significant effect on accuracy of GEBV ( $P < 0.01$ ). For live body weight, all the factors in this study had highly significant effect on accuracy of GEBV ( $P < 0.01$ ).

### Effect of Marker Density Panel on Accuracy of GEBV

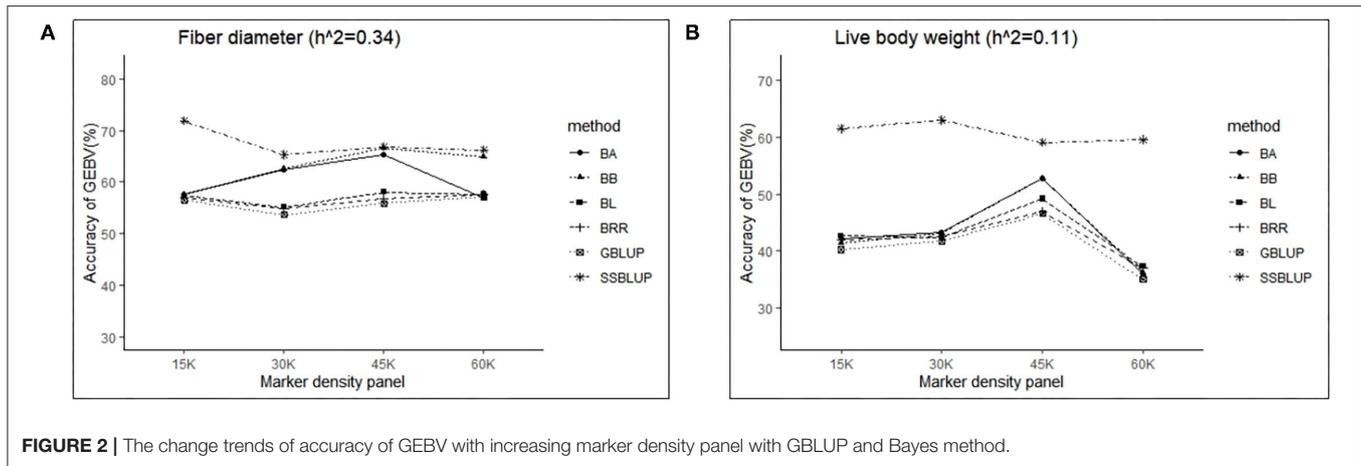
Controlling other factors at the same level, the accuracy of GEBV in four marker density panels (15, 30, 45, and 60 K) with the GBLUP, ssGBLUP, and Bayes methods is shown in **Table 3**. The results obtained by GLM using the least square means method demonstrated that the marker density panel had a significant effect on the accuracy of GEBV for both traits. For FD, the accuracy of GEBV at 45 K is significantly higher than the accuracy of GEBV at 15 and 30 K. The results obtained at 45 and 60 K were not significantly different for FD. However, the accuracy of genomic selection for low heritability traits at 45 K was obviously higher than the accuracy of genomic selection at 15, 30, and 60 K. In general, an increasing trend was observed for the accuracy of GEBV with the marker density panel from 15 to 45 K in both traits (**Figure 2**). However, it is interesting that the accuracy of GEBV at 60 K for LBW is significantly lower than the accuracy of GEBV at 45 K (**Figure 2B**). Under the best marker density panel, the correlation coefficients between the predicted and realized phenotypes for the validation population with medium and low heritability traits reached 66.7 and 52.7% under the BayesB method, respectively. The accuracy of GEBV for medium heritability traits was discovered to be higher than the accuracy of GEBV for low heritability traits.

### Effect of Reference Population Size on Accuracy of GEBV

Similarly, the effects of reference population size on the accuracy of GEBV were analyzed by controlling other factors at the same level. The results are shown in **Tables 4, 5**. Reference population size had a significant effect on the prediction accuracy of FD and LBW. The accuracy of GEBV with reference population sizes of 1,500, 2,000, and 3,000 was significantly higher than the accuracy of GEBV with the reference populations of 500 and 1,000 for both traits. No significant difference in prediction accuracy was discovered among the 1,500, 2,000, and 3,000 training sets. Increasing trends of GEBV accuracy were observed in FD and LBW (**Figure 3**). The average genomic accuracy ranged from 55.27 to 67.4% for the medium heritability trait (**Table 4**) and from 40.39 to 59.09% for the low heritability trait with the ssGBLUP method (**Table 5**). Meanwhile, the accuracy value with the ssGBLUP methods was found to be higher than the accuracy value with the GBLUP methods.

### Effect of Number of QTLs on Accuracy of GEBV

In this study, three levels of QTLs (50, 100, and 150) were considered for analysis. For FD, the accuracy of GEBV in QTLs of 100 was significantly lower than the accuracy of GEBV in the other two levels, and the value was best when the number of QTLs was 50 (**Table 6**). Similarly, except for the ssGBLUP method, the



**TABLE 4 |** Accuracy of GEBV in five reference population sizes with heritability of fiber diameter under different models.

Methods	Reference population size ( $h^2 = 0.34$ )				
	500	1,000	1,500	2,000	3,000
BA	0.5061 ± 0.0351 <sup>c</sup>	0.5427 ± 0.0545 <sup>bc</sup>	0.5703 ± 0.0328 <sup>bc</sup>	0.6205 ± 0.034 <sup>ab</sup>	0.6734 ± 0.0185 <sup>a</sup>
BB	0.5292 ± 0.0307 <sup>b</sup>	0.5514 ± 0.0574 <sup>b</sup>	0.6493 ± 0.0175 <sup>a</sup>	0.6527 ± 0.0247 <sup>a</sup>	0.6876 ± 0.0108 <sup>a</sup>
BL	0.5190 ± 0.0378 <sup>b</sup>	0.5455 ± 0.064 <sup>ab</sup>	0.5767 ± 0.0266 <sup>ab</sup>	0.5885 ± 0.0197 <sup>ab</sup>	0.6312 ± 0.0135 <sup>a</sup>
BRR	0.5272 ± 0.0426 <sup>b</sup>	0.5502 ± 0.0594 <sup>b</sup>	0.5757 ± 0.0272 <sup>ab</sup>	0.5829 ± 0.0208 <sup>ab</sup>	0.6262 ± 0.0108 <sup>a</sup>
GBLUP	0.5080 ± 0.0328 <sup>b</sup>	0.5483 ± 0.0330 <sup>ab</sup>	0.5718 ± 0.0221 <sup>ab</sup>	0.5641 ± 0.0190 <sup>ab</sup>	0.6044 ± 0.0113 <sup>a</sup>
ssGBLUP	0.5527 ± 0.0655 <sup>b</sup>	0.6575 ± 0.0488 <sup>a</sup>	0.6623 ± 0.0309 <sup>a</sup>	0.6459 ± 0.0250 <sup>a</sup>	0.6740 ± 0.0108 <sup>a</sup>

BA, BayesA; BB, BayesB; BL, Bayesian Lasso; BRR, Bayesian Ridge Regression; GBLUP, Genomic Best Linear Unbiased Prediction; ssGBLUP, Single Step Genomic Best Linear Unbiased Prediction. <sup>a,b</sup> represent significant differences. The difference is significant with different letters.

trend of accuracy of GEBV with the number of QTLs showed that it decreased first and then increased (Figure 4A). For LBW, the accuracy of GEBV in QTLs of 100 was significantly higher than the accuracy of GEBV in the other two levels, and the value was up to 59.09% with the ssGBLUP method (Table 6). A first increasing and then decreasing trend was observed for the effect of the number of QTLs on the accuracy of GEBV (Figure 4B). The accuracy of GEBV with the number of QTLs in the GBLUP method was relatively lower than the accuracy of GEBV with the number of QTLs in the Bayesian methods. The accuracy of GEBV for medium- and low-heritability traits was better when the number of QTLs was 50 and 100, respectively.

## Effect of the Number of Males in the Reference Population on the Accuracy of GEBV

The variance analysis of the number of males in the reference population is presented in Tables 7, 8. For fiber diameter and live body weight, the number of males in the reference population had highly significant effect on accuracy of GEBV in goats ( $P < 0.01$ ). The multiple comparative analysis demonstrated that the higher the number of males size in the reference population, the more effective information can be provided, and the higher the accuracy of GEBV. Results from the effect of the reference population size have shown that 1,500 individuals were optimal

to obtain significant genetic progress in both FD and LBW traits. Therefore, six groups were classified based on the ratio male (M) to female (F) in reference population (100 M + 1,400 F, 200 M + 1,300 F, 400 M + 1,100 F, 800 M + 700 F, 1,200 M + 300 F, 1,500 M). The effect of buck population size in the reference population on prediction accuracy was analyzed. The results indicated that the accuracy of GEBV was the highest when the buck population size was up to 400 (400 M + 1,100 F), which was significantly higher than the buck population size in other groups for FD (Table 9). The results indicated that the accuracy of GEBV was the highest when the buck population size was up to 100 (100 M + 1,400 F), which was significantly higher than the buck population size in other groups for LBW (Table 10). The accuracy values of GEBV in FD and LBW are 70.91 and 58.21%, respectively. In general, for FD and LBW, the trend of accuracy of GEBV with buck population size first increases and then remains stable (Figure 5). The trend for LBW was found to be basically consistent in each method. However, the trend for FD is irregular by BayesB method.

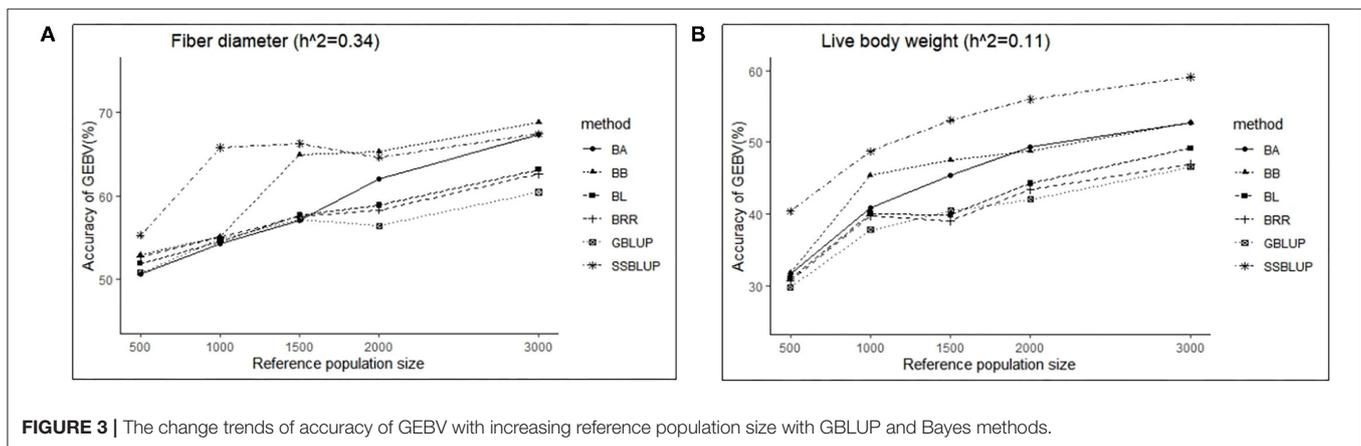
## DISCUSSION

Previous studies illustrated that the genetic evaluation for fiber diameter in Inner Mongolia White Cashmere goats by using phenotype records of 1- and 2-year-olds could ensure the accuracy of EBV (38, 39). Considering the prohibitive cost

**TABLE 5** | Accuracy of GEBV in five reference population sizes with heritability of live body weight under different models.

Methods	Reference population size ( $h^2 = 0.11$ )				
	500	1,000	1,500	2,000	3,000
BA	0.3171 ± 0.0870 <sup>c</sup>	0.4082 ± 0.0321 <sup>b</sup>	0.4543 ± 0.0344 <sup>ab</sup>	0.4929 ± 0.0212 <sup>ab</sup>	0.5269 ± 0.0340 <sup>a</sup>
BB	0.3190 ± 0.0915 <sup>b</sup>	0.4545 ± 0.0213 <sup>ab</sup>	0.4757 ± 0.0342 <sup>ab</sup>	0.4875 ± 0.0274 <sup>a</sup>	0.5270 ± 0.0249 <sup>a</sup>
BL	0.3103 ± 0.1020 <sup>b</sup>	0.4003 ± 0.0330 <sup>ab</sup>	0.3990 ± 0.0459 <sup>ab</sup>	0.4427 ± 0.0153 <sup>ab</sup>	0.4919 ± 0.0307 <sup>a</sup>
BRR	0.3078 ± 0.1020 <sup>b</sup>	0.3976 ± 0.0312 <sup>ab</sup>	0.3902 ± 0.0399 <sup>ab</sup>	0.4340 ± 0.0170 <sup>ab</sup>	0.4696 ± 0.0304 <sup>a</sup>
GBLUP	0.2983 ± 0.0862 <sup>b</sup>	0.3780 ± 0.0273 <sup>ab</sup>	0.4048 ± 0.0773 <sup>ab</sup>	0.4204 ± 0.0246 <sup>ab</sup>	0.4658 ± 0.0206 <sup>a</sup>
ssGBLUP	0.4039 ± 0.0763 <sup>c</sup>	0.4875 ± 0.0389 <sup>bc</sup>	0.5308 ± 0.0658 <sup>ab</sup>	0.5597 ± 0.0216 <sup>ab</sup>	0.5909 ± 0.0071 <sup>a</sup>

BA, BayesA; BB, BayesB; BL, Bayesian Lasso; BRR, Bayesian Ridge Regression; GBLUP, Genomic Best Linear Unbiased Prediction; ssGBLUP, Single Step Genomic Best Linear Unbiased Prediction. <sup>a,b</sup> represent significant differences. The difference is significant with different letters.

**FIGURE 3** | The change trends of accuracy of GEBV with increasing reference population size with GBLUP and Bayes methods.

and long duration of performance testing, genomic selection is gradually applied to animal breeding. The greatest advantage of genomic selection is that early selection can be achieved by collecting DNA samples at an early stage and genotyping individuals. The accuracy of genomic selection was influenced by many factors, such as SNP chip marker density, QTL numbers, genomic selection model, and so on. Therefore, it is necessary to discuss the factors affecting the accuracy of genomic selection in goats.

When more markers were distributed on the chromosome with a certain length, more favorable information could be provided. Detection of markers that are in linkage disequilibrium with QTLs is easier. Generally, a higher accuracy of genomic selection would be obtained with greater marker density. However, when marker density reaches a certain number, the accuracy of genomic selection will not increase significantly to some extent, or the growth rate will slow down. Increasing marker density from 24 to 728 K SNPs resulted in a small increase in the accuracy of GEBV in three cow breeds with the GBLUP method, and the values of 24 and 728 K were 0.22 and 0.24, respectively (40). The cost of genotyping is closely related to the marker density panel. However, some studies have shown that the accuracy of GEBV by using genotype information with high-density chips is higher than the accuracy of GEBV by using genotype information with low-density chips (41, 42). The excessive cost of genotyping limited the realization of genomic

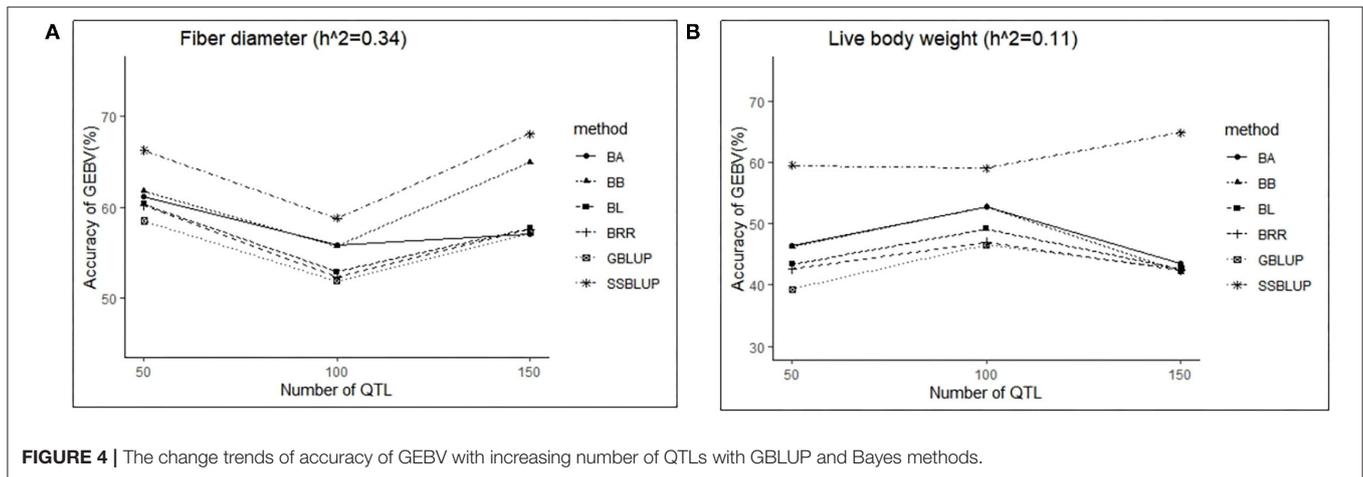
selection in low-income species, such as goats. In production practice, the goal of breeders is to achieve equivalent effects using low-density SNP chips, as well as high-density chips, which can reduce sequence cost and improve the accuracy of selection. The results obtained from our study were similar to the results from previous reports. The prediction accuracy increased with increasing marker density chips. For both traits, the medium marker density chip was most effective in genomic selection of goats. Solberg et al. reported that the accuracy of genome estimation breeding value increased significantly with increasing marker density by simulation (43). The linkage disequilibrium between adjacent SNP markers was positively correlated with the accuracy of genomic breeding value. The degree of linkage disequilibrium depends on the marker density (44).

Many studies have reported that the influence of reference population size has an effect on genomic selection accuracy. Generally, a higher GEBV prediction accuracy was obtained with a larger reference population size. When the population size is small, the genomic relationship matrix cannot hold enough genomic information (independent chromosome segments). Therefore, the accuracy can be lower with the smaller size. Therefore, the number of individuals in the reference group should be increased as much as possible when genomic selection is performed. However, the optimal reference population size must be considered because of high sequencing costs. Zengting Liu et al. estimated the genomic breeding value of milk yield

**TABLE 6** | Accuracy of GEBV in three QTLs with heritability of fiber diameter and live body weight under different methods.

Methods	Number of QTLs ( $h^2 = 0.34$ )			Number of QTLs ( $h^2 = 0.11$ )		
	50	100	150	50	100	150
BA	0.6117 ± 0.0216 <sup>a</sup>	0.5583 ± 0.0045 <sup>b</sup>	0.5703 ± 0.0328 <sup>ab</sup>	0.4636 ± 0.0251 <sup>a</sup>	0.5269 ± 0.0340 <sup>a</sup>	0.4360 ± 0.0062 <sup>a</sup>
BB	0.6173 ± 0.0137 <sup>ab</sup>	0.5568 ± 0.0262 <sup>b</sup>	0.6493 ± 0.0175 <sup>a</sup>	0.4619 ± 0.0250 <sup>ab</sup>	0.5270 ± 0.0249 <sup>a</sup>	0.4217 ± 0.0263 <sup>b</sup>
BL	0.6032 ± 0.0156 <sup>a</sup>	0.5290 ± 0.0191 <sup>b</sup>	0.5767 ± 0.0266 <sup>ab</sup>	0.4341 ± 0.0190 <sup>a</sup>	0.4919 ± 0.0307 <sup>a</sup>	0.4257 ± 0.0200 <sup>a</sup>
BRR	0.6015 ± 0.0136 <sup>a</sup>	0.5231 ± 0.0192 <sup>b</sup>	0.5757 ± 0.0272 <sup>ab</sup>	0.4259 ± 0.0215 <sup>a</sup>	0.4696 ± 0.0304 <sup>a</sup>	0.4241 ± 0.0148 <sup>a</sup>
GBLUP	0.5850 ± 0.0121 <sup>a</sup>	0.5190 ± 0.0165 <sup>a</sup>	0.5718 ± 0.0221 <sup>a</sup>	0.3932 ± 0.0462 <sup>a</sup>	0.4658 ± 0.0206 <sup>a</sup>	0.4264 ± 0.0041 <sup>a</sup>
ssGBLUP	0.6623 ± 0.0309 <sup>a</sup>	0.5877 ± 0.0344 <sup>a</sup>	0.6799 ± 0.0240 <sup>a</sup>	0.5954 ± 0.0193 <sup>a</sup>	0.5909 ± 0.0071 <sup>a</sup>	0.6489 ± 0.0154 <sup>a</sup>

BA, BayesA; BB, BayesB; BL, Bayesian Lasso; BRR, Bayesian Ridge Regression; GBLUP, Genomic Best Linear Unbiased Prediction; ssGBLUP, single step Best Linear Unbiased Prediction. <sup>a,b</sup> represent significant differences. The difference is significant with different letters.



**FIGURE 4** | The change trends of accuracy of GEBV with increasing number of QTLs with GBLUP and Bayes methods.

**TABLE 7** | Variance analysis of the ratio of male to female on the accuracy of GEBV for fiber diameter.

Source	DF	SS	MS	F	P-value
The ratio of male to female	5	0.0196	0.0039	3.67	<0.01
Methods	5	0.3438	0.0688	64.32	<0.01
Error	97	0.1037	0.0011		
Corrected Total	107	0.4671			

DF, degree of freedom; SS, St dev square; MF, mean square.

**TABLE 8** | Variance analysis of the ratio of male to female on the accuracy of GEBV for live body weight.

Source	DF	SS	MS	F	P-value
The ratio of male to female	5	0.0399	0.0080	5.50	<0.01
Methods	5	0.3263	0.0653	44.91	<0.01
Error	97	0.1409	0.0015		
Corrected Total	107	0.5071			

DF, degree of freedom; SS, St dev square; MF, mean square.

in dairy cattle, which indicated that the additive effect variance increased five times when the size of the reference population increased from 734 to 5,025 (45). In our study, when the reference population increased from 500 to 3,000, the accuracy of GEBV in the medium and low heritability traits increased by 15 and 20%, respectively. (46) evaluated the effect of reference population size on genomic selection in dairy goats. The population size was demonstrated to have an important effect on GEBV accuracies, from 2 to 31% with the reference population from 1,966 to 2,651 (46). However, Moser et al. reported the genomic selection of milk protein in Holstein cattle, which explained that the GEBV accuracy showed no obvious change for the reference population from 1,239 to 1,822 (47). All

these results indicated that the accuracy of genomic selection can be effectively ensured when the reference population reaches a certain level.

Generally, the genetic variance was assumed to be one regardless of the number of QTLs. When the number of QTLs is large, the variance proportion of each QTL decreases. That is, the contribution of each QTL to phenotypic value decreases, and the probability of the effect (or variance) of each QTL correctly estimated will be relatively low, which will lead to an increase in deviation and a decrease in the accuracy of GEBV. When the number of QTLs is small, it will be hard to estimate the additive genetic variance or heritability, assuming no polygenic effects, especially for a small data size. The results from our

**TABLE 9 |** Accuracy of GEV in six levels ratio of male to female with heritability of fiber diameter under different models.

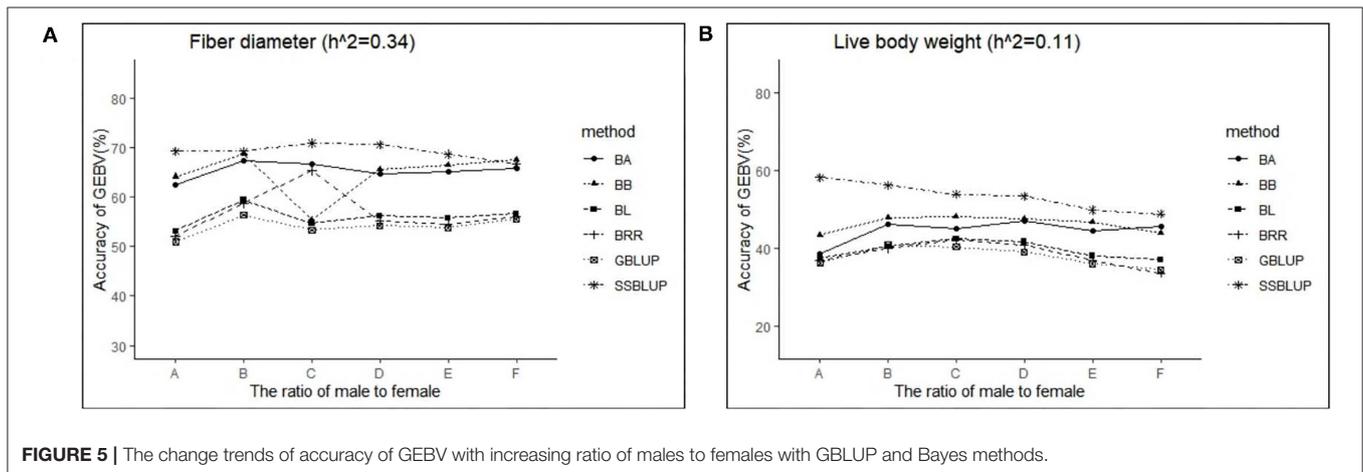
Methods	The ratio of male to female ( $h^2 = 0.34$ )					
	A 100 M + 1,400 F	B 200 M + 1,300 F	C 400 M + 1,100 F	D 800 M + 700 F	E 1,200 M + 300 F	F 1,500 M
BA	0.6239 ± 0.0147 <sup>b</sup>	0.6740 ± 0.0117 <sup>a</sup>	0.6676 ± 0.0139 <sup>ab</sup>	0.6477 ± 0.0171 <sup>ab</sup>	0.6518 ± 0.0240 <sup>ab</sup>	0.6591 ± 0.0026 <sup>ab</sup>
BB	0.6416 ± 0.0205 <sup>b</sup>	0.6867 ± 0.0161 <sup>a</sup>	0.5550 ± 0.0089 <sup>c</sup>	0.6569 ± 0.0184 <sup>ab</sup>	0.6647 ± 0.0265 <sup>ab</sup>	0.6761 ± 0.0029 <sup>a</sup>
BL	0.5312 ± 0.0352 <sup>b</sup>	0.5939 ± 0.0132 <sup>a</sup>	0.5472 ± 0.0115 <sup>b</sup>	0.5620 ± 0.0359 <sup>ab</sup>	0.5580 ± 0.0280 <sup>ab</sup>	0.5669 ± 0.0049 <sup>ab</sup>
BRR	0.5216 ± 0.0330 <sup>b</sup>	0.5862 ± 0.0140 <sup>ab</sup>	0.6535 ± 0.0155 <sup>a</sup>	0.5523 ± 0.0368 <sup>b</sup>	0.5449 ± 0.0273 <sup>b</sup>	0.5611 ± 0.0030 <sup>b</sup>
GBLUP	0.5096 ± 0.0366 <sup>b</sup>	0.5635 ± 0.0110 <sup>a</sup>	0.5330 ± 0.0115 <sup>ab</sup>	0.5422 ± 0.0432 <sup>ab</sup>	0.5377 ± 0.0386 <sup>ab</sup>	0.5552 ± 0.0129 <sup>ab</sup>
ssGBLUP	0.6930 ± 0.0423 <sup>ab</sup>	0.6932 ± 0.0129 <sup>ab</sup>	0.7091 ± 0.0136 <sup>a</sup>	0.7063 ± 0.0136 <sup>ab</sup>	0.6868 ± 0.0172 <sup>ab</sup>	0.6671 ± 0.0228 <sup>b</sup>

BA, BayesA; BB, BayesB; BL, Bayesian Lasso; BRR, Bayesian Ridge Regression; GBLUP, Genomic Best Linear Unbiased Prediction; ssGBLUP, Single Step Genomic Best Linear Unbiased Prediction; M, male; F, female. <sup>a,b</sup> represent significant differences. The difference is significant with different letters.

**TABLE 10 |** Accuracy of GEV in six levels ratio of male to female with heritability of live body weight under different models.

Methods	The ratio of male to female ( $h^2 = 0.11$ )					
	A 100 M + 1,400 F	B 200 M + 1,300 F	C 400 M + 1,100, F	D 800 M + 700 F	E 1,200 M + 300 F	F 1,500 M
BA	0.3866 ± 0.0185 <sup>b</sup>	0.4628 ± 0.0524 <sup>ab</sup>	0.4503 ± 0.0506 <sup>ab</sup>	0.4698 ± 0.0243 <sup>a</sup>	0.4460 ± 0.0590 <sup>ab</sup>	0.4565 ± 0.0406 <sup>ab</sup>
BB	0.4325 ± 0.0131 <sup>a</sup>	0.4796 ± 0.0262 <sup>a</sup>	0.4829 ± 0.0393 <sup>a</sup>	0.4757 ± 0.0152 <sup>a</sup>	0.4665 ± 0.0422 <sup>a</sup>	0.4394 ± 0.0313 <sup>a</sup>
BL	0.3732 ± 0.0137 <sup>a</sup>	0.4042 ± 0.0287 <sup>a</sup>	0.4248 ± 0.0372 <sup>a</sup>	0.4171 ± 0.0366 <sup>a</sup>	0.3803 ± 0.0610 <sup>a</sup>	0.3705 ± 0.0271 <sup>a</sup>
BRR	0.3687 ± 0.0144 <sup>ab</sup>	0.3997 ± 0.0306 <sup>a</sup>	0.4211 ± 0.0414 <sup>a</sup>	0.4084 ± 0.0382 <sup>a</sup>	0.3676 ± 0.0518 <sup>ab</sup>	0.3363 ± 0.0138 <sup>b</sup>
GBLUP	0.3619 ± 0.0249 <sup>a</sup>	0.4090 ± 0.0382 <sup>a</sup>	0.4033 ± 0.0423 <sup>a</sup>	0.3906 ± 0.0363 <sup>a</sup>	0.3605 ± 0.0598 <sup>a</sup>	0.3455 ± 0.0436 <sup>a</sup>
ssGBLUP	0.5821 ± 0.0384 <sup>a</sup>	0.5622 ± 0.0263 <sup>ab</sup>	0.5393 ± 0.0334 <sup>ab</sup>	0.5340 ± 0.0338 <sup>ab</sup>	0.4988 ± 0.0722 <sup>b</sup>	0.4884 ± 0.0369 <sup>b</sup>

BA, BayesA; BB, BayesB; BL, Bayesian Lasso; BRR, Bayesian Ridge Regression; GBLUP, Genomic Best Linear Unbiased Prediction; ssGBLUP, Single Step Genomic Best Linear Unbiased Prediction; M, male; F, female. <sup>a,b</sup> represent significant differences. The difference is significant with different letters.



study were consistent with this point of view. Zhang et al. reported that the accuracy of GEV decreased with the increase in the number of QTLs from 50 to 1,000. The BayesB methods seemed to be more sensitive to the number of QTLs than the GBLUP method (48), which is similar to our study. Daetwyler et al. compared the genomic selection accuracy between the GBLUP and BayesB methods, which demonstrated that the prediction accuracy with the BayesB method was greatest at low NQTLs and decreased with increasing NQTLs. However, as  $N_{QTL}$  increased, the difference between the two methods

decreased, and eventually, both approaches achieved very similar accuracy (49).

Sex chromosomes play a significant role in key evolutionary processes such as speciation and adaptation (50). The male to female ratio could affect accuracy of GEV because it changes the effective population size as well as LD. Previous studies have shown that the accuracy of genomic selection can be improved by increasing male size in reference populations. Avendano (51) reported that the accuracy of GEV decreased with an increase in the ratio of males to females in chickens, and the values

increased from 0.33 to 0.5 (51). Céline (46) showed that the GEBV accuracies with increasing male size in the reference population were not improved (46). Our results demonstrated that the accuracy of GEBV showed no significant change when the buck population size was up to 200. The obvious results in different studies were explained by breeds, methods, and other factors.

Genomic selection is an effective way to accelerate the genetic improvement of traits with low heritability and unmeasurable traits. In our study, two traits with medium and low heritability were used for analysis. The results showed that the GEBV accuracy for medium heritability traits was higher than the GEBV accuracy for low heritability traits, which is consistent with previous reports. Villumsen et al. evaluated the effects of heritability on genomic estimated breeding value. The accuracy of the genomic estimated breeding value was found to increase from 0.69 to 0.86 when heritability increased from 0.02 to 0.30 (20). Zhang et al. reported the accuracies of GEBV by different methods and various heritability traits (48). By decreasing the heritability from 0.90 to 0.05, the prediction accuracies with all methods decreased significantly. The accuracy of GEBV for heritability of 0.1 and 0.3 was slightly higher than the accuracy of GEBV in our study, with heritability traits of 0.11 and 0.34.

Many methods, including GBLUP, ssGBLUP, and Bayes methods, have been used to perform genomic selection in plants and animals. To some extent, the methods affected the accuracy of the prediction accuracy. Gao et al. compared the efficiency of four Bayesian models and the GBLUP model on the GEBV accuracy, which indicated that the superiority of the Bayesian models over the GBLUP model was more profound (52). Sun et al. compared the accuracy of GEBV obtained by BayesB, RRBLUP, and GBLUP using simulated datasets. The prediction accuracy with BayesB was found to be higher than the prediction accuracy with RRBLUP and GBLUP. There were no significant differences among the methods (53). Clark et al. (21) compared the impact of ABLUP, GBLUP, and BayesB on the accuracy of genetic evaluation. For the ABLUP method, the numeric relationship matrix (NRM) was calculated by pedigree. The results showed that the BayesB method would be more accurate if important QTLs had an effect on the traits. However, Clark et al. reported that Bayes and GBLUP methods had similar prediction accuracy when each QTL had a small effect (21). The ssGBLUP method used both genotype information and pedigree information to construct the relationship matrix when GEBV was obtained, which is an ideal alternative for genomic genetic evaluation compared with other methods. Lourenco et al. reported that predictive ability of genomic EBV for growth traits and calving ease when using single-step genomic BLUP (ssGBLUP) in Angus cattle was higher than that in using BLUP (54). Teissier et al. illustrated that the accuracy of GEBV for milk production traits, udder type traits, and somatic cell scores in French dairy goats was higher than that using other methods. Similarly, the accuracy of GEBV in ssGBLUP for FD and LBW was higher than that with other methods in our

study (55). In addition, the computation efficiency for ssGBLUP was also relatively good by comparing with the Bayes methods. Therefore, the ssGBLUP method was suggested to perform genomic selection in goats.

## CONCLUSIONS

All the results in this study determined the optional level of factors influencing the accuracy of genomic estimated breeding value of FD and LBW in goats. The medium marker density panels were designed for genotyping, which can effectively ensure the accuracy of genomic selection of goats. When the reference population size was up to 1,500, genomic selection of cashmere and meat goats was performed. The accuracy of GEBV for FD and LBW was better when the number of QTLs was 50 and 100, respectively, indicating that both traits were controlled by minor genes. Meanwhile, the accuracy of GEBV was discovered to be good when the buck population size in the reference population was up to 200. All these factors will make a reasonable judgement on the factors affecting genomic selection and lay a foundation for the subsequent realization of genomic selection in cashmere and meat goat breeding.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

## AUTHOR CONTRIBUTIONS

ZhiyW, RS, and XiaocY conceived of and coordinated the study. YZ, ZhixW, ZL, JZ, RW, CD, and JL helped in conceive of the study. GG, FW, LZ, and QL simulated the data. TZ, LL, YY, GY, YH, HM, HL, YL, WL, and XiaomY analyzed the data. XiaocY and ZhiyW wrote the manuscript. All authors read and approved the final manuscript.

## FUNDING

The authors are grateful for the grants supported by Natural Science for Youth Foundation (31702086), IMAR (Inner Mongolia Autonomous Region) Natural Science Foundation (2019MS03070), Inner Mongolia Autonomous Region Science and Technology Research Project (2021GG0086), Inner Mongolia Agricultural University for second levels of outstanding doctorate (NDYB2016-05), Science and technology major project of Inner Mongolia Autonomous Region (2021ZD0012), China Agriculture Research System of MOF and MARA (No. CARS-39), Scientific Research Projects of Institutions of Higher Learning in Inner Mongolia Autonomous Region (NJZY19104), and IMAR (Inner Mongolia Autonomous Region) Natural Science Foundation (2019MS08121).

## REFERENCES

- Zhou HM, Allain D, Li JQ, Zhang WG, Yu XCJJ, Genetics AB. Genetic parameters of production traits of inner mongolia cashmere goats in China. *J Anim Breed Genet.* (2002) 119:385–90. doi: 10.1046/j.1439-0388.2002.00362.x
- Barazandeh A, Moghbeli SM, Vatankhah M, Mohammadabadi MJ. Estimating non-genetic and genetic parameters of pre-weaning growth traits in Raini Cashmere goat. *Trop Anim Health Produc.* (2012) 44:811–7. doi: 10.1007/s11250-011-9971-5
- Newman SAN and Paterson DJNZJoAR. Estimates of environmental effects for liveweight and fleece characteristics of New Zealand cashmere goats. *New Zealand J Agricul Res.* (1996) 39:379–86. doi: 10.1080/00288233.1996.9513197
- Meuwissen TH, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. *Genetics.* (2001) 157:1819–29. doi: 10.1093/genetics/157.4.1819
- Mouresan EF, Altarriba J, Moreno C, Munilla S, González-Rodríguez A, Varona L. Performance of genomic selection under a single-step approach in autochthonous Spanish beef cattle populations. *J Anim Breed Genet.* (2017) 134:289–99. doi: 10.1111/jbg.12253
- Nirea KG, Meuwissen TH. Improving production efficiency in the presence of genotype by environment interactions in pig genomic selection breeding programmes. *J Anim Breed Genet.* (2017) 134:119–28. doi: 10.1111/jbg.12250
- Hernández-Montiel W, Martínez-Núñez MA, Ramón-Ugalde JP, Román-Ponce SI, Calderón-Chagoya R, Zamora-Bustillos R. Genome-wide association study reveals candidate genes for litter size traits in pelibuey sheep. *Animals.* (2020) 10:434. doi: 10.3390/ani10030434
- Jiang Y, Tang S, Xiao W, Yun P, Ding X. A genome-wide association study of reproduction traits in four pig populations with different genetic backgrounds. *Asian Australas J Anim Sci.* (2020) 33:1400–10. doi: 10.5713/ajas.19.0411
- Liu X, Liu L, Wang J, Cui H, Chu H, Bi H, et al. Genome-Wide association study of muscle glycogen in jingxing yellow chicken. *Genes.* (2020) 11:497. doi: 10.3390/genes11050497
- Mucha S, Mrode R, MacLaren-Lee I, Coffey M, Conington J. Estimation of genetic breeding values for milk yield in UK dairy goats. *J Dairy Sci.* (2015) 98:8201–8. doi: 10.3168/jds.2015-9682
- Teissier M, Larroque H, Robert-Granié C. Weighted single-step genomic BLUP improves accuracy of genomic breeding values for protein content in French dairy goats: a quantitative trait influenced by a major gene. *GSE.* (2018) 50:31. doi: 10.1186/s12711-018-0400-3
- Meuwissen T, Goddard M. Accurate prediction of genetic values for complex traits by whole-genome resequencing. *Genetics.* (2010) 185:623–31. doi: 10.1534/genetics.110.116590
- Colombani C, Legarra A, Fritz S, Guillaume F, Croiseau P, Ducrocq V, et al. Application of bayesian least absolute shrinkage and selection operator (LASSO) and BayesC $\pi$  methods for genomic selection in French holstein and montbéliarde breeds. *J Dairy Sci.* (2013) 96:575–91. doi: 10.3168/jds.2011-5225
- Esfandyari H, Sørensen A, Bijma P. A crossbred reference population can improve the response to genomic selection for crossbred performance. *GSE.* (2015) 47:76. doi: 10.1186/s12711-015-0155-z
- Liu H, Zhou H, Wu Y, Li X, Zhao J, Zuo T, et al. The impact of genetic relationship and linkage disequilibrium on genomic selection. *PLoS ONE.* (2015) 10:e0132379. doi: 10.1145/2818302
- Muir WM. Comparison of genomic and traditional BLUP-estimated breeding value accuracy and selection response under alternative trait and genomic parameters. *J Anim Breed Genet.* (2015) 124:342–55. doi: 10.1111/j.1439-0388.2007.00700.x
- Ma P, Lund MS, Aamand GP, Su G. Use of a Bayesian model including QTL markers increases prediction reliability when test animals are distant from the reference population. *J Dairy Sci.* (2019) 102:7237–47. doi: 10.3168/jds.2018-15815
- Lillehammer M, Sonesson AK, Meuwissen T. Use of field data in pig genomic selection schemes: a simulation study. *Animal.* (2016) 10:1025–32. doi: 10.1017/S1751731115002669
- Wolc A, Arango J, Settar P, Fulton JE, O'Sullivan NP, Jankowski T, et al. Accounting for complex population structure in pedigree and genomic analyses of laying chickens. *Animal Industry Rep.* (2015) 661:65. doi: 10.31274/ans\_air-180814-1323
- Villumsen T, Janss L, Lund M. The importance of haplotype length and heritability using genomic selection in dairy cattle. *J Anim Breed Genet Zeitschrift Tierzucht Zuchtungsbiol.* (2009) 126:3–13. doi: 10.1111/j.1439-0388.2008.00747.x
- Clark S, Hickey J, van der Werf J. Different models of genetic variation and their effect on genomic evaluation. *GSE.* (2011) 43:18. doi: 10.1186/1297-9686-43-18
- P'Erez P and Campos G. Genome-wide regression and prediction with the BGLR statistical package. *Genetics.* (2014) 198:483–95. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4196607/
- Sargolzaei M, Schenkel F. QMSim: a large-scale genome simulator for livestock. *Bioinformatics.* (2009) 25:680–1. doi: 10.1093/bioinformatics/btp045
- Carvajal-Rodríguez A. Simulation of genomes: a review. *Curr Genom.* (2008) 9:155–9. doi: 10.2174/138920208784340759
- Dong Y, Xie M, Jiang Y, Xiao N, Du X, Zhang W, et al. Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (*Capra hircus*). *Nat Biotechnol.* (2013) 31:135–41. doi: 10.1038/nbt.2478
- Wang Z, Shen B, Jiang J, Li J, Ma L. Effect of sex, age and genetics on crossover interference in cattle. *Sci Rep.* (2016) 6:37698. doi: 10.1038/srep37698
- Hosseini S, Foroutanifar S, Abdolmohammadi A. Comparison of combined, crossbred, and purebred reference populations for genomic selection in small populations. *Small Ruminant Res.* (2020) 190:106171. doi: 10.1016/j.smallrumres.2020.106171
- Habier D, Fernando RL, Dekkers JC. The impact of genetic relationship information on genome-assisted breeding values. *Genetics.* (2007) 177:2389–97. doi: 10.1534/genetics.107.081190
- VanRaden P. Efficient methods to compute genomic predictions. *J Dairy Sci.* (2008) 91:4414–23. doi: 10.3168/jds.2007-0980
- De L, Naya H, Gianola D, Crossa J, Legarra A, Manfredi E, et al. Predicting quantitative traits with regression models for dense molecular markers and pedigree. *Genetics.* (2009) 182:375–85. doi: 10.1534/genetics.109.101501
- Park T, Casella G. The bayesian lasso. *J Am Stat Assoc.* (2008) 103:681–6. doi: 10.1198/016214508000000337
- Legarra A, Aguilar I, Misztal I. A relationship matrix including full pedigree and genomic information. *J Dairy Sci.* (2009) 92:4656–63. doi: 10.3168/jds.2009-2061
- Misztal I, Legarra A, Aguilar I. Computing procedures for genetic evaluation including phenotypic, full pedigree, genomic information. *J Dairy Sci.* (2009) 92:4648–55. doi: 10.3168/jds.2009-2064
- Aguilar I, Misztal I, Johnson DL, Legarra A, Tsuruta S, Lawlor TJ. Hot topic: a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *J Dairy Sci.* (2010) 93:743–52. doi: 10.3168/jds.2009-2730
- Christensen OF, Lund MS. Genomic prediction when some animals are not genotyped. *Genet Sel Evol.* (2010) 42:2. doi: 10.1186/1297-9686-42-2
- Legarra A, Ducrocq V. Computational strategies for national integration of phenotypic, genomic, and pedigree data in a single-step best linear unbiased prediction. *J Dairy Sci.* (2012) 95:4629–45. doi: 10.3168/jds.2011-4982
- Ahmed SE. Univariate and multivariate general linear modelstheory and applications with SAS. *Technometrics.* (2006) 50:97–8. doi: 10.1198/tech.2008.s544
- Wang Z, Wang R, Li J, Zhang W, Wuiriliga, Su R, et al. Modeling genetic covariance structure across ages of fleece traits in an Inner Mongolia cashmere goat population using repeatability and multivariate analysis. *Livestock Sci.* (2014) 161:1–5. doi: 10.1016/j.livsci.2013.11.028
- Zhiying W, Zhixin W, Yan L, Ruijun W, Yanjun Z. Genetic evaluation of fiber length and fiber diameter from inner mongolia white cashmere goats at different ages. *Small Ruminant Res.* (2015) 123:22–6. doi: 10.1016/j.smallrumres.2014.11.015

40. Hayes BJ, Corbet NJ, Allen JM, Laing AR, Fordyce G, Lyons R, et al. Towards multi-breed genomic evaluations for female fertility of tropical beef cattle. *J Anim Sci.* (2019) 97:55–62. doi: 10.1093/jas/sky417
41. Dassonneville R, Brøndum R, Druet T, Fritz S, Guillaume F, Guldbbrandtsen B, et al. Effect of imputing markers from a low-density chip on the reliability of genomic breeding values in Holstein populations. *J Dairy Sci.* (2011) 94:3679–86. doi: 10.3168/jds.2011-4299
42. Bolormaa S, Gore K, van der Werf J, Hayes B, Daetwyler H. Design of a low-density SNP chip for the main Australian sheep breeds and its effect on imputation and genomic prediction accuracy. *Animal Genet.* (2015) 46:544–56. doi: 10.1111/age.12340
43. Solberg TR, Sonesson A, Woolliams J, Meuwissen THE. *Genomic Selection Using Different Markers and Density.* In: *World Congress on Genetics Applied to Livestock Production* (2006). pp. 22–13.
44. Calus M, Meuwissen T, de Roos A, Veerkamp R. Accuracy of genomic selection using different methods to define haplotypes. *Genetics.* (2008) 178:553–61. doi: 10.1534/genetics.107.080838
45. Liu Z, Seefried FR, Reinhardt F, Rensing S, Thaller G, Reents R. Impacts of both reference population size and inclusion of a residual polygenic effect on the accuracy of genomic prediction. *Genet Sel Evol.* (2011) 43:19. doi: 10.1186/1297-9686-43-19
46. Carillier-Jacquin C. *Evaluation of a Reference Population in Dairy Goats for Genomic Selection.* Rapport de stage de fin d'étude de master, AgroParis Tech (2012).
47. Moser G, Khatkar MS, Hayes BJ, Raadsma HW. Accuracy of direct genomic values in holstein bulls and cows using subsets of SNP markers. *Genet Sel Evol.* (2010) 42:37. doi: 10.1186/1297-9686-42-37
48. Zhang Z, Liu J, Ding X, Bijma P, de Koning D, Zhang Q. Best linear unbiased prediction of genomic breeding values using a trait-specific marker-derived relationship matrix. *PLoS ONE.* (2010) 5:e12648. doi: 10.1371/journal.pone.0012648
49. Daetwyler H, Pong-Wong R, Villanueva B, Woolliams J. The impact of genetic architecture on genome-wide evaluation methods. *Genetics.* (2010) 185:1021–31. doi: 10.1534/genetics.110.116855
50. Vicoso B, Charlesworth B. Evolution on the X chromosome: unusual patterns and processes. *Nat Rev Genet.* (2006) 7:645–53. doi: 10.1038/nrg1914
51. Avendao S, Watson KA, Kranis A. *Genomics In Poultry Breeding – From Utopias To Deliverables.* In *Proceedings of the 9th World Congress on Genetics Applied to Livestock Production (WCGALP)*, Leipzig (2010) p. 1–6.
52. Gao H, Su G, Janss L, Zhang Y, Lund M. Model comparison on genomic predictions using high-density markers for different groups of bulls in the Nordic Holstein population. *J Dairy Sci.* (2013) 96:4678–87. doi: 10.3168/jds.2012-6406
53. Sun W, Ibrahim J, Zou F. Genomewide multiple-loci mapping in experimental crosses by iterative adaptive penalized regression. *Genetics.* (2010) 185:349–59. doi: 10.1534/genetics.110.114280
54. Lourenco DAL, Tsuruta S, Fragomeni BO, Masuda Y, Misztal I. Genetic evaluation using single-step genomic best linear unbiased predictor in American Angus. (2015) 93:2653–62. doi: 10.2527/jas.2014-8836
55. Teissier M, Larroque H, Robert-Granie C. Accuracy of genomic evaluation with weighted single-step genomic best linear unbiased prediction for milk production traits, udder type traits, and somatic cell scores in French dairy goats. *J Dairy Sci.* (2019) 102:3142–54. doi: 10.3168/jds.2018-15650

**Conflict of Interest:** TZ is only employed by the company Inner Mongolia Bigvet Co., Ltd.

The remaining 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.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Yan, Zhang, Liu, Yu, Yang, Han, Gong, Wang, Zhang, Liu, Li, Yan, Mao, Li, Du, Li, Zhang, Wang, Lv, Wang, Zhang, Liu, Wang and Su. 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.