Indian Journal of Biotechnology and Pharmaceutical Research
Journal homepage : www.ijbpr.in
ISSN: 2347-3266
2015, 3(3): 1- 4

Short Communication

Accuracy of BayesCπ for genomic breeding value prediction for Intramuscular fat in Hanwoo cattle

Aditi Sharma[1], Yongmin Cho[1]*, Bong-Hwan Choi[1], Dajeong Lim[1], Han-Ha Chai[1], Seung-Hwan Lee[2]*

[1]Animal Genomics and Bioinformatics Division, National Institute of Animal Science, RDA, Jeonju 565-851, Korea
[2]Division of Animal and Dairy Science, Chungnam National University, Daejeon 305-764, Korea

ABSTRACT

Hanwoo cattle play a major role in beef production in Korea. So it is of great interest to determine if genomic selection can, in any way, excel genetic improvement of various production traits in this breed. Because of the reduction of cost of genotyping and increased availability of genome-wide sets of molecular markers, genomic estimated breeding values have become an essential resource in animal breeding. The genomic estimated breeding values are used in genomic selection to predict the genetic merit of the candidate. Genomic selection refers to incorporation of DNA marker information, often the whole genome SNP data, to predict the genomic breeding values (GEBV) to make selection decisions. Genomic prediction is believed to provide better genetic gain for quantitative traits than could be achieved by phenotypic data alone (Meuwissen et. al, 2001). Selection based on genomic data can be applied early in life without sacrificing the selection candidates, which is apparently the most important advantage of this method. Bayesian methods for genomic breeding value (GEBV) estimation have proven to be accurate and efficient. In this study we report the accuracy of prediction using Bayes Cπ method for intramuscular fat in Hanwoo cattle. The accuracy of prediction was found to be 0.38 (N=778) which decreased as the proportion of markers included in Bayes Cπ (Habier et al., 2011) increased.

KEY WORDS: Hanwoo, IMF, Genomic prediction

Introduction

Hanwoo cattle is integral to the beef industry of Korea and there is keen interest in determining if genomic selection can excel the genetic improvement of carcass and meat quality traits in this cattle breed. Cost reduction and availability of whole genome genotype data has given the much necessary boost to field of animal genetics and breeding. Predicting the genetic merit of an individual based on the genotypes and the accuracy of prediction forms the basis of genomic selection strategies. Combined with the reproduction techniques, genomic selection can substantially increase the genetic gain and shorten the generation interval (Meuwissen et. al., 2001).

Intramuscular fat content of cattle muscle is an important component that influence eating quality, such as meat tenderness, juiciness, and taste (Hovenier et al., 1993). In the present study we report genomic prediction for intramuscular fat in Hanwoo cattle using BayesCπ.

Received: 15.10.2015, Revised: 05.11.2015, Accepted: 21.11.2015
***Correspondence:**
[1]Yongmin Cho, Tel.: +82-632387301.
E-mail: variance@korea.kr
[2]Seung-Hwan Lee, Tel.: +82-428215772.
E-mail: slee46@korea.kr

While conventional breeding programs select animals based on phenotype and pedigree information Bayesian methods select animals based on the SNP markers and the LD between them. The accuracy of prediction depends of number of factors like the size of the reference population, heritability of the trait,number of QTLs influencing the trait, relationship between the predicted animals and the animals in the training set etc. A number of statistical models have been used in previous studies for genomic predictions including GBLUP, BayesC, BayesB, Bayes Cπ etc. (Rolf et al., 2015; Zeng et. al., 2012; Habier et al., 2011).

For the present study the data set included a total of 778 animals (706 animals with phenotypes, genotypes and pedigree, 72 animals were bulls of 706 animals) that were genotyped with BovineSNP50 BeadChip (Matukumalli et al., 2009). The data set was split into training and validation data sets for the genetic prediction. The training data set consisted of 706, and the validation data set consisted of 72 genotyped animals with no phenotype information (Table1). Intramuscular fat was chemically measured. In the final dataset, SNP were removed if the call rate was less than 95%, if the Illumina Gentrain score was less than 0.7, if the minor allele frequency was less than 0.01, if the SNP was not in Hardy-Weinberg equilibrium (a P-value cut-off of 1×10−15), if the genome location was unknown or if the SNP showed complete linkage disequilibrium (r2 > 0.99) with another SNP on the chip. Missing genotypes were imputed using fastPHASE (Sheet and Stephens, 2006).

Table 1 Summary statistics for intramuscular fat

Population	n	Mean	S.D.	Min	Max
Tested population	706	10.87	3.57	3.90	25.21
Validation population	80	11.35	4.00	2.29	24.00
Total	786	10.92	3.61	2.29	25.21

Since several studies have shown that giving more weight to SNPs with large effects may improve the accuracy of genomic predictions (Calus et al., 2014), therefore in this study we utilized the Bayes Cπ statistical model (Habier et al., 2011) as implemented in GS3 (Legarra et al., 2010) to study the accuracy of prediction and the proportion of QTLs that explains the genetic variance. In Bayes Cπ the proportion of markers included in the model is assigned a prior distribution and estimated during the analysis. Four assumptions were made i.e., 1) Number of QTL's that effect the phenotype= 50, 2) Number of QTL's that effect the phenotype=100, 3) Number of QTL's that effect the phenotype = 500, 4) Number of QTL's that effect the phenotype =1000.

$$y_i = \text{othereffects} + \sum_{j=1}^{n} (Z_{ij} a_j \delta_j) + e_i$$

y_i is the phenotype of the i-th animal, Z_{ij} is an indicator covariate for the i-th animal and the j-th marker locus , and e_i is a residual term. δ is indicator variable stating whether the marker has any effect with $\delta_j = (0, 1)$. The distribution of $\delta = (\delta_1 \ldots \delta_n)$ can be posited as a binomial, with probability π. Posterior distribution of variances was computed using a Monte Carlo Markov chain of 50,000 iterations, with a burn-in of 10,000 iterations.

In our analysis, Bayes Cπ performed the best for assumption 1 i.e., some QTLs (50) have a relatively large effect on the phenotype while others have small effect (Table 2). The predictive accuracy of Bayes Cπ however decreased as

Table 2 The correlation between the phenotype and GEBV for intramuscular fat

Number of QTL	50	100	500	1000
Accuracy	0.3795	0.3283	-0.0417	0.0037

the model of variation became more polygenic. The accuracy with 50 QTL model was 0.38 while for 1000 QTL model it was 0.0037. Habier et al. (2007) identified two components contributing to prediction accuracy in genomic selection, one due to LD between markers and QTL and the other due to genetic relationships between individuals that can be captured in the absence of LD. We found that there were fewer QTL actually responsible for large proportions of the genetic variance. In the 50 QTL model, the largest QTL explained ~2% of the variation. For the

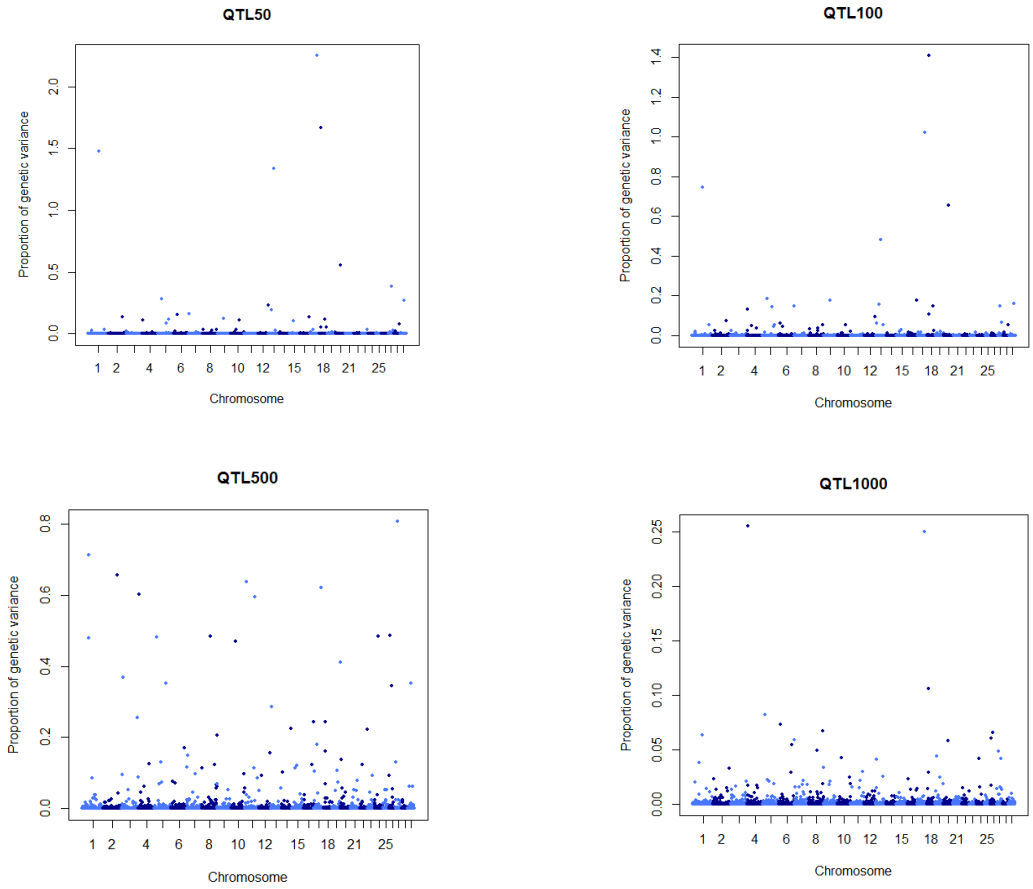

Figure 1 Manhattan plot of the proportion of markers included in Bayes Cπ

1,000 QTL model, the largest QTL explained 0.25% of the genetic variation (Figure 1). Though the accuracy is low as compared with those observed in other meat type cattle yet genomic selection would be beneficial for such traits that are hard to measure.

Acknowledgements

This work was carried out with the support of "Cooperative Research Program for Agriculture Science & Technology Development (Project No. PJ01022002)" Rural Development Administration, Republic of Korea and (2015) Postdoctoral Fellowship Program of (National Institute of Animal science), Rural Development Administration, Republic of Korea.

References

Calus M. P. L., Schrooten C. and Veerkamp R. F., 2014. Genomic prediction of breeding values using previously estimated SNP variances. *Genetics Selection Evolution*, **46**(1):52. doi:10.1186/s12711-014-0052-x

Habier, D., Fernando R. L., Kizilkaya K. and Garrick D. J., 2011.Extension of the Bayesian alphabet for genomic selection. *BMC Bioinformatics* **12**: 186.

Hovenier R., Kanis E., Asseldink T. Van and Westerink N. G,1993. Breeding for pig meat quality in halothane negative populations-a review. *Pig News Info*. 14:17N-35N.

Legarra A., Ricardi A., Filangi O., 2011.GS3: Genomic Selection, Gibbs Sampling, Gauss-Seidel (and BayesCπ). http://snp.toulouse.inra.fr/~alegarra/

Matukumalli L. K., Lawley C. T., Schnabel R. D., Taylor J. F., Allan M. F., Heaton M. P., O'Connell J., Moore S. S., Smith T. P. L., Sonstegard T. S., and Tassell C. P. V. 2009. Development and characterization of a high density SNP genotyping assay for cattle. *PLoSONE* **4**(4):e5350.

Meuwissen T. H. E., Hayes B. J. and Goddard M. E, 2001.Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics* **157**: 1819-1829.

Rolf M. M., Garrick D. J., Fountain T., Ramey H. R. Weaber R. L., Decker J. E., Pollak E. J., Schnabel R. D. and Taylor J. F., 2015.Comparison of Bayesian models to estimate direct genomic values in multi-breed commercial beef cattle. *Genetics Selection Evolution* **47**: 23; doi:10.1186/s12711-015-0106-8

Sheet P. and Stephens M.A. 2006. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *American Journal of Human Genetics* **78**, 629-644.

Zeng J., Pszczola M., Wolc A., Strabel T., Fernando R. L., Garrick D. J. and Dekkers J. C. M., 2012. Genomic breeding value prediction and QTL mapping of QTLMAS2011 data using Bayesian and GBLUP methods. *BMC Proceedings* **6**(2):S7.

How to Cite This Article:

Aditi Sharma, Yongmin Cho, Bong-Hwan Choi, Dajeong Lim, Han-Ha Chai, Seung-Hwan Lee 2015. Accuracy of BayesCπ for genomic breeding value prediction for Intramuscular fat in Hanwoo cattle. Ind. J. Biotechnol. Pharma. Res. 3(3): 1-4.