Genomic prediction of carcass traits using different haplotype block partitioning methods in beef cattle

Hongwei Li | Zezhao Wang | Lei Xu | Qian Li | Han Gao | Haoran Ma | Wentao Cai | Yan Chen | Xue Gao | Lupei Zhang | Huijiang Gao | Bo Zhu | Lingyang Xu | Junya Li

Abstract
Genomic prediction (GP) based on haplotype alleles can capture quantitative trait loci (QTL) effects and increase predictive ability because the haplotypes are expected to be in linkage disequilibrium (LD) with QTL. In this study, we constructed haploblocks using LD-based and the fixed number of single nucleotide polymorphisms (fixed-SNP) methods with Illumina BovineHD chip in beef cattle. To evaluate the performance of different haplotype block partitioning methods, we constructed haploblocks based on LD thresholds (from $r^2 > 0.2$ to $r^2 > 0.8$) and the number of fixed-SNPs (5, 10, 20). The performance of predictive methods for three carcass traits including liveweight (LW), dressing percentage (DP), and longissimus dorsi muscle weight (LDMW) was evaluated using three approaches (GBLUP and BayesB model based on the SNP, $G_1$BLUP, and BayesBH models based on the haploblock, and $G_1$BLUP+GBLUP and BayesBH+BayesB models based on the combined haploblock and the nonblocked SNPs, which were located between blocks). In this study, we found the accuracies of LD-based and fixed-SNP haplotype Bayesian methods outperformed the Bayesian models (up to 8.54±7.44% and 5.74±2.95%, respectively). $G_1$BLUP showed a high improvement (up to 11.29±9.87%) compared with GBLUP. The Bayesian models have higher accuracies than BLUP models in most scenarios. The average computing time of the BayesBH+BayesB model can reduce by 29.3% compared with the BayesB model. The prediction accuracies using the LD-based haplotype method showed higher improvements than the fixed-SNP haplotype method. In addition, to avoid the influence of rare haplotypes generated from haplotype construction, we compared the performance of GP by filtering four types of minor haplotype allele frequency (MHAF) (0.01, 0.025, 0.05, and 0.1) under different conditions (LD levels were set at $r^2 > 0.3$, and the fixed number of SNPs was 5). We found the optimal MHAF threshold for LW was 0.01, and the optimal MHAF threshold for DP and LDMW was 0.025.
1 | INTRODUCTION

Genetic improvement for carcass traits is greatly important in the beef cattle breeding industry. Single nucleotide polymorphisms (SNP) are routinely applied to increase the prediction accuracy at a young age compared with the approach using pedigree information in farm animals (VanRaden, 2008). A haplotype block (haploblock) defines a region of the genome, comprising a set of neighboring SNPs whose alleles are likely to be inherited together (Hess et al., 2017). Haplotype methods have been previously proposed in human genetics (Chapman et al., 2003; Curtis, 2007; Curtis et al., 2001; Gabriel et al., 2002; North et al., 2006), and this strategy was widely used for the prediction of genomic breeding values in animal breeding studies (Calus et al., 2008, 2009; Cuyabano et al., 2014; Cuyabano, Su, & Lund, 2015; Hess et al., 2017; Li et al., 2021; Villumsen & Janss, 2009). However, the prediction accuracy can be influenced by the number of neighboring SNPs contained in a haploblock and the length of the haploblock. When the boundaries of haploblocks are determined appropriately, the haplotype-based methods can produce high accuracy over SNP-based methods for both real (Cuyabano, Su, Rosa, et al., 2015; Hess et al., 2017) and simulated data (Villumsen et al., 2009).

In general, the methods for constructing haploblocks can be divided into two types, including the fixed-length haplotypes and the variable-length haplotypes. The fixed-length haploblock methods include the fixed-SNP number (Calus et al., 2009; Hayes et al., 2007; Villumsen et al., 2009) and the fixed length (Hess et al., 2017) in a block, this method is easy to implement and costs less time; however, the population-specific linkage disequilibrium (LD) information were not included, which may greatly influence the haploblock construction. A variable-length method was proposed to obtain co-segregation of allele combinations based on LD information within the population, which can result in a decrease in the number of haplotype alleles (Rinaldo et al., 2005).

The accuracy of GP can be affected by the population size, the relationships of individuals, and the phasing approach. Determination of the haplotype phase is important in the development of large-scale sequencing (Browning & Browning, 2011). Many programs, such as PHASE (Crawford et al., 2004), BEAGLE (Browning & Browning, 2009), DAGPHASE (Drut & Georges, 2010), SHAPEIT (O’Connell et al., 2014), and LINKPHASE3 (Druet & Georges, 2015) were proposed and compared in previous studies (Choi et al., 2018). Among them, the BEAGLE (Browning & Browning, 2009) found GP based on fixed-length haplotype alleles rather than SNPs can increase the accuracy of genomic prediction (GP) up to 5.5%. The haplotype methods performed better than SNP methods in GP and are less sensitive to prior and hyperparameter settings (Villumsen & Janss, 2009). Previous studies mainly focused on the estimation of prediction accuracy using a single haploblock construction method. However, the accuracy of GP for economically important traits using different haploblock methods has not been fully explored and compared in beef cattle.

The objectives of the current study were to (1) assess the performance of GP for carcass traits and computation time based on different haploblock methods (the fixed-SNPs and the LD-based) using GBLUP and BayesB model; (2) compare the prediction accuracies under different LD thresholds and different number of SNPs; (3) evaluate the prediction accuracy with different MHAF thresholds for three traits.

2 | MATERIALS AND METHODS

2.1 | Data

The phenotypic data were generated from 1478 Chinese Simmental cattle, which were born between 2008 and 2020 from Ulga, Xilingol League, and Inner Mongolia, China. After weaning, all individuals were moved to Jinweifuren Co., Ltd. for fattening under the same feeding and management conditions. All individuals were slaughtered at an average age of 20±2.2 months. More detailed description of the management processes was reported in previous studies (Zhu et al., 2016, 2017). Carcass traits were measured by the guidelines proposed by the Institute of Meat Purchase Specifications established by the Agricultural Marketing Service (AMS) of the USDA.

### TABLE 1 Statistical description and heritability estimation for three carcass traits

| Trait | Number of phenotypes | Mean ± SD | Maximum | Minimum | $h^2 ± SE$ |
|-------|-----------------------|-----------|---------|---------|-----------|
| LW    | 1387                  | 518.55 ± 80.95 | 790.00 | 318.00 | 0.40 ± 0.06 |
| DP    | 1366                  | 0.54 ± 0.03    | 0.62   | 0.45   | 0.22 ± 0.06 |
| LDMW  | 1457                  | 32.78 ± 8.35  | 59.00  | 9.89   | 0.16 ± 0.05 |

Abbreviations: DP, dressing percentage; LDMW, longissimus dorsi muscle weight; LW, liveweight.
Liveweight (LW), dressing percentage (DP), and longissimus dorsi muscle weight (LDMW) were selected for subsequent analysis. After removing missing values and outliers, the basic statistics of phenotype data were shown in Table 1.

2.2 | Genotyping and phasing

The DNA was extracted from blood samples. The Illumina BovineHD BeadChips (777,962 SNPs) were used for genotyping. Before phasing the quality control of the original SNP dataset was carried out using PLINK (v1.07) (Chang et al., 2015; Purcell et al., 2007). Individuals and autosomal SNPs were filtered by the following criteria, SNPs call rate (<0.90), minor allele frequency (MAF <0.01), Hardy–Weinberg Equilibrium (p < 10^{-3}), and individual call rate (<0.90). Missing genotypes were imputed using BEAGLE (v4.1) (Browning & Browning, 2016). Consequently, 1478 individuals and 672,060 SNPs remained and further phased using BEAGLE (v4.1) (Browning & Browning, 2016) with default parameters.

2.3 | Heritability and variance component estimation

Multivariate variance analysis was used to determine the fixed effects of sex, year, fattening farm, and slaughterhouse and the covariates (body weight upon entering the fattening farm and the number of fattening days). The phenotypes were adjusted by the significant factors using glm function in R programming. Variance components were estimated using the following univariate animal model in ASREML (v4.1).

\[ y = 1_n \mu + Za + e, \] (1)

where \( y \) is the vector of the adjusted phenotypes, \( 1_n \) is an \( n \times 1 \) vector with entries equal to 1; \( \mu \) is the overall mean; \( a \sim N(0, \sigma^2_a G) \) is a vector of random additive genetic effect, where \( G \) is the additive genomic relationship matrix constructed using all SNPs, and \( \sigma^2_z \) is the additive genetic variance, \( Z \) is the incidence matrix linking \( a \) to \( y \); and \( e \sim N(0, \sigma^2_e I) \) is a vector of random residuals, where \( I \) is the identity matrix, and \( \sigma^2_e \) is the residual variance. The heritability was estimated as \( h^2 = \sigma^2_a / (\sigma^2_a + \sigma^2_e) \).

2.4 | Haplotype construction

2.4.1 | LD-based method

The LD information was used to determine the boundary of haploblocks for each chromosome. Due to the variation in local recombination rates, the breakdown of LD is often discontinuous and presents a haploblock-like structure (Daly et al., 2001). Therefore, it is important to analyze the haploblock structure and haplotypes that underlie LD.

A group of SNPs was defined as a haploblock if the LD between every two consecutive SNPs in the group was greater than or equal to the square of correlation coefficient (\( r^2 \)). There were no overlapping SNPs between adjacent haploblocks. For two bi-allelic loci \((A_1 / A_2 \text{ and } B_1 / B_2)\), \( r^2 \) was calculated as

\[ r^2 = \frac{D^2}{(p_{A_1}p_{A_2}p_{B_1}p_{B_2})}, \] (2)

where \( D = p_{A_1}p_{A_2}p_{B_1}p_{B_2} - p_{A_1}p_{A_2}p_{B_1}p_{B_2} \), \( D \) is the degree of LD for bi-allelic loci \((A_1 / A_2 \text{ and } B_1 / B_2)\); \( p_{A_1}p_{A_2}p_{B_1}p_{B_2} \) and \( p_{A_1}p_{A_2}p_{B_1}p_{B_2} \) are the alleles \((A_1, A_2, B_1, \text{ and } B_2)\) frequencies; \( p_{A_1}p_{A_2}p_{B_1}p_{B_2} \) and \( p_{A_1}p_{A_2}p_{B_1}p_{B_2} \) is the genotype \((A_1B_1, A_1B_2, A_2B_1, \text{ and } A_2B_2)\) frequencies (Hill, 1981; Hill & Robertson, 1968).

Seven different LD levels \((r^2 > 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, \text{ and } 0.8)\) were set as the thresholds in this study. To reduce the dimensionality of LD-based haplotype analyses, haplotype alleles with a minor allele frequency of less than 0.01 were discarded.

2.4.2 | Fixed-SNP method

The fixed-SNP haplotype was defined by the fixed number of SNPs in one haploblock (e.g., 5-SNPs per haploblock). Three different levels (5-, 10-, and 20-SNPs) were set to construct the haploblock. There were no overlapping SNPs between adjacent haploblocks. Consistent with the LD method, haplotype alleles with a minor allele frequency of less than 0.01 were discarded in the current study.

2.4.3 | MHAF thresholds set

To compare the haplotype structure distribution and performance of GP with different MHAF thresholds, the LD levels were set at \( r^2 > 0.3 \) and the fixed number of SNPs was 5, while the quality control of haplotype alleles was set at four different levels (0.01, 0.025, 0.05 and 0.1).

2.5 | GP models

Haplotype effects were modeled using numerical dosage coding strategies (Calus et al., 2008; Cuyabano et al., 2014; Cuyabano, Su, & Lund, 2015; Da, 2015; Meuwissen et al., 2014). A group of SNPs was contained in a haploblock. In the numerical dosage model, the SNPs were regarded as a unit and coded as the number of copies, which were termed as the haplotype alleles. The GP was performed using the GBLUP model and BayesB model, fitting covariates for either SNPs or haplotype alleles. The approaches include GBLUP and BayesB model (only based on the SNPs), \( G_B \text{ BLUP} \) and BayesB model (only based on haploblock), and \( G_B \text{ BLUP+GBLUP} \) and BayesB+Bayes model (contain both the haploblock and the nonblocked SNPs) were considered for predictions. Seven different \( r^2 \) thresholds and three levels of a fixed number of SNPs were used
for haplblock construction. The Bayesian algorithm was performed using GCTB (v2.0) (Zeng et al., 2018), a single Markov-chain Monte Carlo (MCMC) with a length of 30,000, while the first 10,000 cycles were discarded as the burn-in period.

2.5.1 | GBLUP model

The GBLUP model (including only the SNP effect) was used for LW, DP, and LDMW as described in Equation (1).

We performed GP using GBLUP for all SNPs, and the genomic relationship matrix was calculated as $G = \frac{(M-\tilde{p}M^\top - \tilde{p}^2)}{2\tilde{p}(1-\tilde{p})}$, where $M$ means the (0, 1, 2)-encoded genotype matrix, $p$ is the MAF of marker $i$, $m$ is the number of markers, and $P$ is a matrix with columns equal to $2p_i$ (Amin et al., 2007; Leutenegger et al., 2003).

The haplotype-based genomic best linear unbiased prediction (G$_m$BLUP) was performed for all SNPs. The haplotype-based genomic relationship matrix $G_h$ was expressed as $G_h = \frac{(M_h - \tilde{p}_hM_h^\top - \tilde{p}_h^2)}{2\tilde{p}_h(1-\tilde{p}_h)}$, where $M_h$ is the pseudo-markers matrix with entries 0, 1, and 2 representing the number of copies of each haplotype allele in a haploblock, and $p_h$ is the MAF of haplotype alleles $i$, where $n$ is the number of haplotype alleles of the whole genome, and $P_{hy}$ is a matrix with columns equal to $2p_h$. In the G$_m$BLUP+GBLUP model:

$$y = \beta_0 + Za + Za_2 + \epsilon,$$

which included the haplotype effects and the SNP effects estimated from outside the haplblocks (nonblocked SNPs). $\alpha \sim N(0, \sigma_v^2 G_s)$ is a vector of random additive genetic effect, where $G_s$ is the additive genetic relationship matrix constructed using haplotype alleles and $\sigma_v^2$ is the additive genetic variance based on the haplotype alleles, $Z$ is the incidence matrix associating $\alpha$, $a_2 \sim N(0, \sigma_v^2 G_s)$ is a vector of random additive genetic effect, where $G$ is the additive genetic relationship matrix constructed using nonblocked SNPs and $\sigma_v^2$ is the additive genetic variance based on the nonblocked SNPs, $Z_{ai}$ is the incidence matrix associating $a_i$ and $a_2$, and $\epsilon$ is composed of SNP effects estimated from outside the haplblocks.

2.5.2 | BayesB model

The BayesB, BayesBH, and BayesBH+BayesB models were defined by the Equation (4)

$$y = u + \sum_{j=1}^{M} Z_j \delta_j + \epsilon,$$

where $y$ is the vector of the adjusted phenotypes; $u$ denotes the overall mean; $M$ is the number of covariates for SNPs (BayesB model) or haplotype alleles (BayesBH model) or SNPs and haplotype alleles (BayesBH+BayesB model), $Z_j$ is a vector of alleles counts (0/1/2) at SNP $j$ or haplotype allele $j$. BayesB assumes that some of the $a_i$ have no effect. This is defined by the binary variable $\delta_j$ that represents whether covariate $j$ was fitted in the model according to hyperparameter $x$, such that $\delta = 1$ with probability $1-\pi$, or $\delta = 0$ with probability $\pi$. In this study, $\pi$ was set to 0.96, 0.995, and 0.995 for LW, DP, and LDMW, respectively, according to the previous studies (Wang et al., 2019). $\epsilon \sim N(0, \sigma_e^2)$ is a vector of residual effects.

2.6 | Evaluation of prediction performance

Both the accuracy and bias of GP were assessed based on 5-fold cross-validation (CV). The CV procedure is that the whole population was randomly divided into 5 groups, one of which was taken as the validation group each time, and the remaining 4 groups were used as the reference group. This procedure was randomly repeated 5 times for GBLUP and Bayesian models, respectively. The prediction accuracy is the average Pearson correlation coefficient between the adjusted phenotypic values and genomic estimated breeding values (GEBVs) in the validation groups (Bolormaa et al., 2013). The formula is as follows,

$$\text{Prediction accuracy} = \text{cor}(y, \text{gebv}),$$

where $y$ is the vector of the adjusted phenotypes, and $\text{gebv}$ is the vector of the GEBVs.

The regression coefficient of the adjusted phenotype on GEBVs for individuals in the validation group was obtained to measure the degree of inflation/deflation of prediction and was defined as follows,

$$b = \frac{\text{cov}(\text{gebv}, y)}{\text{var}(\text{gebv})},$$

where $y$ and $\text{gebv}$ is the same as that in Equation (5), the regression coefficient of unbiased models is expected to be close to 1, whereas values greater than 1 indicate a biased deflation prediction of GEBVs, and values smaller than 1 indicate a biased inflation prediction of GEBVs (Xu et al., 2020).

To determine whether haplotype methods can significantly improve prediction accuracy over the SNP methods, we used the one-sided paired t-test to examine the significance of the differences. The level of statistical significance was set at $p < 0.05$.

3 | RESULTS

3.1 | Heritability estimation and haplblock construction

As shown in Table 1, the estimated heritability of LW, DP, and LDMW using the univariate animal model were 0.40, 0.22, and 0.16, respectively. The statistical description of haplblock structure was shown in Table 2. Notably, we observed the number of haplblocks decreased with LD level (from 68,345 to 42,097), while the number of haplotype alleles decreased from 1,010,588 to
Under different LD levels (from $r^2 > 0.2$ to $r^2 > 0.8$), the average number of SNPs within each haploblock ranged from 3.53 to 5.17, while the maximum number of SNPs within each haploblock was from 53 to 99.

As for the fixed-SNP method, the number of haploblocks was from 33,603 to 134,412, while the number of fixed-SNPs was between 5 and 20. Details about haploblocks (the total number of haplotype alleles based on the haplotype-based method with $r^2 > r^2 > 0.3$ (0.3LD)) decreased from 605,908 to 403,841, while the number of fixed-SNPs were 5 (5-SNPs) decreased from 755,697 to 382,073. The number of alleles based on the two methods decrease by 33.35% and 49.44%. Our results revealed that 5-SNP methods contain more alleles with low MHAF. Details about haplotype structure at different MHAF thresholds were shown in Table 3.

### 3.2 | Performance of GP

#### 3.2.1 | Performance of GP based on different models

To assess the performances of GP, we compared six models (GBLUP, $G_J$BLUP, $G_J$BLUP+GBLUP, BayesB, BayesBH, and BayesBH+BayesB) using different haplotype construction methods. In this study, our results revealed the accuracies of GP decreased with the trait heritability. The accuracies of haplotype methods showed great improvement for all three traits under some LD compared with the GBLUP model. Among these traits, we observed the highest accuracies for LW, followed by DP and LDMW (Figure 1). Compared with the GBLUP model, the $G_J$BLUP showed $-6.35 \pm 9.39\%$ and $-11.29 \pm 9.87\%$ higher average accuracy for LDMW using LD-based and fixed-SNP method (Tables 4 and 5), while the $G_J$BLUP+GBLUP showed $-1.1 \pm 7.3\%$ higher average accuracy for LW across all LD levels (Table 4). The Bayesian models have high accuracies in most scenarios compared with BLUP models. Among the six models, BayesBH+BayesB had the best performance for all three traits (Figure 1) and showed increasing accuracy of GP for LDMW (up to 8.54% higher than SNP models) (Table 4). However, as for DP and LDMW, BayesBH model cannot improve the accuracy and have larger biases compared with the SNP-based method in most scenarios.

#### 3.2.2 | Accuracies and biases based on different LD levels

We constructed haploblocks and compared accuracies and bias based on different LD levels (from $r^2 > 0.2$ to $r^2 > 0.8$). Comparing with BayesB model, BayesBH+BayesB model with $r^2 > 0.3$ for LW ($h^2 = 0.40$) and DP ($h^2 = 0.22$) increased $-6.53 \pm 3.02\%$ and $0.72 \pm 5.64\%$, while this model increased $-8.54 \pm 7.44\%$ for LDMW ($h^2 = 0.16$) under $r^2 > 0.4$ (Table 4). The BayesBH had a better performance for LW in most LD levels than the BayesB model (Figure 1; Table 4). $G_J$BLUP models with $r^2 > 0.2$ for LDMW showed up to $6.35 \pm 9.39\%$ accuracy compared with the GBLUP model, while $G_J$BLUP models for LW and DP did not increase the accuracies at all LD levels (Table 4). As shown in Figure 1, the biases of Bayesian models were larger than the GBLUP models for all three traits, while the biases of $G_J$BLUP+GBLUP were similar to $G_J$BLUP models (Figure 1; Table S3). In addition, the low accuracy and large bias of GP were found for LW and LDMW with the strict LD (Figure 1).

### Table 2 | Haploblock structure of different haploblock construction methods

| Method       | $r^2$ and Number of SNPs | Number of haploblocks | Number of SNPs in blocks | Number of SNPs out of blocks | Number of SNPs per haploblock | Number of alleles |
|--------------|--------------------------|-----------------------|--------------------------|-------------------------------|-------------------------------|------------------|
|              | $r^2$ and Number of SNPs | Number of haploblocks | Number of SNPs in blocks | Number of SNPs out of blocks | Number of SNPs per haploblock | Number of alleles |
| LD           | 0.2                      | 68,345                | 353,356                  | 318,704                       | 5.17                          | 1,010,588        |
|              | 0.3                      | 64,765                | 300,455                  | 371,605                       | 4.64                          | 914,760          |
|              | 0.4                      | 60,471                | 260,530                  | 411,530                       | 4.31                          | 856,417          |
|              | 0.5                      | 56,035                | 227,982                  | 444,078                       | 4.07                          | 817,085          |
|              | 0.6                      | 51,704                | 199,825                  | 472,235                       | 3.86                          | 788,801          |
|              | 0.7                      | 47,134                | 174,074                  | 497,986                       | 3.69                          | 766,424          |
|              | 0.8                      | 42,097                | 148,271                  | 523,789                       | 3.52                          | 747,477          |
| Fixed-SNPs   | 0                        | 134,412               | 672,060                  | 0                             | 5                             | 1,526,223        |
|              | 10                       | 67,206                | 672,060                  | 0                             | 10                            | 2,122,577        |
|              | 20                       | 33,603                | 672,060                  | 0                             | 20                            | 2,950,340        |

Abbreviations: Fixed-SNPs, fixed-SNPs Haploblock construction method; LD, LD-based Haploblock construction method; LD, linkage disequilibrium; SNP, single nucleotide polymorphisms.

*The seven different LD thresholds set from $r^2 > 0.2$ to $r^2 > 0.8$ to construct LD-based haploblocks.

*The three levels of the number of SNPs (5, 10, and 20) to construct fixed-SNP haploblocks.
3.2.3 Accuracies and biases of GP based on different numbers of fixed-SNPs

To evaluate the accuracies of GP using the fixed-SNP method, we set three different levels (5, 10, and 20-SNPs) based on the number of fixed-SNPs. Four models (GBLUP, G\textsubscript{H}BLUP, BayesB, and BayesBH) were used to investigate the performance of GP. We found that the 5-SNPs G\textsubscript{H}BLUP model can significantly increase (p < 0.05) the accuracies of GP for all three traits compared with the GBLUP model, while the bias between G\textsubscript{H}BLUP based on fixed-SNPs and GBLUP model significantly increased with the number of fixed-SNPs (Figure 2; Table 5; Tables S2 and S3). As shown in Figure 2 and Table 5, all three fixed-SNP BayesBH models (5-SNPs BayesBH, 10-SNPs BayesBH, 20-SNPs BayesBH) tended to have higher accuracies than BayesB.
**TABLE 4** Percent change of predictive accuracies using the LD-based haplotype model rather than SNPs for three traits in Chinese Simmental beef cattle (±SE)

| Trait         | LD-based method | 0.2   | 0.3   | 0.4   | 0.5   | 0.6   | 0.7   | 0.8   | Mean  |
|---------------|-----------------|-------|-------|-------|-------|-------|-------|-------|-------|
| \( G_{1} \)BLUP | LW              | -1.24 | -1.26 | -1.89 | -2.09 | -1.74 | -2.681| -3.83 | -2.1  |
|               | DP              | -0.073| -0.41 | 0.10  | -0.73 | -1.06 | -1.367| -2.008| -0.79 |
|               | LDMW            | 6.35  | 2.83  | 2.71  | 3.68  | 1.96  | 3.251 | 5.12  | 3.7   |
| \( G_{1} + G \) | LW              | -2.25 | -0.086| 1.21  | 1.76  | 2.41  | 2.72  | 1.90  | 1.1   |
|               | DP              | 0.44  | 0.65  | -0.019| -0.043| -0.088| -0.395| -0.59 | -0.01 |
|               | LDMW            | 1.85  | 2.42  | 1.83  | 1.44  | 1.48  | 1.224 | 0.97  | 1.6   |
| BayesBH       | LW              | 5.18  | 5.31  | 5.42  | 5.16  | 5.43  | -1.08 | -1.26 | 3.45  |
|               | DP              | -1.89 | -0.97 | -0.17 | -1.03 | -1.21 | -2.16 | -1.88 | -1.33 |
|               | LDMW            | -4.27 | -3.75 | -1.13 | -0.93 | -4.52 | -0.67 | -1.54 | -2.4  |
| BH+B          | LW              | 5.81  | 5.63  | 5.41  | 1.15  | 0.8   | 1.79  | 0.77  | 3.18  |
|               | DP              | 0.22  | 0.72  | -0.16 | -0.66 | -0.08 | -0.48 | 0.48  | 0.01  |
|               | LDMW            | 0.64  | 0.25  | 8.54  | 8.13  | -1.13| -0.64 | -0.73 | 2.15  |

Abbreviations: BayesBH, BayesB model based on haplotype; BH+B, BayesBH+BayesB model; DP, dressing percentage; \( G_{1} + G \), \( G_{1} \)BLUP+GBLUP model; \( G_{1} \)BLUP, GBLUP model based on haplotype; LD, linkage disequilibrium; LDMW, longissimus dorsi muscle weight; LW, liveweight; SNP, single nucleotide polymorphisms.

*The seven different LD thresholds set from \( r^2 \) > 0.2 to \( r^2 \) > 0.8 to construct LD-based haploblocks.

*Higher accuracy than the SNP model \((p<0.05)\); **Higher accuracy than the SNP model \((p<0.01)\).
TABLE 5 Percent change of predictive accuracies using fixed-SNPs haplotype model rather than SNPs for three traits in Chinese Simmental beef cattle (±SE).

| Model     | Trait | Fixed-SNP method* |
|-----------|-------|-------------------|
|           |       | 5     | 10    | 20    | Mean      |
| $G_{BH}$BLUP | LW    | 1.09±6.41* | 1.03±6.47 | 0.57±6.5 | 0.9±6.46  |
|           | DP    | 0.78±6.66**  | 0.29±6.59 | 0.07±6.62 | 0.38±6.63 |
|           | LDMW  | 10.52±9.68**  | 11.29±9.87** | 10.24±9.54** | 10.68±9.7  |
| BayesBH   | LW    | 5.74±2.95**  | 1.03±3   | 0.61±3.05 | 2.46±3    |
|           | DP    | 0.2±5.49     | −0.45±5.44 | −0.76±5.37 | −0.34±5.43 |
|           | LDMW  | 3.61±6.41    | 3.7±6.62  | 3.65±6.64 | 3.65±6.56 |

Abbreviations: DP, dressing percentage; LDMW, longissimus dorsi muscle weight; LW, liveweight; SNP, single nucleotide polymorphisms.

*The three levels of number of SNPs (5, 10, and 20) to construct fixed-SNPs haplotype.

**Higher accuracy than the SNP model ($p<0.05$); ***Higher accuracy than the SNP model ($p<0.01$).

3.2.4 Comparison of performance of GP with different MHAF

To evaluate the prediction accuracy with different MHAF thresholds, the haplotype alleles frequency less than 0.01, 0.025, 0.05, and 0.1 were removed, respectively. The prediction performance of GP using four models, such as BayesBH+BayesB, $G_{BH}$BLUP+GBLUP, BayesBH, and $G_{BH}$BLUP were compared at different MHAF thresholds. For LW, 0.3LD and 5-SNPs haplotype using BayesBH+BayesB model showed the highest when the MHAF threshold was greater than 0.01 (Figures 3 and 4; Tables S4 and S5). As for DP and LDMW, when MHAF was greater than 0.025, 0.3LD haplotype and 5-SNPs haplotype using the BayesBH+BayesB model showed the highest accuracy. The accuracies improved by 1.05%–1.57% for DP and 4.04%–10.46% for LDMW, respectively. In addition, 0.3LD and 5-SNPs haplotype method can reduce the computing time by 30.59%–31.6% and 9.64%–16.49%, compared with the SNP method (Tables S4 and S5).

FIGURE 2 Prediction accuracies and bias of different number of SNPs in a block for three traits based on the SNP and haplotype alleles using GBLUP and BayesB models. DP, dressing percentage; LD, linkage disequilibrium; LDMW, longissimus dorsi muscle weight; LW, liveweight; SNP, single nucleotide polymorphisms.
Number of covariates and computation time of different models

The average number of covariates fitted in models for three traits and the corresponding computation time were shown in Table 6. In the study, both the number of covariates and computation time decreased with the LD (from $r^2 > 0.2$ to $r^2 > 0.8$) and the increase of the number of fixed-SNPs (from 5 to 20). The computation time of GP using the LD-based BayesBH+BayesB model was similar to that of the fixed-SNPs BayesBH model. The computation time of the $G_p$BLUP+GBLUP model was slightly larger than $G_p$BLUP, and GBLUP model due to long time was required to calculate the covariance of the two $G$ matrices in the $G_p$BLUP+GBLUP model. Because the MCMC sampling often takes at least an order of magnitude, the Bayesian models cost much more time to fit covariates compared with the GBLUP models (Table 6). The Bayesian haplotype methods have fewer covariates than the SNP methods with 1% SNP allele frequency filter. The LD-based BayesBH+BayesB model can reduce by 29.3% of the average computation time compared with the BayesB SNP model. However, the accuracy of BayesBH+BayesB models was larger than the BayesB model in most scenarios (Figure 2 and Table 5). It is worth mentioning that the LD-based BayesBH model can drastically reduce the computation time because the few covariates were fitted in the model (Table 6), while the accuracy of GP for LW using the LD-based BayesBH model is higher than that of the BayesB model (Figure 2 and Table 5).

4 | DISCUSSION

The genomic selection has been widely applied to many farm animals (Meuwissen et al., 2001). Various methods based on the
| Method   | Computation time (h) | Number of covariates |
|----------|----------------------|----------------------|
|          | GBLUP | $G_{H}^{*}$BLUP | $G_{H}^{*}$+G | BayesB | BayesBH | BH+B | GBLUP | $G_{H}^{*}$BLUP | $G_{H}^{*}$+G | BayesB | BayesBH | BH+B |
| SNP      |      |                |                |        |         |      | 0.006 |                |                |        |         |      |
| LD       |      |                |                |        |         |      | 19.41 |                |                |        |         |      |
| 0.2      |      | 0.025           | 0.05            | 5.37   | 14.425  |      | 691.884 | 1,010,588 |        | 280,575 |        | 598,160 |
| 0.3      |      | 0.022           | 0.05            | 4.23   | 13.71   |      | 543.155 | 914,760   |        | 235,534 |        | 605,908 |
| 0.4      |      | 0.02            | 0.054           | 4.02   | 14.16   |      | 444.887 | 856,417   |        | 200,621 |        | 610,845 |
| 0.5      |      | 0.024           | 0.057           | 3.27   | 13.23   |      | 373.007 | 817,085   |        | 172,198 |        | 614,902 |
| 0.6      |      | 0.02            | 0.062           | 2.7    | 13.59   |      | 316.566 | 788,801   |        | 148,697 |        | 619,493 |
| 0.7      |      | 0.025           | 0.058           | 2.4    | 13.56   |      | 268.438 | 766,424   |        | 127,20  |        | 623,674 |
| 0.8      |      | 0.024           | 0.06            | 1.83   | 13.38   |      | 223.688 | 747,747   |        | 105,810 |        | 628,034 |
| Fix-SNPs |      |                |                |        |         |      | 0.031 |                |                |        |         |      |
| 5        |      |                |                |        |         |      | 19.73 |                |                |        |         |      |
| 10       |      |                |                |        |         |      | 17.93 |                |                |        |         |      |
| 20       |      |                |                |        |         |      | 16.67 |                |                |        |         |      |

Abbreviations: Fixed-SNPs, fixed-SNPs Haploblock construction method; LD, LD-based Haploblock construction method; LD, linkage disequilibrium; SNP, single nucleotide polymorphisms.

- The seven different LD thresholds set from $r^2 > 0.2$ to $r^2 > 0.8$ to construct LD-based haploblocks.
- The three levels of number of SNPs (5, 10, and 20) to construct fixed-SNP haploblocks.
- Computation time for running the analysis based on GBLUP and Bayesian model with a chain length of 30,000.
- The number of SNPs or haplotype alleles fitted in models.
SNPs were used to improve the accuracies of GP (Gianola, 2013; VanRaden, 2008; Yin et al., 2020). Haploblock comprising a set of neighboring SNPs (likely to be inherited together) has a higher LD with quantitative trait loci (QTLs) than SNPs, thus GP method integrating haplotype should be effective to estimate the effect of QTLs and can further improve the prediction accuracies. Our study evaluated the performance of GP using several models (BayesBH+BayesB, G_HBLUP+GBLUP, BayesBH, and G_HBLUP) based on the different haplotype construction methods in beef cattle. We found the haplotype-based GP generated better performance compared with SNPs, which was consistent with previous studies (Cuyabano et al., 2014; Cuyabano, Su, & Lund, 2015; Hess et al., 2017; Li et al., 2021). In addition, GP based on haplotype was an effective method to reduce the number of variables and compress the computation time, thus this method is beneficial for GP using whole-genome sequence data.

4.1 | Analysis of haplotype structure using different construction methods

We found the haplotype structure showed diverse patterns using different haplotype construction methods based on LD and fixed number of SNPs. The number of haploblocks decreased under thresholds with LD (from $r^2 > 0.2$ to $r^2 > 0.8$) and the number of fixed-SNPs (from 5-SNPs to 20-SNPs). These results were similar to previous studies (Hess et al., 2017). We also observed that the number of haplotype alleles (covariates) using fixed-SNP methods was larger than that of LD-based methods (Table 6), this can be explained that the fixed-SNP method used all the SNPs thus it produced more haplotype alleles. Moreover, the previous study presented that variable-length haploblocks can reflect the true haplotype structure compared with fixed-length haploblocks due to different LD levels among populations, which cannot be reflected by fixed-length haploblocks (Hess et al., 2017).

4.2 | Haplotype-based GP using different construction methods

We evaluated the performance of GP for three traits (with heritability ranging from 0.16 to 0.40) based on diverse LD levels and different numbers of fixed-SNPs. We found that the haplotype methods can produce high accuracies across the three traits in most scenarios when LD thresholds were $r^2 > 0.3$ or $r^2 > 0.4$ and the number of fixed-SNPs was five (Figures 1 and 2; Tables 4 and 5). As shown in Table 6, the average number of SNPs per haplloblock was 4.64 and 4.31 under the thresholds $r^2 > 0.3$ or $r^2 > 0.4$. Therefore, the haplotype structure using two construction methods may be similar under $r^2 > 0.3$ or $r^2 > 0.4$ and the number of fixed-SNPs was five (Table 6). Cuyabano et al. (2014) reported that the LD-based haploblocks can lead to the highest accuracies for milk protein, fertility, and mastitis traits when $D' \geq 0.45$. Villumsen and Janss (2009) also revealed the 5-SNP method can produce the highest prediction accuracy when GP based on haplotypes were investigated in simulated data. In addition, Hess et al. (2017) found the short haploblock (250kb) can generate an optimal set of variables in admixed dairy cattle population, and the average number of SNPs per haplloblock in 250kb haplotype was four, which was similar to our study.

In this study, the LD-based methods showed slightly higher accuracy under some thresholds compared with the accuracies of GP using fixed-SNP methods. One possible explanation is that LD-based haploblocks can reflect the true block structure of the population compared with fixed-SNPs haploblocks. However, LD-based methods did not consistently show improvement over SNPs or fixed-SNPs haploblocks as expected for all three traits (Figure 1), which may be caused by the different genetic architectures among traits.

4.3 | Performance of different GP models

The accuracies for GP using most haplotype methods are improved compared with the SNP methods in the current study. One possible explanation was the high LD between the haplotype alleles and linked QTLs (Zondervan & Cardon, 2004). Sun et al. (2013) showed that the effects of low-MAF QTLs can be effectively captured by haplotype alleles. Moreover, the haplotype contained a group of SNPs, which may capture short-range interaction among SNPs and improve the accuracies of GP (Kühn et al., 2004; Littlejohn et al., 2014). We also found that the BayesBH+BayesB models were robust and suitable for haplotype analyses in most scenarios. Similar to our results, the previous study reported that the Bayesian model with haploblocks and nonblocked SNPs can produce the highest accuracy for mastitis, which was the most unstable trait for genetic prediction (Cuyabano et al., 2014). The BayesBH+BayesB models included the haplotype alleles in blocks and SNPs alleles outside the blocks. Many previous studies have proved that the Bayesian model can exploit LD information between SNPs and their linked QTLs compared with the GBLUP model (Gao et al., 2013; Habier et al., 2010; Ostersen et al., 2011). The advantage can further be amplified with Bayesian haplotype models due to the high LD between QTLs and haplotypes (Zondervan & Cardon, 2004). As for all three traits, the fixed-SNP $G_{H}$BLUP model has a high improvement for GP compared with the GBLUP model (Figure 2; Table 5). The reason can be explained that the true IBD relationship was captured by the haplotype alleles (combinations of phased SNP alleles) than the individual SNP alleles (Ferdosi et al., 2016; Hickey et al., 2013). Therefore, $G_{H}$ matrix improved the ability to detect ancestral relationships and led to higher accuracy of GP than G matrix. However, haploblock length can greatly influence the ability to capture IBD relationships, if the haploblock length was not suitable, they cannot reflect the true relationships between relatives (Ferdosi et al., 2016). Thus, the improvement of accuracy using the 20-SNPs $G_{H}$BLUP model was not significant for LW and DP compared with GBLUP (Figure 2; Table 5).
In the present study, haplotype-based BayesBH methods have larger biases than the SNP-based methods in most scenarios. The LD-based methods were used to preselect SNPs with a certain degree of LD. The SNP effects in each of the haploblock may be overestimated due to “Beavis effect” (Xu, 2003), which indicates that the reported effects of SNPs tend to be larger than the true effects of them, which may lead to potential bias for GEBV (Zollner & Pritchard, 2007).

4.4 Number of covariates and computation time

The computation time for the BayesB model was more than the GBLUP model (Table 6). This was because that more time was required to fit SNPs in each round of a Markov-chain Monte Carlo (MCMC) algorithm and estimated the SNP effects (Meuwissen et al., 2001). Our results also showed that the haplotype methods can produce more haplotype alleles with low frequency (rare haplotype alleles), especially for the long haploblocks (Tables 2 and 6). These rare haplotype alleles would affect prediction accuracy because the rare haplotype alleles are not possible to explain more genetic variance due to their low frequency, therefore the effect of rare alleles may be shrunk to zero (Gianola, 2013). The haplotype alleles with MHAF less than 0.01 were discarded in the fixed-SNPs Bayesian haplotype model. We found the number of covariates of haplotype methods was reduced to 17%–50% for all haplotype alleles (Tables 2 and 6). The prediction accuracies of haplotype methods were slight decrease as a large number of low-frequency haplotype alleles removed. This may be the potential reason why we observed the drop in the accuracy of GP for long haploblocks (i.e., 20-SNPs haploblocks) (Figure 2).

4.5 Comparison of fixed-SNPs and LD-based haplotype

Two methods (fixed-SNPs and LD-based) were used to construct the haploblock, the LD-based methods produce a high accuracy of GP compared with the fixed-SNPs methods (Tables S1 and S2). A previous study has reported that using methods by defining haploblocks with LD information (variable-length haplotypes) can generate higher prediction accuracies than fixed-length haploblocks (Hess et al., 2017), which was agreed with our findings. The LD between haplotype alleles and QTL may vary because of the recombination and artificial selection (Goddard et al., 2010). The optimal haploblock lengths may differ across the genome, thus application of variable-length haplblock is likely to be an effective method that can contain the LD or recombination information. By contrast, the fixed-SNP methods can not reveal the variation of LD. Many methods based on recombination (Heinemann, 2019), IBD probabilities (Calus et al., 2008, 2009), and smoothing spline techniques (Beissinger et al., 2015) were proposed to define the variable-length haploblocks and further promote the improvement of GP. These methods can greatly enrich the methods of haplotype construction. The length of these haploblocks construction methods can vary along the genome of different populations.

4.6 Performance of GP with different MHAF

In this study, the optimal MHAF threshold for LW was set at 0.01, and the optimal MHAF threshold for DP and LDMW was set at 0.025. Under the optimal MHAF threshold, we found the accuracies of GP increased using the haplotype-based method compared with the SNP method, while the computation time was reduced. Hess et al. (2017) suggested that haplotype-based GP with the most frequent haplotype alleles can reduce computation time for short haplotypes in dairy cattle, while only a slight decrease in prediction accuracy was observed. Moreover, many studies showed that haplotype-based GP with higher frequency was more accurate than that with lower frequency (Boleckova et al., 2012; Kolbehdari et al., 2007). By contrast, previously suggested that computation time was reduced with the little expected decrease in prediction accuracy by removing rare haplotype alleles because the effect of haplotype alleles is shrunk toward zero in Bayesian linear regression models (Gianola, 2013).

5 CONCLUSION

Our study showed that haplotype-based Bayesian and GBLUP methods can improve the prediction accuracy up to 8.54% and 9.93% compared with the individual SNP approach. The LD-based haplotype approach has larger improvements than that of fixed-SNPs. When LD thresholds were \( r^2 > 0.3 \) or \( r^2 > 0.4 \) and the number of fixed-SNPs was five, the Bayesian haplotype methods (BayesBH+BayesB) showed high accuracies across the three traits in most scenarios. Meanwhile, the average computation time of the LD-based BayesBH+BayesB method can be reduced by 29.3% compared with the BayesB SNP method. LD-based GBLUP+GBLUP and fixed-SNPs-based GBLUP methods using the 5-SNPs can increase the accuracy of GP compared with GBLUP for all three traits. The optimal MHAF threshold for LW was 0.01, and the optimal MHAF threshold for DP and LDMW was 0.025. Further studies were required to evaluate the relationships within and between training and validation sets and further improve the accuracy of GP.

ACKNOWLEDGMENTS

The authors would like to thank all staff at the cattle experimental unit in Beijing and Ulagai for animal care and sample collection.

FUNDING INFORMATION

This work was supported by the National Natural Science Foundation of China (31802049, 31872975, and 31972554), the Chinese Academy of Agricultural Sciences of Technology Innovation Project (CAAS-ZDRW202102, CAAS-XTCX2016010, CAAS-ZDXT2018006 and ASTIP-IAS03, ASTIP-IAS-TS-16), the
Beijing Natural Science Foundation (6154032), the National Beef Cattle Industrial Technology System (CARS-37), the China Agriculture Research System of MOF and MARA, and the Science and Technology Project of Inner Mongolia Autonomous Region (2020GG0210).

CONFLICT OF INTEREST
The authors declare that they have no competing interests.

DATA AVAILABILITY STATEMENT
Partial genotype data have been submitted to Dryad: doi:10.5061/dryad.4qc06. The phenotype data used during the current study are available from the corresponding author upon reasonable request after the manuscript is accepted for publication.

ORCID
Junya Li https://orcid.org/0000-0001-8051-050X

REFERENCES
Amin, N., van Duijn, C. M., & Aulchenko, Y. S. (2007). A genomic background based method for association analysis in related individuals. PLoS One, 2(12), e1274. https://doi.org/10.1371/journal.pone.0001274
Beissinger, T. M., Rosa, G. J., Kaeppler, S. M., Gianola, D., & de Leon, N. (2015). Defining window-boundaries for genomic analyses using smoothing spline techniques. Genetics Selection Evolution, 47(1), 30. https://doi.org/10.1186/s12711-015-0105-9
Boleckova, J., Christensen, O., Sørensen, P., & Sahana, G. (2012). Strategies for haplotype-based association mapping in a complex pedigreed population. Czech Journal of Animal Science, 57(1), 1–9.
Bolormaa, S., Pryce, J. E., Kemper, K., Savin, K., Hayes, B. J., Barendse, W., Zhang, Y., Reich, C. M., Mason, B. A., Bunch, R. J., Harrison, B. E., Reverter, A., Herd, R. M., Tier, B., Graser, H. U., & Goddard, M. E. (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 cattle1. Journal of Animal Science, 91(7), 3088–3104. https://doi.org/10.2527/jsas.2012-5827
Brownning, B. L., & Browning, S. R. (2009). A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. American Journal of Human Genetics, 84(2), 210–223. https://doi.org/10.1016/j.ajhg.2009.01.005
Brownning, B. L., & Browning, S. R. (2016). Genotype imputation with millions of reference samples. American Journal of Human Genetics, 98(1), 116–126. https://doi.org/10.1016/j.ajhg.2015.11.020
Brownning, S. R., & Browning, B. L. (2011). Haplotype phasing: Existing methods and new developments. Nature Reviews Genetics, 12(10), 703–714. https://doi.org/10.1038/nrg3054
Calus, M. P., Meuwissen, T. H., de Roos, A. P., & Veerkamp, R. F. (2008). Accuracy of genomic selection using different methods to define haplotypes. Genetics, 178(1), 553–561. https://doi.org/10.1534/genetics.107.080838
Calus, M. P., Meuwissen, T. H., Windig, J. J., Knol, E. F., Schrootens, C., Vereijken, A. L., & Veerkamp, R. F. (2009). Effects of the number of markers per haplotype and clustering of haplotypes on the accuracy of QTL mapping and prediction of genomic breeding values. Genetics Selection Evolution, 41(1), 11. https://doi.org/10.1186/1297-9686-41-11
Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. Gigascience, 4(7). https://doi.org/10.1186/s13742-015-0047-8
Chapman, J. M., Cooper, J. D., Todd, J. A., & Clayton, D. G. (2003). Detecting disease associations due to linkage disequilibrium using haplotype tags: A class of tests and the determinants of statistical power. Human Heredity, 56(1–3), 18–31. https://doi.org/10.1159/000073729
Choi, Y., Chan, A. P., Kirkness, E., Telenti, A., & Schork, N. J. (2018). Comparison of phasing strategies for whole human genomes. PLoS Genetics, 14(4), e1007308. https://doi.org/10.1371/journal.pgen.1007308
Crawford, D. C., Banghale, T., Li, N., Hennenthal, G., Rieder, M. J., Nickerson, D. A., & Stephens, M. (2004). Evidence for substantial fine-scale variation in recombination rates across the human genome. Nature Genetics, 36(7), 700–706. https://doi.org/10.1038/ng1376
Curtis, D. (2007). Comparison of artificial neural network analysis with other multimarker methods for detecting genetic association. BMC Genetics, 8, 49. https://doi.org/10.1186/1471-2156-8-49
Curtis, D., North, B. V., & Sham, P. C. (2001). Use of an artificial neural network to detect association between a disease and multiple marker genotypes. Annals of Human Genetics, 65(1 Pt 1), 95–107. https://doi.org/10.1046/j.1469-1809.2001.6510093.x
Cuyabano, B. C., Su, G., & Lund, M. S. (2014). Genomic prediction of genetic merit using LD-based haplotypes in the Nordic Holstein population. BMC Genomics, 15(1), 1171. https://doi.org/10.1186/1471-2164-15-1171
Cuyabano, B. C., Su, G., & Lund, M. S. (2015). Selection of haplotype variables from a high-density marker map for genomic prediction. Genetics, Selection, Evolution, 47(1), 61. https://doi.org/10.1186/s12711-015-0143-3
Cuyabano, B. C., Su, G., Rosa, G. J., Lund, M. S., & Gianola, D. (2015). Bootstrap study of genome-enabled prediction reliabilities using haplotype blocks across Nordic Red cattle breeds. Journal of Dairy Science, 98(10), 7351–7363. https://doi.org/10.3168/jds.2015-9360
Da, Y. (2015). Multi-allelic haplotype model based on genetic partition for genomic prediction and variance component estimation using SNP markers. BMC Genetics, 16, 144. https://doi.org/10.1186/s12863-015-0301-1
Daly, M. J., Rioux, J. D., Schaffner, S. F., Hudson, T. J., & Lander, E. S. (2001). High-resolution haplotype structure in the human genome. Nature Genetics, 29(2), 229–232. https://doi.org/10.1038/ng1001-229
Druet, T., & Georges, M. (2010). A hidden markov model combining linkage and linkage disequilibrium information for haplotype reconstruction and quantitative trait locus fine mapping. Genetics, 184(3), 789–798. https://doi.org/10.1534/genetics.109.108431
Druet, T., & Georges, M. (2015). UNKPHASE3: An improved pedigree-based phasing algorithm robust to genotyping and map errors. Bioinformatics, 31(10), 1677–1679. https://doi.org/10.1093/bioinformatics/btu859
Ferdosi, M. H., Henshall, J., & Tier, B. (2016). Study of the optimum haplotype length to build genomic relationship matrices. Genetics Selection Evolution, 48(1), 75. https://doi.org/10.1186/s12711-016-0253-6
Gabriel, S. B., Schaffner, S. F., Nguyen, H., Moore, J. M., Roy, J., Blumenstiel, B., Higgins, J., DeFelice, M., Lochner, A., Faggart, M., Liu-Cordero, S. N., Rotimi, C., Adeyemo, A., Cooper, R., Ward, R., Lander, E. S., Daly, M. J., & Altshuler, D. (2002). The structure of haplotype blocks in the human genome. Science, 296(5576), 2225–2229. https://doi.org/10.1126/science.1069424
Gao, H., Su, G., Janss, L., Zhang, Y., & Lund, M. S. (2013). Model comparison on genomic predictions using high-density markers for different groups of bulls in the Nordic Holstein population. Journal
North, B. V., Sham, P. C., Knight, J., Martin, E. R., & Curtis, D. (2006). Investigation of the ability of haplotype association and logistic regression to identify associated susceptibility loci. *Annals of Human Genetics, 70*(Pt 6), 893–906. https://doi.org/10.1111/j.1469-1809.2006.00301.x

O’Connell, J., Gurdasani, D., Delaney, O., Pirastu, N., Ulivi, S., Cocca, M., Traglia, M., Huang, J., Huffman, J. E., Rudan, I., McQuillan, R., Fraser, R. M., Campbell, H., Polasek, O., Aslak, G., Ekoru, K., Hayward, C., Wright, A. F., Vitart, V., ... Marchini, J. (2014). A general approach for haplotype phasing across the full spectrum of relatedness. *PLoS Genetics, 10*(4), e1004234. https://doi.org/10.1371/journal.pgen.1004234

Ostersen, T., Christensen, O. F., Henryon, M., Nielsen, B., Su, G., & Madsen, P. (2011). Deregressed EBV as the response variable yield more reliable genomic predictions than traditional EBV in pure-bred pigs. *Genetics Selection Evolution, 43*(1), 38. https://doi.org/10.1186/1297-9686-43-38

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., de Bakker, P. I., Daly, M. J., & Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics, 81*(3), 559–575. https://doi.org/10.1086/519795

Rinaldo, A., Bacanu, S. A., Devlin, B., Sonpar, V., Wasserman, L. M., & Roeder, K. (2005). Characterization of multilocus linkage disequilibrium. *Genetic Epidemiology, 28*(3), 193–206. https://doi.org/10.1002/gepi.20056

Sun, X., Fernando, R. L., Garrick, D. J., & Dekkers, J. (2013). Genomic prediction using linkage disequilibrium and co-segregation. *Dissertations & Theses Gradworks.*

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *Journal of Dairy Science, 91*(11), 4414–4423. https://doi.org/10.3168/jds.2007-0980

Villumsen, T. M., & Janss, L. (2009). Bayesian genomic selection: The effect of haplotype length and priors. *BMC Proceedings, 3*(Suppl 1), S11. https://doi.org/10.1186/1753-6561-3-s1-s11

Villumsen, T. M., Janss, L., & Lund, M. S. (2009). The importance of haplotype length and heritability using genomic selection in dairy cattle. *Journal of Animal Breeding and Genetics, 126*(1), 3–13. https://doi.org/10.1111/j.1439-0388.2008.00747.x

Wang, X., Miao, J., Chang, T., Xia, J., An, B., Li, Y., Xu, L., Zhang, L., Gao, X., Li, J., & Gao, H. (2019). Evaluation of GBLUP, BayesB and elastic net for genomic prediction in Chinese Simmental beef cattle. *PLoS One, 14*(2), e0210442. https://doi.org/10.1371/journal.pone.0210442

Xu, L., Gao, N., Wang, Z., Xu, L., Liu, Y., Chen, Y., Xu, L., Gao, X., Zhang, L., Gao, H., Zhu, B., & Li, J. (2020). Incorporating genome annotation into genomic prediction for carcass traits in Chinese Simmental beef cattle. *Frontiers in Genetics, 11*, 481. https://doi.org/10.3389/fgene.2020.00481

Xu, S. (2003). Theoretical basis of the Beavis effect. *Genetics, 165*(4), 2259–2268. https://doi.org/10.1093/genetics/165.4.2259

Yin, L., Zhang, H., Zhou, X., Yuan, X., Zhao, S., Li, X., & Liu, X. (2020). KAML: Improving genomic prediction accuracy of complex traits using machine learning determined parameters. *Genome Biology, 21*(1), 146. https://doi.org/10.1186/s13059-020-02052-w

Zeng, J., de Vlaming, R., Wu, Y., Robinson, M. R., Lloyd-Jones, L. R., Yengo, L., Yap, C. X., Xue, A., Sidorenko, J., McRae, A. F., Powell, J. E., Montgomery, G. W., Metspalu, A., Esko, T., Gibson, G., Wray, N. R., Visscher, P. M., & Yang, J. (2018). Signatures of negative selection in the genetic architecture of human complex traits. *Nature Genetics, 50*(5), 746–753. https://doi.org/10.1038/s41588-018-0101-4

Zhu, B., Niu, H., Zhang, W., Wang, Z., Liang, Y., Guan, L., Guo, P., Chen, Y., Zhang, L., Guo, Y., Ni, H., Gao, X., Hao, X., Xu, L., & Li, J. (2017). Genome wide association study and genomic prediction for fatty acid composition in Chinese Simmental beef cattle using
high density SNP array. *BMC Genomics*, 18(1), 464. https://doi.org/10.1186/s12864-017-3847-7

Zhu, B., Zhu, M., Jiang, J., Niu, H., Wang, Y., Wu, Y., Xu, L., Chen, Y., Zhang, L., Gao, X., Gao, H., Liu, J., & Li, J. (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(5), e0154118. https://doi.org/10.1371/journal.pone.0154118

Zollner, S., & Pritchard, J. K. (2007). Overcoming the winner’s curse: Estimating penetrance parameters from case-control data. *American Journal of Human Genetics*, 80(4), 605–615. https://doi.org/10.1086/512821

Zondervan, K. T., & Cardon, L. R. (2004). The complex interplay among factors that influence allelic association. *Nature Reviews Genetics*, 5(2), 89–100. https://doi.org/10.1038/nrg1270

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Li, H., Wang, Z., Xu, L., Li, Q., Gao, H., Ma, H., Cai, W., Chen, Y., Gao, X., Zhang, L., Gao, H., Zhu, B., Xu, L., & Li, J. (2022). Genomic prediction of carcass traits using different haplotype block partitioning methods in beef cattle. *Evolutionary Applications*, 15, 2028–2042. https://doi.org/10.1111/eva.13491