Genetics
Selection
Evolution

RESEARCH ARTICLE

CrossMark

Predictive performance of genomic selection methods for carcass traits in Hanwoo beef cattle: impacts of the genetic architecture

Hossein Mehrban[1], Deuk Hwan Lee[2*], Mohammad Hossein Moradi[3], Chung IlCho[4], Masoumeh Naserkheil[5] and Noelia Ibáñez-Escriche[6]

Abstract

Background: Hanwoo beef is known for its marbled fat, tenderness, juiciness and characteristic flavor, as well as for its low cholesterol and high omega 3 fatty acid contents. As yet, there has been no comprehensive investigation to estimate genomic selection accuracy for carcass traits in Hanwoo cattle using dense markers. This study aimed at evaluating the accuracy of alternative statistical methods that differed in assumptions about the underlying genetic model for various carcass traits: backfat thickness (BT), carcass weight (CW), eye muscle area (EMA), and marbling score (MS).

Methods: Accuracies of direct genomic breeding values (DGV) for carcass traits were estimated by applying fivefold cross-validation to a dataset including 1183 animals and approximately 34,000 single nucleotide polymorphisms (SNPs).

Results: Accuracies of BayesC, Bayesian LASSO (BayesL) and genomic best linear unbiased prediction (GBLUP) methods were similar for BT, EMA and MS. However, for CW, DGV accuracy was 7% higher with BayesC than with BayesL and GBLUP. The increased accuracy of BayesC, compared to GBLUP and BayesL, was maintained for CW, regardless of the training sample size, but not for BT, EMA, and MS. Genome-wide association studies detected consistent large effects for SNPs on chromosomes 6 and 14 for CW.

Conclusions: The predictive performance of the models depended on the trait analyzed. For CW, the results showed a clear superiority of BayesC compared to GBLUP and BayesL. These findings indicate the importance of using a proper variable selection method for genomic selection of traits and also suggest that the genetic architecture that underlies CW differs from that of the other carcass traits analyzed. Thus, our study provides significant new insights into the carcass traits of Hanwoo cattle.

Background

Hanwoo (*Bos taurus coreanae*) is an indigenous cattle breed in Korea that has been intensively bred for meat during the last 30 years [1]. Until the 1980s, Hanwoo cattle were used extensively for farming, transportation and religious sacrifices [2] but they have now become popular for meat production owing to their rapid growth and high-quality meat. It is now one of the most economically important species in Korea. The extensive marbling of the Hanwoo beef is an important factor that influences the perception of meat quality in commercial beef production [3]. Hanwoo beef is known for its marbled fat, tenderness, juiciness and characteristic flavor. In addition, it has a lower cholesterol content and higher omega 3 fatty acid content, which makes it healthier than the meat from other bovine breeds [4]. In spite of its high price, i.e. almost three times that of imported beef meat from other breeds [5], Hanwoo beef is very popular both

*Correspondence: dhlee@hknu.ac.kr
[2] Department of Animal Life and Environment Science, Hankyong National University, Jungang-ro 327, Anseong-si, Gyeonggi-do 456-749, Korea
Full list of author information is available at the end of the article

BioMed Central

Mehrban *et al. Genet Sel Evol* (2017) 49:1

Page 2 of 13

among Korean consumers and abroad because of these invaluable traits [6].

The main aim of the Hanwoo beef industry is to increase both the quality (marbling, tenderness and flavor) and the quantity (carcass weight) of the meat. Estimated breeding values for backfat thickness (BT), carcass weight (CW), eye muscle area (EMA), and marbling score (MS) are commonly used as selection criteria in attempts to increase meat yield and quality, and subsequently to improve the income generated from steer feedlots and calf sales [7]. The recently developed genomic selection approach is beginning to revolutionize animal breeding. It refers to a genetic evaluation method that uses phenotypic data and genotypes of dense single nucleotide polymorphisms (SNPs) to estimate effects of SNPs from a training population and subsequently to predict the genetic values of selection candidates based on their genotypes [8]. It has been widely applied to dairy cattle breeding [9–11] and is now beginning to be used in other livestock species [12, 13]. Genomic predictions for beef cattle are attractive because many traits that affect the profitability of beef production, such as carcass traits, are difficult to select for because they are expensive to measure or are measured only on the relatives of breeding bulls [14]. Accurate genomic estimated breeding values would lead to greater genetic gain for these traits [15].

Accuracy of genomic prediction is key to the success of genomic selection [13]. Several analytical approaches have been proposed to predict genetic values based on genomic data, among which genomic (ridge regression) best linear unbiased prediction (GBLUP or RRBLUP), Bayesian shrinkage (e.g. BayesA) and variable selection models [e.g. BayesB, BayesCπ, BayesC and BayesL (LASSO)] have been widely used [13, 16]. The main differences between these models are their assumptions concerning the distributions of the effects of genetic markers. GBLUP (or equivalent RRBLUP procedures) models assume that all effects of SNPs are drawn from the same normal distribution and thus, that all SNPs have small effects [8]. The Bayesian approaches allow the variances of the SNP effects to differ from one another. However, Gianola et al. [17] argued that for BayesA and BayesB models there is a strong dependency on the prior distributions of the marker variance because, in this case, the posterior variance is estimated with only one marker, thus its posterior distribution has only one more degree of freedom than its prior distribution. BayesCπ, is less sensitive to the prior assumption of the marker variance compared with BayesA and BayesB models because all SNPs have a common variance and the proportion of SNPs with no effect (π) has a uniform prior distribution

that is estimated during the analysis [18]. In BayesC, π is considered to be a fixed value [19], which leads to more accurate detection of quantitative trait loci (QTL) than BayesCπ, especially for traits with a moderate to high heritability and when sufficient numbers of records are available [20]. However, one drawback of the Bayesian methods is the need for the definition of priors. The requirement of a prior for the parameter π is circumvented in the BayesL method, which requires less information [21, 22].

Several studies have compared the performance of statistical methods applied to genomic selection and reported that genomic evaluation is more accurate than conventional genetic evaluation, see for example in dairy cattle [23, 24], beef cattle [25–27], pigs [28], sheep [29] and chickens [13, 30]. However, to date the performance of genomic selection in Hanwoo cattle has not been investigated. In addition, genomic prediction methods may perform differently for different traits and, thus lead to results that may differ because the genetic architecture that underlies a trait varies with the trait considered [9, 18]. Several studies have shown that Bayesian approaches produce higher accuracies than linear models when traits are influenced by genes with large effects [16, 31–34].

The aim of our study was to evaluate methods for genomic prediction in Hanwoo cattle. Three different methods, GBLUP, BayesC and BayesL, which differed in assumptions about the genetic architecture of traits, were used to compare the accuracy of genomic predictions for the traits BT, CW, EMA and MS.

Methods

Phenotypic and pedigree data

Phenotypic data from 5218 purebred Hanwoo steers produced by 590 young bulls were collected by the Hanwoo Improvement Center of the National Agricultural Cooperative Federation (NACF) between 1996 and 2012 in South Korea during a progeny testing program. Pedigree data from 44,538 individuals were used in the animal model. The four carcass traits included in the analysis, BT, CW, EMA and MS, were recorded at about 24 months of age on samples collected 24 h postmortem between the 13th rib and the 1st lumbar vertebra, according to the Korean carcass grading procedure by the National Livestock Cooperatives Federation. MS was assessed using a categorical system of nine classes that range from 1 (no marbling) to 9 (abundant marbling). Because MS data were skewed, they were transformed by a natural logarithm to lnMS after adding 1 to all records. Table 1 summarizes the statistics used for each trait to estimate variance components.

Mehrban *et al. Genet Sel Evol* (2017) 49:1

Page 3 of 13

Table 1 Summary statistics for the phenotypic data used to estimate variance components

Trait (unit)	Number of animals in the pedigree	Number of animals with records	Mean (SE)	Min.	Max.	SD
BT (mm)	44,538	5218	8.60 (0.05)	1	35	3.74
CW (kg)	44,538	5217	341.01 (0.63)	158	518	45.26
EMA (cm^2)	44,538	5213	78.73 (0.13)	40	123	9.18
lnMS (Score)	44,538	3382	1.38 (0.01)	0.69	2.30	0.37

BT backfat thickness, *CW* carcass weight, *EMA* eye muscle area, *MS* marbling score

Genotypes

A total of 1679 animals were genotyped using the Illumina BovineSNP50 K (n = 959) and HD 777 K (n = 720) Beadchips (Illumina Inc., San Diego, CA, USA). Common SNPs between the 50 K and 777 K SNP chips were selected which resulted in 43,852 SNPs. All animals with more than 10% missing data (N = 68) and those with an inconsistency between pedigree and genomic relationships (N = 5) were excluded from further analyses. Phenotypic records were available for 1183 of the remaining 1606 animals that were genotyped (Table 2). To ensure overall quality of the samples and a consistent set of genotypes, quality control procedures were applied to the initial data [35]. SNPs were excluded from further analyses if their minor allele frequency (MAF) was lower than 0.01 (6679 SNPs) or if the percentage of calls (the proportion of SNP genotypes over all animals, calculated by the Illumina GenCall analysis software) was less than 0.98 (2677 SNPs). For the remaining SNPs, any outliers [that departed from the Hardy–Weinberg equilibrium (p < 10^{-6}) across all animals from one breed] were used to identify genotyping errors (302 SNPs). Missing genotypes were imputed using BEAGLE [36]. Finally, 34,194 SNPs remained for analyses.

Statistical analysis

Estimation of heritability

Heritability for each carcass trait (Table 1) was estimated using the restricted maximum likelihood method (REML) for animal models, using BLUPF90 (AIREMLF90) software [37]. The mixed model used was:

Table 2 Summary statistics for the phenotypic data used in the genomic analysis

Trait (unit)	Number of animals	Mean (SE)	Min.	Max.	SD
BT (mm)	1183	8.24 (0.10)	2	24	3.53
CW (kg)	1183	360.18 (1.16)	183	476	39.85
EMA (cm^2)	1183	82.99 (0.26)	55	121	8.78
lnMS (Score)	1183	1.34 (0.01)	0.69	2.30	0.34

BT backfat thickness, *CW* carcass weight, *EMA* eye muscle area, *MS* marbling score

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e},$$

where $\mathbf{y}$ is the vector of observations; $\mathbf{b}$ is the vector of fixed effects including slaughter date and batch effects as a contemporary group (369, 369, 368 and 176 levels for BT, CW, EMA and MS, respectively), and slaughter age (days from birth to slaughter) as a covariate; $\mathbf{u}$ is the vector of random animal effects and is assumed to follow a normal distribution $N\left(\mathbf{0}, \mathbf{A}\sigma_a^2\right)$, $\mathbf{A}$ and σ_a^2 are the numerator relationship matrix and polygenic variance, respectively; $\mathbf{e}$ is the vector of random residual effects and is assumed to follow a normal distribution $N\left(\mathbf{0}, \mathbf{I}\sigma_e^2\right)$, where $\mathbf{I}$ is an identity matrix including all animals with records and σ_e^2 is the error variance; and $\mathbf{X}$ and $\mathbf{Z}$ are design matrices that relate records to fixed effects and random animal effects, respectively.

Genomic prediction

Genomic predictions were performed for animals that had both genotype and phenotype records using three different models, i.e. GBLUP, BayesL [38] and BayesC [19]. GBLUP was applied using AIREMLF90 software [37] as follows:

$$\mathbf{y}_c = \mathbf{1}\mu + \mathbf{Zg} + \mathbf{e},$$

where $\mathbf{y}_c$ is a vector of the trait of interest, which was adjusted for fixed effects (slaughter date and batch effects as a contemporary group, and slaughter age as a covariate) based on the full dataset (see, Table 1); 1 is a vector of 1 s; μ is the overall mean; $\mathbf{Z}$ is the incidence matrix of direct genomic breeding values (DGV) and $\mathbf{g}$ is the vector of DGV and is assumed to follow a normal distribution $N\left(\mathbf{0}, \mathbf{G}\sigma_g^2\right)$, where $\mathbf{G}$ is the marker-based genomic relationship matrix as a genomic relationship matrix and σ_g^2 the genetic variance captured by the markers; $\mathbf{e}$ is a vector of random residual effects and is assumed to follow a normal distribution $N\left(\mathbf{0}, \mathbf{I}\sigma_\varepsilon^2\right)$, where $\mathbf{I}$ is an identity matrix; and σ_ε^2 is the residual variance.

The $\mathbf{G}$-matrix was built using the information from genome-wide dense SNPs [39] with the default options (except for a MAF of 0.01) in the preGSf90 program [40]. In the Bayesian framework, genomic analyses were

Mehrban *et al. Genet Sel Evol* (2017) 49:1

Page 4 of 13

performed using GS3 software [38]. The allelic substitution effect of each SNP was estimated using BayesL and BayesC, which were fitted with values in the covariate codes as 0, 2 (for homozygotes) and 1 (for heterozygotes) using the following model:

$$\mathbf{y}_c = 1\mu + \sum_{i=1}^{m} \mathbf{z}_i \alpha_i \delta_i + \boldsymbol{\varepsilon},$$

where $\mathbf{y}_c$ is a vector of corrected phenotypes as defined before, 1 is a vector of 1s; μ is the overall mean, m is the number of SNPs; $\mathbf{z}_i$ is the vector of genotype covariates for SNP$_i$, α_i is the allelic substitution effect of SNP$_i$, δ_i is an indicator variable for the presence (1) or absence (0) of the ith SNP in the model (for the BayesL method, δ_i is equal to 1 for all (i); $\boldsymbol{\varepsilon}$ is the vector of random residual effects assumed to follow a normal distribution $N(\mathbf{0}, \mathbf{I}\sigma_{\varepsilon}^2)$, where $\mathbf{I}$ is an identity matrix; and σ_{ε}^2 is the residual variance.

In the BayesL method, the prior distribution for α_i (with $\delta_i = 1$) follows a normal distribution $N(\mathbf{0}, \mathbf{I}\sigma_{\alpha}^2)$ and the prior distribution was as follows [38]:

$$\Pr\left(\alpha_i | \tau^2\right) = N\left(0, \tau_i^2\right),$$
$$\Pr\left(\tau_i^2\right) = \frac{\lambda^2}{2} \exp\left(-\lambda^2 \left|\tau_i^2\right|\right).$$

The prior distribution for σ_{α}^2 for all methods, was an inverted χ^2 distribution with two degrees of freedom and expectation was equal to $\sigma_a^2/(1-\pi)\sum_{i=1}^{m} 2p_iq_i$ as proposed by Habier et al. [18] where σ_a^2 is the estimated additive genetic variance using the animal model and p and q are the allelic frequencies at the ith SNP. In the BayesC method, the value of π is fixed. To identify the most suitable proportion of SNPs with no effect, the parameter π was considered to be equal to 0.999 and π values ranging from 0.91 to 0.99 in 0.02 increments (six values of π) were used. The residual variance was also assigned an inverted χ^2 distribution with two degrees of freedom and the expected value was equal to the residual variance as estimated using the animal model. The Markov chain Monte Carlo (MCMC) process was run for 550,000 cycles with 50,000 iterations as burn-in with a thinning interval of 50, so the effect of SNPs was estimated as a posterior mean of 10,000 samples.

The DGV for each animal in the validation set was estimated as the sum of the cross-product of animal genotype and the estimated SNP effect over all SNPs.

To confirm results of Bayesian analyses, a single-marker regression was run by using the Wombat software [41] with the following model:

$$\mathbf{y}_c = 1\mu + \mathbf{w}_i \mathbf{s}_i + \mathbf{Z}\mathbf{u} + \mathbf{e},$$

where $\mathbf{y}_c$ is a vector with adjusted phenotypes as defined before, 1 is a vector of 1s; μ is the overall mean; $\mathbf{w}_i$ is the vector of genotype covariates for SNP$_j$; $\mathbf{s}_i$ is the allelic substitution effect of the ith SNP; $\mathbf{u}$ is the vector of random animal effects and is assumed to follow a normal distribution $N(\mathbf{0}, \mathbf{A}\sigma_a^2)$, where $\mathbf{A}$ and σ_a^2 are the numerator relationship matrix and polygenic variance, respectively; $\mathbf{e}$ is the vector of random residual effects and is assumed to follow a normal distribution $N(\mathbf{0}, \mathbf{I}\sigma_e^2)$, where $\mathbf{I}$ is an identity matrix including all animals with records and σ_e^2 is the error variance; and $\mathbf{Z}$ is a design matrix that relate records to random animal effects.

To adjust for multiple testing, a Bonferroni-corrected threshold of 0.05/N (=1.46×10^{-6}) was used, where N is the number of SNPs used for the analyses.

Validation of models

The dataset was randomly split into five approximately equal subsets (fivefold cross-validation). Four subsets were used as training populations ($\approx$946) and the fifth subset as a validation sample ($\approx$237). The animals for the various subsets were selected randomly, except that paternal half-sibs were always placed in the same subset [42]. Cross-validation was replicated 10 times. Pedigree relationships within folds were on average equal to 0.038 and between fivefolds ranged from 0.023 to 0.031, with an average relationship of 0.026 for 10 replications. The predictive ability of DGV was determined by calculating the correlation between the DGV and the adjusted phenotypes for each of the five subsets. To estimate the prediction accuracy for each trait, predictive ability was divided by the square root of the heritability for that trait [43]. The accuracy for each replicate was obtained as the mean of the accuracies for the fivefold cross-validations of the ten replicates. The slope of the regression of the adjusted phenotypes on DGV was calculated as a measurement of the bias of the DGV in each method and trait. In addition, the mean square error (MSE) was predicted as the mean of the square differences between corrected phenotypes and DGV. In order to investigate the impact of the size of reference population on accuracy of DGV, analyses were also performed with training population sizes of 473 (50%) and 710 (75%) animals that were randomly sampled from the original training set. The validation population size was kept constant for all training sample sizes as in [44]. The means of accuracies and biases for different traits and methods were computed using the 10 replicates of the same cross-validation structure previously described.

Estimation of genomic heritability

In GBLUP, the genomic variance (σ_g^2) is estimated by REML. However, for the BayesC and BayesL methods, σ_g^2 is estimated by $2\sigma_{\alpha}^2(1-\pi)\sum_{i=1}^{m} p_iq_i$ [38], where σ_{α}^2 is the common effect marker variance, π is the proportion of

SNPs with no effect, p_i and q_i are the allelic frequencies at SNP i. Genomic heritability (h_g^2) was estimated according to the following formula [45]:

$$h_g^2 = h^2 \frac{\sigma_g^2}{\sigma_a^2},$$

where h^2 and σ_a^2 are the pedigree-based heritability and additive genetic variance, respectively.

Estimation of effective population size and expected accuracy

The past effective population size (N_e) for the tth generation ($t = (2c_t)^{-1}$), was estimated using the following model [46]:

$$E\left(r^2 - \frac{1}{n}\right) = \frac{1}{4N_e c + \alpha}$$

where r^2 is the pair-wise linkage disequilibrium, n is the number of animals sampled (1606 animals), c is the recombination rate (Morgan) defined for a particular physical distance and α is a correction for the occurrence of mutations ($\alpha = 2$) [47]. Due to the sensitivity of the estimated effective population size to the threshold that is set for MAF [46], we considered two different MAF thresholds, i.e. 0.1 and 0.2.

The expected accuracy of the genomic prediction ($r_{g\hat{g}}$) in our population was calculated using the formula derived by Daetwyler et al. [32], i.e. $r_{g\hat{g}} = \sqrt{\frac{N_P h^2}{N_P h^2 + M_e}}$. This formula depends on h^2 (heritability of the trait), N_P (number of animals in the training population) and M_e (the number of independent chromosome segments). M_e was calculated by using two different approximations: (1) $M_{e1} = \frac{2N_e L}{\ln(4N_e l)}$ [48] and (2) $M_{e2} = 2N_e L$ [49], where N_e is the effective population size, L is the genome length and l is the average chromosome length. Therefore, these two approximations of M_e lead to two different estimates of $r_{g\hat{g}}$.

Results and discussion

Estimation of heritability

The pedigree-based estimates of variance components for the carcass traits are in Table 3. Medium to high heritabilities were estimated for carcass traits in Hanwoo cattle. Estimated heritabilities for CW and EMA agreed with those previously reported in Hanwoo cattle by Lee et al. [7]. However, estimated heritabilities for BT and MS were higher (+9 and +11.3, respectively) than those in the study of Lee et al. [7]. In Japanese Black cattle, Onogi et al. [50] reported similar heritabilities for EMA (0.43) and MS (0.66) but a higher heritability for CW (0.56) than our study. In a study on the Angus breed, Saatchi et al. [25] reported higher heritabilities for CW and EMA

Table 3 Variance components (standard error) estimated using pedigree and phenotypic data

Trait (unit)	σ_a^2	σ_e^2	σ_p^2	h^2
BT (mm)	5.57 (0.62)	5.75 (0.49)	11.32 (0.26)	0.49 (0.05)
CW (kg)	315.28 (46.76)	699.95 (40.51)	1015.23 (22.26)	0.31 (0.04)
EMA (cm²)	26.75 (3.27)	35.33 (2.67)	62.08 (1.42)	0.43 (0.05)
lnMS (Score)	0.08 (0.01)	0.05 (0.008)	0.13 (0.004)	0.61 (0.06)

BT backfat thickness, *CW* carcass weight, *EMA* eye muscle area, *MS* marbling score

$\sigma_a^2, \sigma_e^2, \sigma_p^2, h^2$: additive genetic variance, error variance, phenotypic variance and heritability, respectively
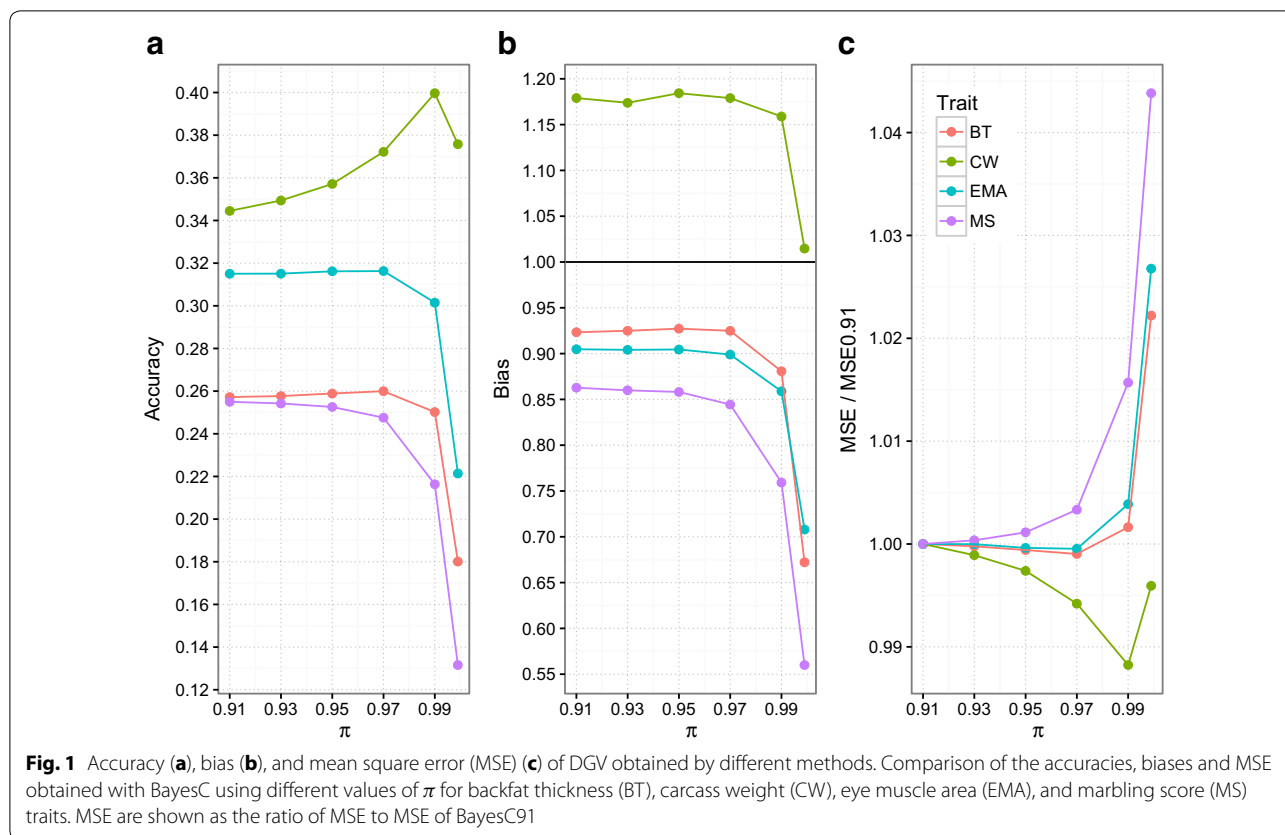
and lower heritabilities for BT and MS than those found here. Our estimated heritabilities for carcass traits were within the range of those obtained for multi-breed commercial beef cattle by Rolf et al. [16].

Estimation of effective population size

We used the average extent of linkage disequilibrium (LD) in the genome to estimate effective population sizes at various times in the past. Estimates of N_e were not influenced by the threshold set for MAF i.e. 0.10 or 0.20 [see Additional file 1: Figure S1]. Therefore, we used a threshold of 0.10 for MAF to estimate N_e. The results showed that N_e declined across generations to reach a value of 224 in the latest generation. The effective population size that was estimated here for Hanwoo cattle was not consistent with that reported by Lee et al. [51], who also found that it declined across generations but to 98, three generations ago. However, we used a sample size that was approximately 6 times larger than that used by Lee et al. [51] and also a much larger number of SNPs to estimate linkage disequilibrium (r^2). Moreover, Li and Kim [52] estimated an effective population size of 402, five generations ago, by using 547 Hanwoo bulls and a 50 K SNP chip, whereas our estimate for that generation was 298. With the exception of the N_e reported by Marquez et al. [53] ($N_e = 445$) for American Red Angus beef cattle and by Saatchi et al. [25] ($N_e = 654$) for American Angus beef cattle, most studies in beef and dairy cattle [54–58] have found smaller N_e than in the present study. According to Godard and Hayes [59], this implies that a larger reference population would be required for Hanwoo cattle than for the above-mentioned breeds [54–58] to obtain a similar accuracy in genomic prediction.

Comparison of models

The parameter π is a fixed value in the BayesC method [19]. We analyzed a range of π values from 0.91 to 0.999 to determine the most accurate π for the BayesC method for each trait. As shown in Fig. 1a, the realized accuracy for BT remained stable across a range of π values from
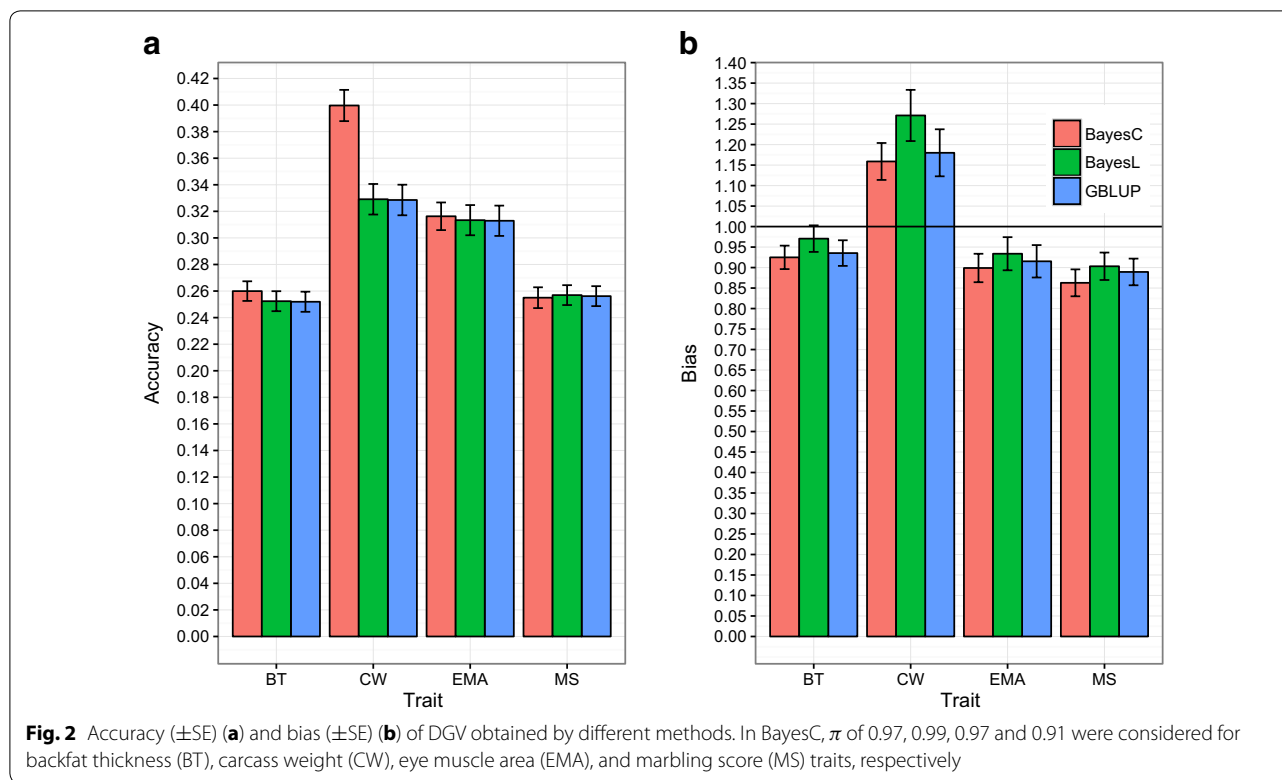
Mehrban *et al. Genet Sel Evol* (2017) 49:1

Page 6 of 13

Fig. 1 Accuracy (**a**), bias (**b**), and mean square error (MSE) (**c**) of DGV obtained by different methods. Comparison of the accuracies, biases and MSE obtained with BayesC using different values of π for backfat thickness (BT), carcass weight (CW), eye muscle area (EMA), and marbling score (MS) traits. MSE are shown as the ratio of MSE to MSE of BayesC91

0.91 to 0.97, and then decreased for π values above 0.97. Similar patterns were observed for EMA and MS, with accuracies decreasing for π values above 0.97 and 0.91, respectively. In contrast, the accuracy of CW improved as π increased to reach a peak for a π value of 0.99 and then declined dramatically. Overall, the values of π for which the BayesC model provided the highest accuracy were 0.97 (BayesC97), 0.99 (BayesC99), 0.97 (BayesC97) and 0.91 (BayesC91) for BT, CW, EMA and MS traits, respectively (Fig. 1a). The lowest bias was obtained with π values of 0.95 for BT, 0.999 for CW, 0.95 for EMA, and 0.91 for MS (Fig. 1b). Thus, for CW there was a conflict between accuracy and bias to determine the most suitable π value. The highest accuracy and lowest bias for CW were obtained for π values of 0.99 and 0.999, respectively. Nevertheless, González-Recio et al. [60] showed that the MSE is a more flexible criterion than correlation and bias for comparing models because it takes both prediction bias and variability into account. Due to the fact that MSE depends on the trait, we used the MSE ratio (ratio between MSE and MSE of BayesC91) to compare across traits and models. The lowest MSE ratio was achieved when π was set to 0.97, 0.99, 0.97, and 0.91 for BT, CW, EMA and MS, respectively (Fig. 1c).

A comparison of the accuracy and bias obtained for CW with the BayesC99, BayesL and GBLUP methods, revealed the superiority of the BayesC99 model (Fig. 2a); the accuracy of this model was higher than those of GBLUP (+0.071) and BayesL (+0.070) and the bias was lower than those of GBLUP (−0.02) and BayesL (−0.11) (Fig. 2b). For the other carcass traits (BT, EMA and MS), the accuracy and bias of BayesC99, BayesL and GBLUP methods were similar.

In terms of MSE, BayesC99 exhibited the best performance (the lowest MSE) for CW, while for the other traits, the differences in MSE between the methods were trivial [see Additional file 2: Table S1].

The predictive performance of the models depended on the trait analyzed. The three methods performed similarly for BT, EMA and MS traits, whereas for CW BayesC clearly outperformed GBLUP and BayesL. This indicates that the infinitesimal model holds for BT, EMA and MS but not completely for CW. In other words, BT, EMA and MS traits would be controlled by several genes, each with a small effect, whereas one or more individual genes would have a large effect on CW. These findings were confirmed by the single-marker method used for the GWAS analysis, which detected genome-wide significant
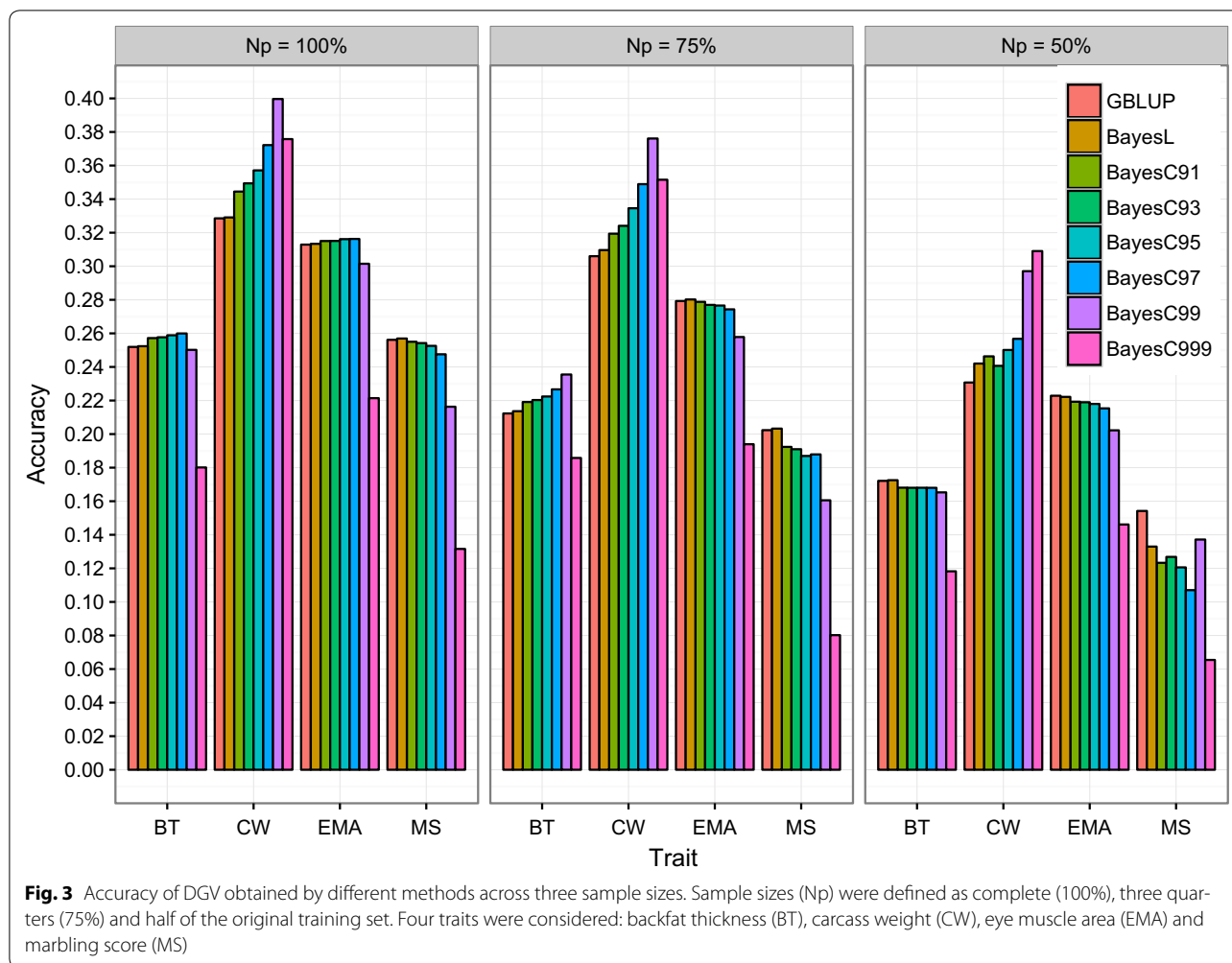
Mehrban *et al. Genet Sel Evol* (2017) 49:1

Page 7 of 13

Fig. 2 Accuracy (±SE) (**a**) and bias (±SE) (**b**) of DGV obtained by different methods. In BayesC, π of 0.97, 0.99, 0.97 and 0.91 were considered for backfat thickness (BT), carcass weight (CW), eye muscle area (EMA), and marbling score (MS) traits, respectively

SNPs on chromosomes 6 and 14 for CW but not for MS, BT and EMA [see Additional file 3: Figure S2]. However, our results could be quite sensitive to the size of the reference population. Gao et al. [61] showed that by increasing the number of animals in the reference population, the difference in accuracy between Bayesian and GBLUP approaches decreased. Therefore, the impact of the size of the training population on accuracy was also investigated. As shown in Fig. 3, the accuracy of prediction for the traits and methods studied decreased as the size of the training population decreased, in agreement with the literature [32, 44, 59]. Nevertheless, the superiority of BayesC compared to GBLUP and BayesL was maintained in terms of accuracy regardless of the size of the training sample for CW but not for BT, EMA, and MS, regardless of the π value (Fig. 3).

Wolc et al. [62] pointed out that mixture models (i.e. BayesB and BayesC) were clearly better than GBLUP for genomic prediction in the presence of QTL with a large effect, especially for small datasets and resulted in more accurate and persistent predictions. In our study, the accuracy of genomic prediction clearly differed between a Bayesian model (BayesC99) and GBLUP for CW with varying sizes of the training population as was also reported by [32].

Our results support a previous study on Hanwoo cattle by Lee et al. [7] that aimed at identifying major loci

associated with several carcass traits (BT, CW, EMA and MS). They demonstrated that six highly significant SNPs on chromosome 14 were associated with CW, but no significant SNPs were identified for the other carcass traits. Another GWAS on Japanese black beef cattle also detected three QTL that had a relatively large effect on CW [63]. Ogawa et al. [64] reported that MS is controlled by QTL that have only relatively small effects compared with the CW trait in Japanese black beef cattle. Other studies have also reported conflicting results. For example, Chen et al. [27] showed that GBLUP and the Bayesian methods were very similar in terms of accuracy for BT, CW, EMA and MS traits in Angus cattle and for CW, EMA and MS traits in Charolais cattle. They found that the BayesB95 ($\pi = 0.95$) model performed more accurately (3%) than GBLUP for BT in Charolais, whereas in contrast, Rolf et al. [16] found that the accuracy of BayesB95 ($\pi = 0.95$) was 3.4% lower than that of RRBLUP for the same trait in multi-breed commercial beef cattle. They showed that RRBLUP was more accurate than BayesB for BT, CW and MS, whereas, for EMA, the accuracy of DGV was the same using either method. Júnior et al. [65] obtained similar results for BT, CW, and EMA in terms of accuracy and MSE using RRBLUP, BayesC and BayesL in Nellore cattle. These observations may also support the argument that the genetic architecture of these traits may differ among breeds because of different population histories. Saatchi

Mehrban *et al. Genet Sel Evol* (2017) 49:1

Page 8 of 13

Fig. 3 Accuracy of DGV obtained by different methods across three sample sizes. Sample sizes (Np) were defined as complete (100%), three quarters (75%) and half of the original training set. Four traits were considered: backfat thickness (BT), carcass weight (CW), eye muscle area (EMA) and marbling score (MS)

et al. [66] showed that one reason that explains the differences in the QTL identified among different populations could be that the genetic architecture that underlies trait variation varies among breeds.

Comparison between the traits analyzed

In spite of their high heritabilities, prediction accuracies for BT and MS were lower than those for CW and EMA (Table 3; Figs. 1a, 2a), which is consistent with the results of Onogi et al. [50]. To investigate further the low prediction accuracy for BT and MS, genomic heritability (h_g^2) was estimated for each trait and with each method (Table 4). The proportion of genomic heritability to pedigree-based heritability (h_g^2/h^2) represents the proportion of genetic variance that was explained by the markers (σ_g^2/σ_a^2) [45]. Our results indicated that the estimated genomic variance (σ_g^2) was lower than the additive genetic variance σ_a^2 (Tables 3, 4) for all traits and with all methods except for CW using BayesC, which was slightly

larger. However, given the large standard error obtained for σ_g^2 (72.12) and σ_a^2 (46.76), the differences between σ_a^2 and σ_g^2 were not significant. Compared to CW and EMA, genomic heritabilities for BT and MS differed largely from pedigree-based heritabilities, regardless of the method (Table 4). With the GBLUP model, the proportion of genetic variance captured by SNPs for BT and MS was equal to 65 and 66%, respectively. In other words, for BT and MS, 35 and 34% of the genetic variance was not explained by SNPs, while for EMA and CW, only 15% and just 5% of the additive genetic variance was unexplained.

This finding may explain the lower prediction accuracies obtained for BT and MS compared with EMA and CW, in spite of their higher heritability. In addition, it was expected that the DGV for MS would be more accurate than those for BT because MS had a higher heritability (Table 3), possibly because MS is a categorical trait. Kizilkaya et al. [67] showed that the accuracy of DGV for an ordinal categorical trait was substantially lower than for a continuous trait

Mehrban *et al. Genet Sel Evol (2017) 49:1*

Page 9 of 13

Table 4 Genomic variance (σ_g^2), marker variance explained (σ_g^2/σ_a^2) and genomic heritability (h_g^2) by fully corrected phenotype and medium-density SNP

Trait (unit)	Method[a]	σ_g^2 (SE)[b]	σ_g^2/σ_a^2	$h_g^2 = h^2\frac{\sigma_g^2}{\sigma_a^2}$
BT (mm)	BayesC[2]	3.71 (0.75)	0.67	0.33
	BayesL	3.63 (0.75)	0.65	0.32
	GBLUP	3.62 (0.73)	0.65	0.32
CW (kg)	BayesC	330.73 (72.12)	1.05	0.33
	BayesL	299.73 (72.96)	0.95	0.30
	GBLUP	300.70 (69.013)	0.95	0.30
EMA (cm^2)	BayesC	23.19 (4.04)	0.87	0.37
	BayesL	23.00 (4.16)	0.86	0.37
	GBLUP	22.84 (4.14)	0.85	0.37
lnMS (Score)	BayesC	0.055 (0.009)	0.69	0.42
	BayesL	0.054 (0.009)	0.68	0.41
	GBLUP	0.053 (0.009)	0.66	0.40

BT backfat thickness, *CW* carcass weight, *EMA* eye muscle area, *MS* marbling score

[a] For BayesC, π values of 0.97, 0.99, 0.97 and 0.91 (the highest accuracy) were considered for BT, CW, EMA and MS, respectively

[b] SE in Bayesian methods were estimated as the standard deviation of the posterior distribution

under the same conditions of heritability, effective and training population sizes, and number of categories.

The low genomic heritabilities achieved for BT and MS indicate that more animals (with genotypes and phenotypes) are necessary to accurately estimate the effects of SNPs compared with CW and EMA. We also observed that the SNPs on the 50 K SNP chip could not capture all the genetic variability for those traits (BT and MS). Therefore, a high-density SNP chip could be used to adequately assess LD and potentially capture a larger proportion of the additive genetic variance than the medium-density chip (i.e. 50,000 SNPs). In order to investigate the performance of SNP density, 570,969 SNPs were imputed from the 50 K chip. Our findings indicate that the genomic variance σ_g^2 and (σ_g^2/σ_a^2) increased as the SNP density increased [see Additional file 4: Table S2]. The accuracy of DGV increased by 4% for BT and 12% for MS; however, for CW and EMA, the accuracy did not improve. Many studies using simulation and real data confirmed that the accuracy of genomic selection improves only slightly when a high-density SNP chip or whole-sequence data are used [34, 68–71].

In general, the realized accuracies of DGV for the four carcass traits, regardless of the method used, were low compared with results from other studies [16, 25, 50]. One of the main reasons for the lower accuracies observed in our study could be due to the small training population size ($N \approx 946$) and the large effective population size ($N_e = 224$) for the Hanwoo breed. Theoretical studies have shown that, to obtain the same accuracy, the number of animals needed in the reference population increases with increasing effective population size [32, 59]. Using the K-means method, Saatchi et al. [25] estimated DGV accuracies of 0.60, 0.47, 0.60 and 0.69 for BT, CW, EMA and MS, respectively, using a training population of approximately 2200 Angus beef cattle. Using a training population of about 2000 animals in multi-breed commercial beef cattle, Rolf et al. [16] observed that the highest accuracies of DGV for BT, CW, EMA and MS were equal to 0.51, 0.78, 0.60 and 0.76, respectively. Onogi et al. [50] reported a predicted ability (correlation between the DGV and the adjusted phenotypes) of 0.44, 0.42 and 0.39 for CW, EMA and MS, respectively. In our study, the genetic relationship between the validation and reference populations was close to zero. This is the most challenging scenario for genomic prediction because a large part of the accuracy of DGV results from genetic relationships captured by SNPs [72]. This could explain that our prediction accuracies were lower than those reported by Onogi et al. [50] for which the number of genotyped animals was larger and the effective population size was smaller [64] than in our study.

An alternative for improving prediction accuracy for Hanwoo cattle, with a deep pedigree, is to apply single-step GBLUP (ssGBLUP) [73, 74]. In this method, accuracy is increased by using information from the pedigree and SNPs simultaneously [73]. However, as we have shown, GBLUP cannot be the best method for genomic prediction in the presence of QTL with a large effect such as the CW trait in our study. Thus, an alternative to increase the prediction accuracy for CW in single-step evaluation could be to use genomic relationship matrices weighted by marker realized variance as suggested by [75, 76].

Comparison of realized and expected accuracy

As shown in Fig. 4, the observed accuracies were lower than the expected accuracies according to the formula derived by Daetwyler et al. [32] when the approximation for M_e (i.e. number of independent chromosome segments) was $M_{e1} = 2N_eL/\ln(4N_el)$ [48] but greater than the expected accuracy when M_e was $M_{e2} = 2N_eL$ [49]. Our results agree with those of Neves et al. [77] who reported that expected accuracies based on M_{e1} were higher than realized accuracies across traits; however, expected accuracies using M_{e2} were lower than realized accuracies in the case of within-family predictions.
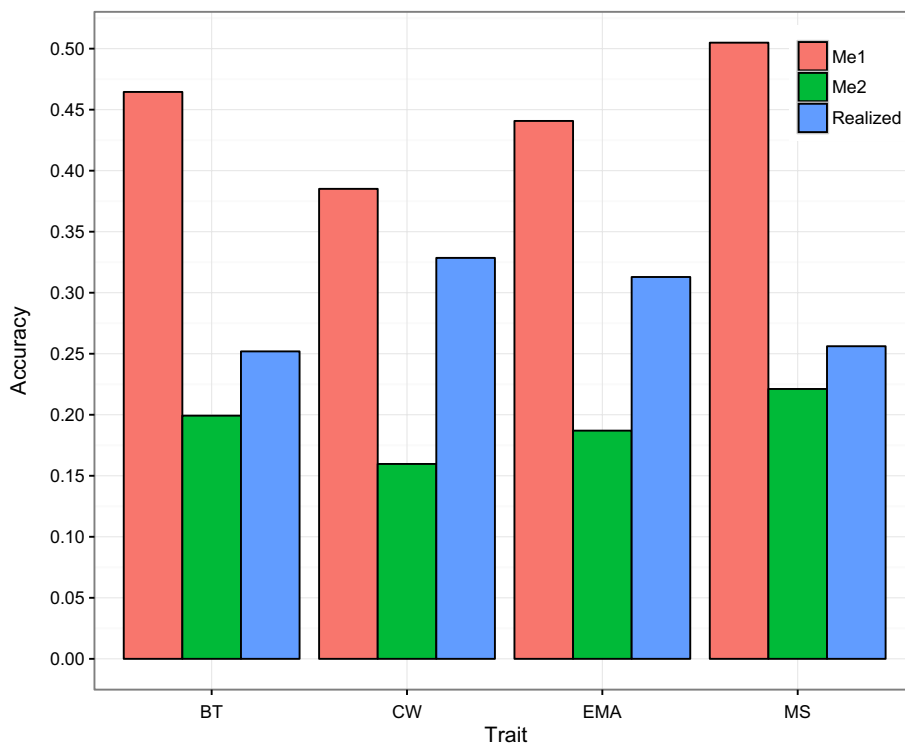
Mehrban *et al. Genet Sel Evol* (2017) 49:1

Page 10 of 13

Fig. 4 Realized and expected accuracy of genomic predictions with GBLUP. Expected accuracies were calculated according to Daetwyler et al. [32] using two different approximations for the number of independent chromosome segment M_e ($M_{e1} = 2N_eL/\ln(4N_eL)$ and $M_{e2} = 2N_eL$). The realized accuracies were averaged over 10 replicates for each trait [backfat thickness (BT), carcass weight (CW), eye muscle area (EMA) and marbling score (MS)]

Hayes et al. [49] pointed out that M_{e1} does not take into account that the small segments may still contain as many mutations in the QTL as the larger segments. Thus, Hayes et al. [49] recommended the use of $M_{e2} = 2N_eL$, which is a compromise between the number of segments ($4 N_eL$) and the number of segments weighted by length ($2N_eL/\log(4N_eL)$ per chromosome). However, M_{e2} is not an optimal approximation and based on our results as well as those of Neves et al. [76], it seems to underestimate the genomic prediction accuracy. However, the formula of Daetwyler et al. [32] assumes that all the genetic variance of the trait is explained by SNPs. Therefore, the formula is expected to overestimate prediction accuracy when SNPs cannot capture all the genetic variability. In our study, the genomic variance was smaller than the additive genetic variance (see Table 4), especially for BT and MS. Consequently, this could explain the differences between expected (M_{e1}) and realized accuracy for BT (0.21) and MS (0.25) and for EMA (0.13) and CW (0.06). This would indicate that when nearly all the total genetic variance is explained by the SNP

array, the realized accuracies of GBLUP are closer to the expected values based on M_{e1} than on M_{e2}.

Conclusions

The performance of the statistical methods used depended on the trait analyzed. The results showed a clear superiority of BayesC compared with GBLUP and BayesL for CW, whereas for the other traits all methods performed similarly. The prediction accuracy of DGV for CW using BayesC was around 7% higher than that obtained with the GBLUP and BayesL methods. This indicates the importance of using a proper variable selection method for genomic selection of traits. In addition, the results also suggest that the genetic architecture underlying CW may differ from that underlying the other carcass traits. This could be due to the fact that BT, EMA and MS seem to be controlled by several genes, each with a small effect, whereas for CW, there are probably several individual genes that each have a large effect. Overall, our results provide the first information for implementing genomic prediction in Hanwoo beef cattle.

Mehrban *et al. Genet Sel Evol* (2017) 49:1

Page 11 of 13

Additional files

Authors' contributions

DHL conceived and designed the study and contributed to the discussion of the results. HM conceived the study, analyzed the data and drafted the manuscript. MHM contributed to the discussion of the results and drafted the manuscript. ClC and MN participated in analyzing the data. NIE conceived the study, evaluated the experiments, contributed to the discussion of the results and edited the manuscript. All authors read and approved the final manuscript.

Author details

[1] Department of Animal Science, Shahrekord University, P.O. Box 115, Shahrekord 88186-34141, Iran. [2] Department of Animal Life and Environment Science, Hankyong National University, Jungang-ro 327, Anseong-si, Gyeonggi-do 456-749, Korea. [3] Department of Animal Science, Faculty of Agriculture and Natural Resources, Arak University, Arāk 38156-8-8349, Iran. [4] Hanwoo Improvement Center, National Agricultural Cooperative Federation, Haeun-ro 691, Unsan-myeon, Seosan-si, Chungnam-do 356-831, Korea. [5] Department of Animal Science, University College of Agriculture and Natural Resources, University of Tehran, P.O. Box 4111, Karaj 31587-11167, Iran. [6] The Roslin Institute, Royal (Dick) School of Veterinary Studies, University of Edinburgh, Roslin, UK.

Competing interests

The authors declare that they have no competing interests.

References

1. Choi JW, Choi BH, Lee SH, Lee SS, Kim HC, Yu D, et al. Whole-genome resequencing analysis of Hanwoo and Yanbian cattle to identify genome-wide SNPs and signatures of selection. Mol Cells. 2015;38:466–73.
2. Lee SH, Choi BH, Cho SH, Lim D, Choi TJ, Park BH, et al. Genome-wide association study identifies three loci for intramuscular fat in Hanwoo (Korean cattle). Livest Sci. 2014;165:27–32.
3. Choi Y, Davis ME, Chung H. Effects of genetic variants in the promoter region of the bovine adiponectin (ADIPOQ) gene on marbling of Hanwoo beef cattle. Meat Sci. 2015;105:57–62.
4. Jo C, Cho SH, Chang J, Nam KC. Keys to production and processing of Hanwoo beef: a perspective of tradition and science. Anim Front. 2012;2:32–8.
5. Korea Rural Economic Institute (KREI). Outlook and Agricultural Statistics Information System(OASIS). http://oasis.krei.re.kr (2015). Accessed 20 Oct 2015.
6. Lee SH, Park BH, Sharma A, Dang CG, Lee SS, Choi TJ, et al. Hanwoo cattle: origin, domestication, breeding strategies and genomic selection. J Anim Sci Technol. 2014;56:2.
7. Lee SH, Choi BH, Lim D, Gondro C, Cho YM, Dang CG, et al. Genome-wide association study identifies major loci for carcass weight on BTA14 in Hanwoo (Korean cattle). PLoS One. 2013;8:e74677.
8. Meuwissen THE, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. Genetics. 2001;157:1819–29.
9. Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME. Invited review: genomic selection in dairy cattle: progress and challenges. J Dairy Sci. 2009;92:433–43.
10. VanRaden PM, Sullivan PG. International genomic evaluation methods for dairy cattle. Genet Sel Evol. 2010;42:7.
11. 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 BayesCpi methods for genomic selection in French Holstein and Montbeliarde breeds. J Dairy Sci. 2013;96:575–91.
12. Duchemin SI, Colombani C, Legarra A, Baloche G, Larroque H, Astruc JM, et al. Genomic selection in the French Lacaune dairy sheep breed. J Dairy Sci. 2012;95:2723–33.
13. Liu T, Qu H, Luo C, Shu D, Wang J, Lund MS, et al. Accuracy of genomic prediction for growth and carcass traits in Chinese triple-yellow chickens. BMC Genet. 2014;15:110.
14. Boerner V, Johnston DJ, Tier B. Accuracies of genomically estimated breeding values from pure-breed and across-breed predictions in Australian beef cattle. Genet Sel Evol. 2014;46:61.
15. Bolormaa S, Pryce JE, Kemper K, Savin K, Hayes BJ, Barendse W, et al. Accuracy of prediction of genomic breeding values for residual feed intake and carcass and meat quality traits in *Bos taurus*, *Bos indicus*, and composite beef cattle. J Anim Sci. 2013;91:3088–104.
16. Rolf MM, Garrick DJ, Fountain T, Ramey HR, Weaber RL, Decker JE, et al. Comparison of Bayesian models to estimate direct genomic values in multi-breed commercial beef cattle. Genet Sel Evol. 2015;47:23.
17. Gianola D, de los Campos G, Hill WG, Manfredi E, Fernando R. Additive genetic variability and the Bayesian alphabet. Genetics. 2009;183:347–63.
18. Habier D, Fernando RL, Kizilkaya K, Garrick DJ. Extension of the bayesian alphabet for genomic selection. BMC Bioinformatics. 2011;12:186.
19. Fernando RL, Garrick D. Bayesian methods applied to GWAS. Methods Mol Biol. 2013;1019:237–74.
20. van den Berg I, Fritz S, Boichard D. QTL fine mapping with Bayes C(pi): a simulation study. Genet Sel Evol. 2013;45:19.
21. de los Campos G, 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.
22. Legarra A, Robert-Granie C, Croiseau P, Guillaume F, Fritz S. Improved Lasso for genomic selection. Genet Res. 2011;93:77–87.
23. Su G, Guldbrandtsen B, Gregersen VR, Lund MS. Preliminary investigation on reliability of genomic estimated breeding values in the Danish Holstein population. J Dairy Sci. 2010;93:1175–83.
24. Lund MS, Roos AP, Vries AG, Druet T, Ducrocq V, Fritz S, et al. A common reference population from four European Holstein populations increases reliability of genomic predictions. Genet Sel Evol. 2011;43:43.
25. Saatchi M, McClure MC, McKay SD, Rolf MM, Kim J, Decker JE, et al. Accuracies of genomic breeding values in American Angus beef cattle using K-means clustering for cross-validation. Genet Sel Evol. 2011;43:40.
26. Neves HH, Carvalheiro R, O'Brien AM, Utsunomiya YT, do Carmo AS, Schenkel FS, et al. Accuracy of genomic predictions in *Bos indicus* (Nellore) cattle. Genet Sel Evol. 2014;46:17.

Mehrban *et al. Genet Sel Evol* (2017) 49:1

Page 12 of 13

27. Chen L, Vinsky M, Li C. Accuracy of predicting genomic breeding values for carcass merit traits in Angus and Charolais beef cattle. Anim Genet. 2015;46:55–9.

28. Tribout T, Larzul C, Phocas F. Efficiency of genomic selection in a purebred pig male line. J Anim Sci. 2012;90:4164–76.

29. Daetwyler HD, Swan AA, van der Werf JH, Hayes BJ. Accuracy of pedigree and genomic predictions of carcass and novel meat quality traits in multi-breed sheep data assessed by cross-validation. Genet Sel Evol. 2012;44:33.

30. Wolc A, Stricker C, Arango J, Settar P, Fulton JE, O'Sullivan NP, et al. Breeding value prediction for production traits in layer chickens using pedigree or genomic relationships in a reduced animal model. Genet Sel Evol. 2011;43:5.

31. Hayes BJ, Pryce J, Chamberlain AJ, Bowman PJ, Goddard ME. Genetic architecture of complex traits and accuracy of genomic prediction: coat colour, milk-fat percentage, and type in Holstein cattle as contrasting model traits. PLoS Genet. 2010;6:e1001139.

32. Daetwyler HD, Pong-Wong R, Villanueva B, Woolliams JA. The impact of genetic architecture on genome-wide evaluation methods. Genetics. 2010;185:1021–31.

33. Zhang Z, Liu J, Ding X, Bijma P, de Koning DJ, Zhang Q. Best linear unbiased prediction of genomic breeding values using a trait specific marker derived relationship matrix. PLoS One. 2010;5:e12648.

34. Clark SA, Hickey JM, van der Werf JH. Different models of genetic variation and their effect on genomic evaluation. Genet Sel Evol. 2011;43:18.

35. Moradi MH, Nejati-Javaremi A, Moradi-Shahrbabak M, Dodds KG, McEwan JC. Genomic scan of selective sweeps in thin and fat tail sheep breeds for identifying of candidate regions associated with fat deposition. BMC Genet. 2012;13:10.

36. Browning SR, Browning BL. Rapid and accurate haplotype phasing and missing data inference for whole genome association studies by use of localized haplotype clustering. Am J Hum Genet. 2007;81:1084–97.

37. Misztal I, Tsuruta S, Strabel T, Auvray B, Druet T, Lee DH. BLUPF90 and related programs (BGF90). In: Proceedings of the 7th world congress on genetics applied to livestock production: 19–23 August 2002; Montpellier. CD-ROM Communication No 28-27. 2002.

38. Legarra A RA, Filangi O. GS3 genomic Selection–Gibbs Sampling–Gauss Seidel (and BayesCπ) (2011). https://qgsp.jouy.inra.fr/index.php?option=com_content&view=article&id=60&Itemid=67.

39. VanRaden PM. Efficient methods to compute genomic predictions. J Dairy Sci. 2008;91:4414–23.

40. Aguilar I, Misztal I, Tsuruta S, Legarra A, Wang H. PREGSF90–POSTGSF90: computational tools for the implementation of single-step genomic selection and genome-wide association with ungenotyped individuals in BLUPF90 programs. In: Proceedings of the 10th world congress on genetics applied to livestock production: 18–22 August 2014; Vancouver. 2014.

41. Meyer K, Tier B. SNP Snappy: a strategy for fast genome-wide association studies fitting a full mixed model. Genetics. 2012;190:275–7.

42. Khansefid M, Pryce JE, Bolormaa S, Miller SP, Wang Z, Li C, et al. Estimation of genomic breeding values for residual feed intake in a multibreed cattle population. J Anim Sci. 2014;92:3270–83.

43. Legarra A, Robert Granié C, Manfredi E, Elsen JM. Performance of genomic selection in mice. Genetics. 2008;180:611–8.

44. Abdollahi-Arpanahi R, Morota G, Valente BD, Kranis A, Rosa GJ, Gianola D. Assessment of bagging GBLUP for whole-genome prediction of broiler chicken traits. J Anim Breed Genet. 2015;132:218–28.

45. de los Campos G, Sorensen D, Gianola D. Genomic heritability: what is it? PLoS Genet. 2015;11:e1005048.

46. Corbin LJ, Liu AY, Bishop SC, Woolliams JA. Estimation of historical effective population size using linkage disequilibria with marker data. J Anim Breed Genet. 2012;129:257–70.

47. Tenesa A, Navarro P, Hayes BJ, Duffy DL, Clarke GM, Goddard ME, et al. Recent human effective population size estimated from linkage disequilibrium. Genome Res. 2007;17:520–6.

48. Rabier CE, Barre P, Asp T, Charmet G, Mangin B. On the accuracy of genomic selection. PLoS One. 2016;11:e0156086.

49. Hayes BJ, Visscher PM, Goddard ME. Increased accuracy of artificial selection by using the realized relationship matrix. Genet Res. 2009;91:47–60.

50. Onogi A, Ogino A, Komatsu T, Shoji N, Simizu K, Kurogi K, et al. Genomic prediction in Japanese Black cattle: application of a single-step approach to beef cattle. J Anim Sci. 2014;92:1931–8.

51. Lee SH, Cho YM, Lim D, Kim HC, Choi BH, Park HS, et al. Linkage disequilibrium and effective population size in Hanwoo Korean cattle. Asian Aust J Anim Sci. 2011;24:1660–5.

52. Li Y, Kim JJ. Effective population size and signatures of selection using bovine 50 K SNP chips in Korean native cattle (Hanwoo). Evol Bioinform Online. 2015;11:143–53.

53. Marquez GC, Speidel SE, Enns RM, Garrick DJ. Genetic diversity and population structure of American Red Angus cattle. J Anim Sci. 2010;88:59–68.

54. Cleveland MA, Blackburn HD, Enns RM, Garrick DJ. Changes in inbreeding of US Herefords during the twentieth century. J Anim Sci. 2005;83:992–1001.

55. Sorensen AC, Sorensen MK, Berg P. Inbreeding in Danish dairy cattle breeds. J Dairy Sci. 2005;88:1865–72.

56. Kim ES, Kirkpatrick BW. Linkage disequilibrium in the North American Holstein population. Anim Genet. 2009;40:279–88.

57. de Roos AP, Hayes BJ, Spelman RJ, Goddard ME. Linkage disequilibrium and persistence of phase in Holstein–Friesian, Jersey and Angus cattle. Genetics. 2008;179:1503–12.

58. Ni GY, Zhang Z, Jiang L, Ma PP, Zhang Q, Ding XD. Chinese Holstein cattle effective population size estimated from whole genome linkage disequilibrium. Yi Chuan. 2012;34:50–8.

59. Goddard ME, Hayes BJ. Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nat Rev Genet. 2009;10:381–91.

60. González-Recio O, Rosa GJ, Gianola D. Machine learning methods and predictive ability metrics for genome-wide prediction of complex traits. Livest Sci. 2014;166:217–31.

61. Gao N, Li J, He J, Xiao G, Luo Y, Zhang H, et al. Improving accuracy of genomic prediction by genetic architecture based priors in a Bayesian model. BMC Genet. 2015;16:20.

62. Wolc A, Arango J, Settar P, Fulton JE, O'Sullivan NP, Dekkers JC, et al. Mixture models detect large effect QTL better than GBLUP and result in more accurate and persistent predictions. J Anim Sci Biotechnol. 2016;7:7.

63. Nishimura S, Watanabe T, Mizoshita K, Tatsuda K, Fujita T, Watanabe N, et al. Genome-wide association study identified three major QTL for carcass weight including the PLAG1-CHCHD7 QTN for stature in Japanese Black cattle. BMC Genet. 2012;13:40.

64. Ogawa S, Matsuda H, Taniguchi Y, Watanabe T, Nishimura S, Sugimoto Y, et al. Effects of single nucleotide polymorphism marker density on degree of genetic variance explained and genomic evaluation for carcass traits in Japanese Black beef cattle. BMC Genet. 2014;15:15.

65. Fernandez Júnior GA, Rosa GJ, Valente BD, Carvalheiro R, Baldi F, Garcia DA, et al. Genomic prediction of breeding values for carcass traits in Nellore cattle. Genet Sel Evol. 2016;48:7.

66. Saatchi M, Schnabel RD, Taylor JF, Garrick DJ. Large-effect pleiotropic or closely linked QTL segregate within and across ten US cattle breeds. BMC Genomics. 2014;15:442.

67. Kizilkaya K, Fernando RL, Garrick DJ. Reduction in accuracy of genomic prediction for ordered categorical data compared to continuous observations. Genet Sel Evol. 2014;46:37.

68. Jensen J, Su G, Madsen P. Partitioning additive genetic variance into genomic and remaining polygenic components for complex traits in dairy cattle. BMC Genet. 2012;13:44.

69. Khatkar MS, Moser G, Hayes BJ, Raadsma HW. Strategies and utility of imputed SNP genotypes for genomic analysis in dairy cattle. BMC Genomics. 2012;13:538.

70. Druet T, Macleod IM, Hayes BJ. Toward genomic prediction from whole-genome sequence data: impact of sequencing design on genotype imputation and accuracy of predictions. Heredity. 2014;112:39–47.

71. Heidaritabar M, Calus MP, Megens HJ, Vereijken A, Groenen MA, Bastiaansen JW. Accuracy of genomic prediction using imputed whole genome sequence data in white layers. J Anim Breed Genet. 2016;133:167–79.

Mehrban *et al. Genet Sel Evol* (2017) 49:1

Page 13 of 13

72. Habier D, Fernando RL, Dekkers JC. The impact of genetic relationship information on genome-assisted breeding values. Genetics. 2007;177:2389–97.
73. Legarra A, Aguilar I, Misztal I. A relationship matrix including full pedigree and genomic information. J Dairy Sci. 2009;92:4656–63.
74. Misztal I, Aggrey SE, Muir WM. Experiences with a single-step genome evaluation. Poult Sci. 2013;92:2530–4.
75. Wang H, Misztal I, Aguilar I, Legarra A, Muir WM. Genome-wide association mapping including phenotypes from relatives without genotypes. Genet Res. 2012;94:73–83.
76. Tiezzi F, Maltecca C. Accounting for trait architecture in genomic predictions of US Holstein cattle using a weighted realized relationship matrix. Genet Sel Evol. 2015;47:24.
77. Neves HH, Carvalheiro R, Queiroz SA. A comparison of statistical methods for genomic selection in a mice population. BMC Genet. 2012;13:100.