J Anim Reprod Biotechnol 2023;38:32-37
pISSN: 2671-4639 • eISSN: 2671-4663
https://doi.org/10.12750/JARB.38.1.32

Original Article

JARB
Journal of Animal Reproduction and Biotechnology
Check for updates

Genetic evaluation and accuracy analysis of commercial Hanwoo population using genomic data

Gwang Hyeon Lee[1], Yeon Hwa Lee[1] and Hong Sik Kong[1,2,*]

[1]Department of Applied Biotechnology, The Graduate School of Hankyong National University, Anseong 17579, Korea
[2]Gyeonggi Regional Research Center, Hankyong National University, Anseong 17579, Korea

Received February 18, 2023
Revised March 10, 2023
Accepted March 14, 2023

***Correspondence**
Hong Sik Kong
E-mail: kebinkhs@hknu.ac.kr

Author's Position and Orcid no.
Lee GH, Graduate student,
https://orcid.org/0000-0002-6598-8430
Lee YH, Graduate student,
https://orcid.org/0000-0002-6798-7837
Kong HS, Professor,
https://orcid.org/0000-0003-1144-016X

ABSTRACT This study has evaluated the genomic estimated breeding value (GEBV) of the commercial Hanwoo population using the genomic best linear unbiased prediction (GBLUP) method and genomic information. Furthermore, it analyzed the accuracy and realized accuracy of the GEBV. 1,740 heads of the Hanwoo population which were analyzed using a single nucleotide polymorphism (SNP) Chip has selected as the test population. For carcass weight (CWT), eye muscle area (EMA), back fat thickness (BFT), and marbling score (MS), the mean GEBVs estimated using the GBLUP method were 3.819, 0.740, -0.248, and 0.041, respectively and the accuracy of each trait was 0.743, 0.728, 0.737, and 0.765, respectively. The accuracy of the breeding value was affected by heritability. The accuracy was estimated to be low in EMA with low heritability and high in MS with high heritability. Realized accuracy values of 0.522, 0.404, 0.444, and 0.539 for CWT, EMA, BFT, and MS, respectively, showing the same pattern as the accuracy value. The results of this study suggest that the breeding value of each individual can be estimated with higher accuracy by estimating the GEBV using the genomic information of 18,499 reference populations. If this method is used and applied to individual selection in a commercial Hanwoo population, more precise and economical individual selection is possible. In addition, continuous verification of the GBLUP model and establishment of a reference population suitable for commercial Hanwoo populations in Korea will enable a more accurate evaluation of individuals.

Keywords: breeding value, genetic evaluation, Hanwoo, single nucleotide polymorphism

INTRODUCTION

Traditionally, selection in the breeding field has been based on the breeding value estimated through the best linear unbiased prediction (BLUP) method using the pedigree and phenotypic information of the individual. (Henderson, 1984; Meuwissen et al., 2001). However, because the BLUP method depends on the pedigree and phenotypic information of the individual, the accuracy of the estimated breeding value is lowered when an error occurs, and the same breeding value is estimated when the parental information of the individual is the same (Heffner et al., 2009). Meuwissen presented the genomic selection (GS) theory and reported that the genetic ability of an individual can be estimated more accurately using their genetic information than the BLUP method using their pedigree information (Meuwissen et al., 2001). GS is an estimation of the genetic ability of an individual using variation markers found in the whole genome, and it is widely used in the breeding field (Meuwissen et al., 2001; Gao et al., 2018). The statistical method for GS is the genomic best linear unbiased prediction (GBLUP) method based on the infinitesimal model, and the breeding value is estimated using the genomic relationship matrix (GRM) instead of the Numerator Relationship Matrix (Garrick, 2007; VanRaden, 2007; Zhang, 2007). The GBLUP method utilizes single nucleotide polymorphism (SNP) information uniformly spread throughout the whole genome using SNP chips and estimates the SNP effect using individual SNP and phenotype information (Meuwissen et al., 2001). Using the estimated SNP effect value, the breeding value can be estimated based on the SNP information of individuals without phenotypic information. Because the GBLUP method uses genetic information analyzed in an individual, it can explain the effect of Mendelian sampling that occurs in an individual. Moreover, because the GBLUP method can construct a GRM, it is possible to accurately estimate the breeding value of an individual, even if the pedigree is the same (Visscher et al., 2006; Lee, 2012). If the GBLUP method is applied to farms to evaluate individuals, genetic evaluation with higher accuracy will be possible compared with the estimation of breeding values based on pedigrees. In addition, due to the large-scale breeding and specialization of Hanwoo cow breeding farms, interest in improving the Hanwoo population for the selection of cows has increased, and breeding values with higher reliability are needed (Son, 2021). Therefore, this study estimated the genomic estimated breeding value (GEBV) of the commercial Hanwoo population using the GBLUP method and genomic information. Furthermore, this study analyzed the accuracy and realized the accuracy of the estimated GEBV.

MATERIALS AND METHODS

Data of analysis

The test population consisted of 1,740 heads commercial Hanwoo raised in Gyeonggi, Gangwon, Chungbuk, and Gyeongnam regions. DNA of the Hanwoo was extracted using their tail hair root samples and was subjected to SNP Chip analysis. Information on individual Hanwoo was collected from the Hanwoo Improvement Main Center, Korea Animal Improvement Association, Animal Products Traceability, and Korea Institute for Animal Products Quality Evaluation, and data such as pedigree, birth year, sex, region of birth, farm number, and carcass characteristics were collected.

The construction of the reference population is essential for estimating the breeding value using the GBLUP method. Therefore, the data of 18,499 heads of the Hanwoo population that were slaughtered from the whole country was used with both phenotype and genome information.

SNP genotype information using the GBLUP method was analyzed using the DNA of a commercial Hanwoo population of 1,740 heads and the Hanwoo 50K SNP Analysis BeadChip (Illumina, San Diego, CA, USA) to analyze the genotype, and the genotype information of 52,195 SNPs was collected.

Genomic information quality control (QC)

To increase the accuracy of the analysis, QC was performed on the SNP information. GenomeStudio 2.0 software (Illumina Inc, 2016) was used to convert the obtained genotype information to Plink version 1.9 (Purcell et al., 2007; Chang et al., 2015) analysis format. Ped and map files were created through the PLINK formatting process using the Perl programming language. The QC criteria were: sample call rate < 90%, minor allele frequency < 1%, SNP call rate < 90%, and Hardy-Weinberg Equilibrium p-value $< 1 \times 10^{-7}$. The final dataset consisted of the genomic information of 45,548 SNPs.

Statistical analysis

The genetic parameters for the evaluation of the genetic ability of the individual were estimated using the REMLF90 program (Misztal et al., 2018), and the breeding values were estimated using the BLUPF90 program (Misztal et al., 2018). To estimate the genetic parameters of the four economic traits [carcass weight (CWT), eye muscle area (EMA), back fat thickness (BFT), and marbling score (MS)], A-matrix and G-matrix were constructed and used, and a single trait analysis was performed for each trait. Breeding values were estimated after determining the genetic and environmental variance for each economic trait. The mixed linear model for estimating the dielectric parameters of the additive genetic effect is as follows (Henderson, 1975):

$$Y_P = X\beta + Zu + e$$

$$\mathrm{Var}\left(\frac{u}{e}\right) = \begin{pmatrix} G\sigma_\alpha^2 & 0 \\ 0 & I\sigma_e^2 \end{pmatrix}$$

where Y_P is the observed phenotype for carcass traits, X is the coefficient matrix of the fixed effects, Z is the random effect vector for individuals, β is the estimated vector for the fixed effects, u is the additive genetic effect, and e is the random error vector (Henderson and Quaas, 1976). G is the genomic relationship matrix between individuals, I is the identity matrix or unit matrix, σ_{α}^{2} is the additive genetic variance, and σ_{α}^{2} is the random environmental variance. The genetic parameters and breeding values for each trait were estimated using the above model.

The heritability was calculated using the following formula using the estimated genetic parameters.

$$\text{heritability } (h^2) = \left(\frac{\sigma_{\alpha}^{2}}{\sigma_{\alpha}^{2} + \sigma_{e}^{2}}\right)$$

where h^2 is heritability, σ_{α}^{2} is the genetic variance, and σ_{α}^{2} is the environmental variance. Heritability can be calculated as the ratio of genetic variance to phenotypic variance, which is the sum of the genetic and environmental variances.

Accuracy analysis

The accuracy of the breeding values estimated using the GLBUP method was estimated. The accuracy was estimated using the prediction error variance of the breeding value estimated for each individual among the solution values generated using the BLUPF90 program (Misztal et al., 2018) analysis and the additive genetic variance estimated using the REMLF90 program (Misztal et al., 2018) analysis. The accuracy was calculated as follows:

$$\text{Accuracy } = \sqrt{1 - \left(\frac{PEV}{\sigma_{\alpha}^{2}}\right)}$$

where accuracy is the accuracy of the estimated breeding value, *PEV* is the prediction error variance of the estimated breeding value, and σ_{α}^{2} is the additive genetic variance.

Realized accuracy was analyzed by determining the correlation coefficient between carcass characteristics of the test population and GEBV using the GBLUP method, and the proximity between carcass characteristics and GEBV was confirmed.

RESULTS AND DISCUSSION

Basic statistical analysis of the reference population

For the reference population for estimating the GEBV using the GBLUP method, the data of 18,499 heads of the Hanwoo population that were slaughtered from the whole country was used. The basic statistics were analyzed in the order of CWT, EMA, BFT, and MS, with mean and standard deviation (SD) of 442.33±50.99 kg, 96.53±12.31 cm^2, 14.26±4.93 mm, and 5.96±1.87 score, respectively (Table 1).

The results of this study were compared with the basic statistics of the GBLUP reference population of previous studies; Byun et al. (2021) constituted the reference population with Hanwoo steers slaughtered from the whole country, and the mean±SD values of carcass characteristics were reported to be 442.33±50.99 kg, 96.53±12.31 cm^2, 14.26±4.93 mm, and 5.96±1.87 score for CWT, EMA, BFT, and MS, respectively. Kim et al. (2021) reported the basic statistics of CWT, EMA, BFT, and MS of the collected reference population for GLBUP analysis as 439.9 ± 51.3 kg, 94.8 ± 12.0 cm^2, 14.2 ± 4.9 mm, and 6.0 ± 1.9 scores, respectively. Jang reported the values 439±879.78 kg, 96.15±29.21 cm^2, 14.23±13.36 mm, and 5.99±1.98 scores for CWT, EMA, BFT, and MS, respectively, using the information on Hanwoo by the National Institute of Animal Science Jang et al. (2022). When comparing the previous studies and the results of this study, basic statistical values were similar, and the reference population used in this study was considered suitable for evaluating the Commercial Hanwoo Population.

Table 1. Basic statistics of reference population according to the GBLUP method

Analysis method	Reference population size	Trait	Maximum	Minimum	Mean	SD
GBLUP method	18,499	CWT (kg)	692	159	442.33	50.99
		EMA (cm^2)	160	20	96.53	12.31
		BFT (cm)	47	1	14.26	4.93
		MSC (score)	9	1	5.96	1.87

GBLUP, genomic best linear unbiased prediction; CWT, carcass weight; EMA, eye muscle area; BFT, back fat thickness; MS, marbling score; SD, standard deviation.

The normality test results for the reference population's phenotypic data are shown in Fig. 1. The quantile-quantile plots for CWT, EMA, BFT, and MS showed a normal distribution pattern, and we included information on outliers in the analysis as they represent diversity among the individuals without deleting them.

Estimation of genetic parameter

A G-matrix was constructed to estimate genetic parameters using the GBLUP method. Genetic variance, environmental variance, and heritability were estimated for genetic parameters and the results are listed in Table 2. For CWT, EMA, BFT, and MS, the genetic variance was estimated to be 866.3, 48.550, 9.012, and 1.515, respectively, and the environmental variance was estimated to be 1346.0, 90.210, 15.010, and 1.772, respectively; the heritability values using genetic and environmental variances were estimated to be 0.392, 0.350, 0.375, and 0.461, respectively.

In previous studies, Byun et al. (2021) reported that for 13,000 heads of the reference population, the estimated heritability of CWT, EMA, BFT, and MS was 0.41, 0.38, 0.38, and 0.44, respectively, and Kim et al. (2021) reported that the estimated heritability was 0.395, 0.347,

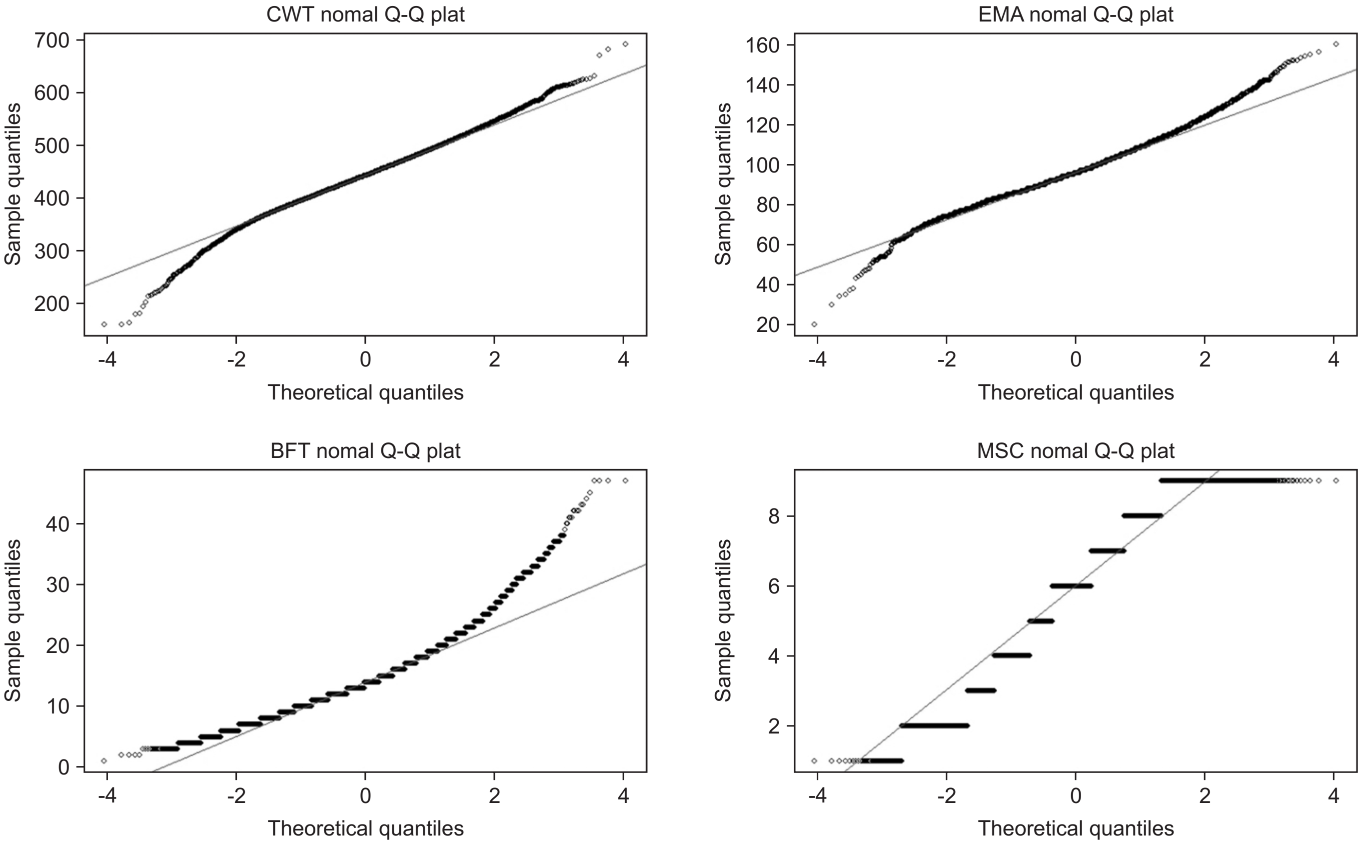

Fig. 1. Quantile-quantile plot (Q-Q plot) for the four carcass traits in the reference population.

Table 2. Heritability and genetic parameters of the reference population

Analysis method	Type	Trait			
		CWT	EMA	BFT	MS
GBLUP method	Genetic variance	866.3	48.550	9.012	1.515
	Residual variance	1346.0	90.210	15.010	1.772
	Heritability	0.392	0.350	0.375	0.461

GBLUP, genomic best linear unbiased prediction; CWT, carcass weight; EMA, eye muscle area; BFT, back fat thickness; MS, marbling score.

0.363, and 0.453, respectively. The heritability of CWT, EMA, BFT, and MS of Hanwoo raised by the National Institute of Animal Science was reported to be 0.39, 0.36, 0.44, and 0.35, respectively (Jang et al., 2022). Comparing the heritability estimated from previous research and the results of this study, it was confirmed that the heritability estimated using traits showed a slight difference but appeared to be almost similar.

Accuracy analysis of breeding value

The GEBV, GEBV accuracy, and realized accuracy of the subjects estimated using the GBLUP method are listed in Table 3. For CWT, EMA, BFT, and MS, the mean GEBV estimated using the GBLUP method were 3.819, 0.740, -0.248, and 0.041, respectively and the accuracy of each trait was 0.743, 0.728, 0.737, and 0.765, respectively. Accuracy was estimated to be the lowest in EMA and the highest in MS, and the accuracy of the breeding value was affected by heritability (Goddard, 2009); therefore, it is thought that accuracy was estimated to be low in EMA with low heritability, and high accuracy was estimated in MS. In previous studies, Byun et al. (2021) reported that the genetic ability evaluation accuracy of the Brindle Cattle was 0.44, 0.43, 0.42, and 0.44, for CWT, EMA, BFT, and MS, respectively. Furthermore, Jang et al., 2022) reported that the mean accuracy of breeding values estimated using GBLUP was 0.41, 0.38, 0.35, and 0.42 for CWT, EMA, BFT, and MS, respectively. These two previous studies showed lower accuracy than the current study because the Hanwoo reference population was used to analyze Brindle Cattle and the size of the reference population used was smaller than that used in this study. Kim et al. (2021) estimated the GEBV using a 16,972 heads of reference population and reported mean accuracies of 0.799, 0.779, 0.786, and 0.810 for CWT, EMA, BFT, and MS, respectively. Furthermore, Kim, (2021) reported that the mean of estimated accuracies in each subject for CWT, EMA, BFT, and MS were 0.68, 0.67, 0.67, and 0.70 respectively in the GBLUP analysis using 14,000 heads of reference population. By comparing the our results with these studies, it has confirmed that the mean accuracy was higher when using 14,000 heads of the reference population, but lower when using 16,972 heads. The reason why our mean accuracy was low compared to the 16,972-head reference paper is that the reference paper evaluated only the steer in the Gyeongnam region as the test population. On the other hand, in this study, Hanwoo population from Gyeonggi, Gangwon, Chungbuk, and Gyeongnam regions were evaluated as the test population. Therefore, it is considered that the accuracy may vary depending on the composition of the test population.

When checking the realized accuracy value with the actual phenotype, the values were reported to be 0.522, 0.404, 0.444, and 0.539 for CWT, EMA, BFT, and MS, respectively. The realized accuracy value was estimated to be the lowest for EMA and the highest for MS. Kim (2021) reported that for CWT, EMA, BFT, and MS, the realized accuracy values were 0.49, 0.40, 0.35, and 0.53, respectively; furthermore, Kim et al. (2021) reported that the realized accuracy values for each trait were 0.673, 0.526, 0.454, and 0.654, respectively. Compared with previous studies, the realized accuracy value also showed a similar pattern to the accuracy value; the larger the size of the reference population, the higher would be the accuracy value; however, it is thought that the difference in accuracy appears depending on the size of the test population and breeding region.

The results of this study suggest that the breeding value of each individual can be estimated with higher accuracy by estimating the GEBV using the genomic information of 18,499 reference populations. If this method is used and applied to individual selection in a commercial Hanwoo population, more precise and economical individual selection is possible. In addition, continuous verification

Table 3. Breeding value, accuracy, and realized accuracy estimated using the GBLUP method

Analysis method	Reference population size	Type	Trait			
			CWT (kg)	EMA (cm^2)	BFT (mm)	MS (score)
GBLUP method	18,499	GEBV	3.819	0.740	−0.248	0.041
		Accuracy	0.743	0.728	0.737	0.765
		Realized accuracy	0.522	0.404	0.444	0.539

GBLUP, genomic best linear unbiased prediction; CWT, carcass weight; EMA, eye muscle area; BFT, back fat thickness; MS, marbling score; GEBV, genomic estimated breeding value.

of the GBLUP model and establishment of a reference population suitable for commercial Hanwoo populations in Korea will enable a more accurate evaluation of individuals. The selection method using a genome requires less time than the traditional selection method, and it is possible to select the individual early in the calf period, suggesting that unnecessary cost reduction is possible.

Author Contributions: Conceptualization, H.S.K.; methodology, G.H.L., Y.H,L., H.S.K.; investigation, G.H.L., Y.H,L.; writing—original draft preparation, G.H.L., Y.H,L.; writing—review and editing, G.H.L., Y.H,L., H.S.K.; supervision, H.S.K.; project administration, H.S.K.; funding acquisition, H.S.K.

Funding: None.

Ethical Approval: The study was approved by the Hankyong National University Animal Ethics Committee (No.2018-1).

Consent to Participate: Not applicable.

Consent to Publish: Not applicable.

Availability of Data and Materials: Not applicable.

Acknowledgements: None.

Conflicts of Interest: No potential conflict of interest relevant to this article was reported.

REFERENCES

Byun SK, Kim DH, Oh JD, Lee HK. 2021. The analysis on evaluation of genomic breeding value in Brindle cattle using reference population of Hanwoo. J. Anim. Breed. Genom. 5:91-98.

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4:7.

Gao N, Teng J, Ye S, Yuan X, Huang S, Zhang H, Zhang X, Li J, Zhang Z. 2018. Genomic prediction of complex phenotypes using genic similarity based relatedness matrix. Front. Genet. 9:364.

Garrick DJ. 2007. Equivalent mixed model equations for genomic selection. J. Anim. Sci. 85:376.

Goddard M. 2009. Genomic selection: prediction of accuracy and maximisation of long term response. Genetica 136:245-257.

Heffner EL, Sorrells ME, Jannink JL. 2009. Genomic selection for crop improvement. Crop Sci. 49:1-12.

Henderson CR. 1975. Best linear unbiased estimation and prediction under a selection model. Biometrics 31:423-447.

Henderson CR. 1984. Applications of Linear Models in Animal Breeding. University of Guelph, Guelph.

Henderson CR and Quaas RL. 1976. Multiple trait evaluation using relatives' records. J. Anim. Sci. 43:1188-1197.

Jang MJ, Lim DJ, Park WC, Park JE. 2022. Estimation of genetic prediction accuracy using convolutional neural network in Hanwoo. J. Korea Acad. Ind. Coop. Soc. 23:516-523.

Kim DH. 2021. Studies on the genetic evaluation of Hanwoo using the genomic information [Doctoral dissertation, Jeonbuk National University]. RISS. http://www.riss.kr/link?id=T15777082

Kim EH, Sun DW, Kang HC, Kim JY, Myung CH, Lee DH, Lee SH, Lim HT. 2021. A study of the genomic estimated breeding value and accuracy using genotypes in Hanwoo steer (Korean cattle). Korean J. Agric. Sci. 48:681-691.

Lee JJ. 2012. Genetic evaluation for carcass traits of Hanwoo using pedigree and SNP marker-derived relationship matrix [Doctoral dissertation, Chungbuk National University]. RISS. http://www.riss.kr/link?id=T12647293

Meuwissen TH, Hayes BJ, Goddard ME. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819-1829.

Misztal I, Tsuruta S, Lourenco D, Masuda Y, Aguilar I, Legarra A, Vitezica Z. 2018. Manual for BLUPF90 Family of Programs. University of Georgia, Athens.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81:559-575.

Son JH, Koo YM, Jeoung YH, Kim JI, Cha DH, Kim SJ, Choi TJ, Park MN, Lee DH, Lee JH. 2021. Estimation of the genetic gain of carcass traits in Hanwoo Korean proven bull and cow. J. Agric. Life Sci. 55:81-88.

VanRaden PM. 2007. Genomic measures of relationship and inbreeding. Interbull Bull. 37:33-36.

VanRaden PM. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91:4414-4423.

Visscher PM, Medland SE, Ferreira MA, Morley KI, Zhu G, Cornes BK, Montgomery GW, Martin NG. 2006. Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. PLoS Genet. 2:e41.

Zhang Z, Todhunter RJ, Buckler ES, Van Vleck LD. 2007. Technical note: use of marker-based relationships with multiple-trait derivative-free restricted maximal likelihood. J. Anim. Sci. 85:881-885.