

Genomic prediction by single-step genomic BLUP using cow reference population in Holstein crossbred cattle in India

N G Nayee¹, G Su², S G Gajjar¹, G Sahana², S Saha¹, K R Trivedi¹, B Guldbrandtsen², M S Lund²

nileshn@nddb.coop (Corresponding Author)

¹ National Dairy Development Board, Anand 388 001, Gujarat, India

² Center for Quantitative Genetics and Genomics, Department of Molecular Biology and Genetics, Aarhus University, , 8830 Tjele, Denmark

Introduction

Many factors can affect the accuracy of genomic prediction, among them two important factors are number of animals with phenotypes and genotype information used as reference population and the heritability of the trait (Goddard, 2008; Hayes et al, 2009).

Dairying in India is done mainly by small holders. Average herd size (cattle and buffalo) in the districts under milk recording in present study is 2.2 animals (19th Livestock census, Govt. of India, 2012). Performance recording and genetic evaluation under this condition poses a big challenge. Due to operational constraints, the annual numbers of progeny tested bulls are limited to 20-40. Advantages of genomic selection in breeds with limited number of progeny tested bulls have been demonstrated by use of females in the reference population (Thomasen et al 2014). As the number of animals with genotype information was limited compared to number of animals with phenotypic records, single-step genomic BLUP (ssGBLUP) was a logical choice to use marker information for genetic evaluation.

The current study explores the feasibility of implementing genomic selection in Holstein crossbred cows under small holder conditions, using cows as reference population.

Material and methods

Phenotype data

Since 1992 a progeny testing program for Holstein Crossbred cattle has been operated by Sabarmati Ashram Gaushala, Bidaj, the largest semen station in India along with three district cooperative milk unions (Panchmahal, Sabarkantha and Surat), in western part of India (Gujarat). For the present study, first lactation test day milk yield records of 10,797 daughters sired by 258 bulls were used. Information on location of animal (village, block and district), date of birth of daughter, calving and milk recording were extracted from the INAPH database (Nayee et al 2016).

Farmers in this area mainly maintain Holstein × *Bos Indicus* crossbred cattle as well as buffaloes for milk production and herd size ranges from 1-5 milking animals per farmer. Animals are reared under stall fed conditions with moderate plane of nutrition.

Genotype data

National Dairy Development Board, Anand, India (NDDB) has developed a custom SNP panel based on Illumina platform for genotyping *Bos Indicus* and their taurine crosses. A total

of 2,194 daughters and 109 sires whose DNA samples were available were genotyped. Data for 36,813 SNPs on autosomes with average call rates above 95% per individual, 90% for a locus, and minor allele frequency (MAF) higher than 0.01 were used.

Model for estimation of breeding values

Breeding values were estimated using a conventional test day model and ssGBLUP model. The two models were described below.

1. Conventional test day random regression model with 3rd order Legendre polynomials for both fixed and random regression.

where is the test-day milk yield of cow k produced - within the h^{th} HA (herd \times age at calving, fixed effect), the i^{th} HYMR (herd \times year of recording \times month of recording, random effect) and j^{th} YS (year \times season of calving, fixed effect); is the fixed regression coefficient of the test day record on the l^{th} order of Legendre polynomial; and are random regression coefficients of the test day record on the l^{th} order of Legendre polynomial for animal additive genetic and permanent environmental effects for animal k ; is the l^{th} Legendre polynomial of the t^{th} days in milk for cow k .

Considering that individual farmers only have a few cows and farmers in the same village have more common management practices compared with farmers in different villages, a herd was defined as all animals in the same village. Age at first calving ranged from 18 months to 72 months. Age class was defined by combining animals less than 2 years of age at first calving in to a single class. Animals in every 12 month age increment at first calving were put in separate age class groups till 5 years. Animals above five years of age at first calving were grouped in a single class.

2. ssGBLUP model

The ssGBLUP (Christensen et. al. 2012) had the same structure and effects as the conventional model. The only difference was that the ssGBLUP used a combined relationship matrix (\mathbf{H}) instead of pedigree-based relationship matrix (\mathbf{A}). The inverse of the \mathbf{H} matrix is

$$\mathbf{H}^{-1} = \begin{bmatrix} \mathbf{G}_w^{-1} - \mathbf{A}_{11}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \mathbf{A}^{-1}$$

Here \mathbf{G}_w was the adjusted \mathbf{G} matrix, $\mathbf{G}_w = 0.8 \cdot \mathbf{G} + 0.2 \cdot \mathbf{A}$, in which \mathbf{G} was the genomic relationship matrix and \mathbf{A} was pedigree based relationship matrix. To get \mathbf{G} in the same scale as \mathbf{A} , the scale of \mathbf{G} was adjusted so that the average of diagonals and average of off-diagonals were the same as those in \mathbf{A} (Christensen et al., 2012). Off diagonal terms in \mathbf{G} were used to detect errors in pedigree relationship between sires and daughters. Consequently, 571 genotyped daughters out of 2194 genotyped were not compatible with genotypes of the sire in the pedigree record. Hence, sire for these cows were set to missing in the pedigree before further processing.

\mathbf{G} was calculated by the VanRaden method 1 (VanRaden, 2008) using Gmatrix (<http://dmu.agrsci.dk/Gmatrix/>). DMU V 5.2 (Madsen et al., 2006; <http://dmu.agrsci.dk/DMU/>) was used to estimate variance components and predicting breeding values.

Validation of conventional and ssGBLUP models

The performance of the conventional test day model and the ssGBLUP model was assessed using a 5-fold cross validation in two scenarios. In scenario 1, half-sib groups in the whole data were randomly divided into 5 subsets (leave-family-out). In scenario 2, individuals in the whole data were randomly divided into 5 subsets. For each fold of validation, one subset was used as the test data set and the remaining four subsets for training. The phenotypic records of animals in the test data were masked, and their breeding values were predicted from the information of training data. In addition, sires were also used as validation animals in scenario 1 where sires did not have daughters in training data.

In scenario 1 the test animals did not have any half sibs in the training data, and thus had more distant relationships to the training animals, whereas in scenario 2 the test animals 2 had half sibs in the training data, and thus had relatively closer relationships with the training animals.

In the procedure of cross validation, breeding values were estimated using the above two models. The variance components obtained by conventional BLUP model were used to calculate corrected phenotype Y_c and EBV. For estimating GEBV, variance components obtained by ssGBLUP model were used.

The performance of the conventional BLUP model and the ssGBLUP model were assessed for their accuracy of prediction of EBV, which was calculated as correlation between 305 day EBV and corrected 305 day phenotypic values (Y_c). The correlation was calculated for genotyped and non-genotyped validation animals separately.

Y_c for 305 day was calculated from conventional test day model by adding breeding value over 305 days (BV305), permanent environment effect over 305 days (PE305) and residuals over 305 days (R305). BV305 and PE305 were obtained from parameters estimated using test day model following the method described by Mrode (2005). R305 was obtained from test day residuals following the logic of the ICAR test interval method for obtaining lactation yield from test day yields.

Results and discussion

Table 1 shows the correlations between Y_c and predicted breeding value from conventional BLUP model (EBV) and between Y_c and predicted breeding value from ssGBLUP model (GEBV) for genotyped and un-genotyped animals in validation scenario 1 of leave-family-out. Correction of pedigree by genotype information improved predictions only for non-genotyped animals. ssGBLUP led to higher accuracy of predictions than conventional BLUP model for both genotyped and non-genotyped animals.

Table 1: Correlations between Y_c and predicted breeding values using different approaches in validation scenario 1.

Genotyping status	No. of animals	Correlations			
		Y_c and EBV prior	Y_c and GEBV prior	Y_c and EBV after	Y_c and GEBV after

		to pedigree correction	to pedigree correction	pedigree correction	pedigree correction
Genotyped	1542	0.247	0.288	0.243	0.276
Not genotyped	7809	0.136	0.149	0.161	0.173
All validation animals	9351	0.150	0.177	0.174	0.194

In scenario 2, ssGBLUP produced higher accuracy of predictions than conventional BLUP model for genotyped animals, but not for non-genotyped animals (Table 2). Correction of pedigree slightly reduced the correlation between predictions and Y_c . The reason for the unexpected result remains unclear.

Table 2: Correlations between Y_c and predicted breeding values using different approaches in validation scenario 2.

Genotyping status	No. of animals	Correlations			
		(Y_c, EBV), prior to pedigree correction	(Y_c, GEBV), prior to pedigree correction	(Y_c, EBV), after pedigree correction	(Y_c, GEBV), after pedigree correction
Genotyped	1814	0.397	0.413	0.387	0.405
Not genotyped	8527	0.294	0.293	0.289	0.289
All validation animals	10341	0.312	0.318	0.306	0.312

Based on scenario 1 and correlated pedigree, correlations between EBV and GEBV of sires with average daughter Y_c , were calculated (Table 3). The correlations from ssGBLUP were much higher than those from conventional BLUP for both genotyped sires and all sires.

Table 3: Correlation between average Y_c of daughters and predicted breeding values of sires.

Bull category	Correlation Avg. Daughter Y_c and EBV	Correlation Avg. Daughter Y_c and GEBV
All sires	0.126	0.202
Genotyped sires	0.127	0.199
Bulls not genotyped	0.029	0.094

In practical dairy cattle breeding, when bulls are selected for progeny test, their female sibs do not have performance records yet. Therefore, validation scenario 1 more closely approaches real-life conditions than scenario 2. The current study showed a large increase of prediction accuracy in both validations on cows (Table 1) and on sires (Table 3) by genomic prediction using the ssGBLUP model. It indicates that the use of genomic information can increase accuracy of predicting breeding values substantially, and speed up genetic progress.

Conclusions

ssGBLUP is suited to estimate genomic breeding values in HF crossbred cattle even though only recent female animals have been genotyped. The predictive ability of ssGBLUP for both bulls and cows has proven to be higher than conventional breeding values in current study.

Acknowledgement

The authors sincerely thank Dr. CT Patel, Sabarmati Ashram Gaushala, Bidaj, Dr. Sanjay Dholariya, Sumul Dairy, Surat, Gujarat, Dr. D V Shah, Panchamrut Dairy, Godhra, Gujara and Dr. Sameer Patel, Sabar Dairy, Himmatnagar, Gujarat, who are implementing Progeny Testing Project for HF Crossbred cattle and have arranged for collection of samples from recorded daughters. Genotyping for the samples was carried out at M/s. Sandor Life Sciences, Hyderabad.

Under Indo-Danish collaboration, Department of Molecular Biology and Genetics (MBG), Aarhus University, Denmark and National Dairy Development Board (NDDDB), Anand, Gujarat, India has collaborated in a joint group to explore and implement genomic selection in Indian cattle. Present work was designed and executed as part of this collaborative effort. We sincerely thank NDDDB management for financial support and MBG scientists for technical support for this work.

List of References

- Christensen, O. F., Madsen, P., Neilsen, B., Ostersen, T. and Su, G. 2012. Single-step methods for genomic evaluation in pigs. *Animal*. 6:10: 1565-1571.
- Goddard, M. E. 2008. Genomic selection: Prediction of accuracy and maximisation of long term response. *Genetica* doi:10.1007/s10709-008-9308-0.
- Madsen, P., Jensen, J., Labouriau, R., Cristensen, O.F. and Sahana, G. 2014. DMU- a package for analyzing multivariate mixed models in quantitative genetics and genomics. In: Proc. 10th World Congress of Genetics Applied to Livestock Production.
- Mrode, R. A. and Thompson, R. J. 2005. In: *Linear models for the prediction of animal breeding values* (Second Edition), CABI Publishing, CAB International, Wallingford, Oxfordshire OX10 8 DE, UK., p143-154.
- Nayee, N. G., Ambaliya, R.M. Saha, S., Trivedi, K. R. and Namjoshi, M. N., 2016. National database for enhancing the productivity of cattle and buffaloes - Suitability of INAPH to create a national database., *Indian Dairyman*.
- Thomassen, J. R., Sørensen, A. C., Lund, M. S. and Guldbrandsten B., 2014. Adding cows to the reference population makes a small dairy population competitive. *J. Dairy Sci.* 97: 5822-5832.
- VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. *Journal of dairy science* 91(11):4414-4423.