



Genomic selection in Gir cattle using female reference population

NILESH NAYEE¹, SWAPNIL GAJJAR¹, A SUDHAKAR¹, SUJIT SAHA¹, KAMLESH TRIVEDI¹ and PRAVIN VATALIYA²

National Dairy Development Board, Anand, Gujarat 388 001 India

Received: 14 September 2020; Accepted: 7 November 2020

ABSTRACT

When a sizeable reference population of proven bulls is not available for implementing Genomic selection for a particular trait, and when a recording of certain traits on large scale is difficult, the use of a female reference population is recommended. Gir, one of the important milk purpose cattle breeds of India falls under this category. There is no large scale Progeny Testing (PT) programme in Gir, so proven bulls based on daughter performance in large numbers are not available. Considering the constraints, a genomic BLUP (GBLUP) model was implemented based on recorded cow reference population in Gir breed. Cows (3491) and 23 bulls were genotyped using INDUSCHIP for this purpose. Due to non-availability of pedigreed data, conventional breeding values (BV) of bulls and their reliabilities were not known. For comparison, assumed theoretical reliability of BV of a bull selected based on its dam's yield was compared with reliability obtained for genomic breeding value (GBV) using a GBLUP model. The reliability estimates for GBVs were 4 times higher than that for BVs. The predictive ability of the model was demonstrated by measuring the correlation between corrected phenotypes and GBVs for animals whose records were masked in a five-fold cross-validation study. The correlation was around 0.45 showing reasonable predictability of the GBLUP model. The GBVs were not biased. The regression coefficient between the corrected phenotype and GBV was 1.045. The present study demonstrates that it is feasible to implement genomic selection in Gir cattle in Indian conditions using a female reference population. It is expected that the bulls can be selected with around 4 fold more accuracy than the current method of selecting based on their dams' yield accelerating expected genetic growth in Gir cattle.

Keywords: Female reference population, Genomic selection, Gir

Implementation of genomic selection is expected to enhance the rate of genetic progress in various cattle and buffalo breeds in India (Rath *et al.* 2017). Ma *et al.* (2015) showed that in dairy cattle, usually, progeny tested bulls are used to form a reference population for genomic selection. In large populations, such as Holsteins, accurate prediction of genetic worth using genomic information has been obtained (VanRaden *et al.* 2009, Lund *et al.* 2011). However, in the absence of large scale progeny testing programmes for indigenous cattle breeds in India, such reference populations are not available at present. Incidentally, the use of females as a reference population has been gaining popularity, especially for novel traits, such as feed efficiency, methane emissions, and some reproductive measures, the traits which are expensive to measure and are therefore available for some animals in the population (Pryce *et al.* 2012). So, for indigenous breeds of cattle in India, genomic selection can be implemented if we have performance records of a sizeable number of cows as a reference population. One such breed is Gir. Under

National Dairy Plan I (<https://www.nddb.coop/ndpi/about/brief>), performance records were collected from a sizeable number of Gir cows reared by farmers in the native tract of the breed. This has opened up a new avenue for implementing genomic selection in Gir cattle under Indian conditions. The present study was undertaken to examine the utility of Genomic Selection based on a female reference population over the currently used selection method for Gir bull calves in India.

MATERIALS AND METHODS

Phenotype records: Performance records of Gir cows calved during the years 2016 to 2019 were used for the present study. A Test Day recording scheme as per the ICAR method AZ44,2X (ICAR Guidelines, 2020) was implemented for collecting milk production records in Gir cattle reared by smallholders in the native breeding tract in the Saurashtra region in the western part of Gujarat State in India, by appointing official milk recorders under the Pedigree Selection project implemented by Sabarmati Ashram Gaushala (SAG), Bidaj under National Dairy Plan I. The performance records were collected through Information Network for Animal Productivity and Health

Present address: ¹National Dairy Development Board, Anand, Gujarat, India. ²Kamdhenu University, Gandhinagar, Gujarat, India. ✉Corresponding author email: Nileshn@Nddb.coop

(INAPH) App developed by NDDB (Nayee *et al.* 2016). The 305-day standard lactation milk yield records were extracted from the INAPH database. The INAPH application internally calculates 305-day milk yield using the test interval method (ICAR Guidelines, 2020). Records for animals having less than 500 kg or more than 6,000 kg of lactation yield were considered as outliers and not included in the analysis. Only animals that had their first record between day 5 and 40 and had subsequent test day records at the intervals of 20 to 40 days apart were retained for the present study. A total of 9,108 Gir cows with single lactation records were available for the study.

Pre-correction of phenotype for environmental factors: Traditionally, de-regressed proofs of bulls and females are used as a dependent variable in the model (Garrick *et al.* 2009, Campos *et al.* 2018). However, due to non-availability of pedigree derived breeding values in the current study, corrected phenotypes were used as a dependent variable for genomic breeding value estimation.

The data were pre-corrected for fixed effects using a linear model for genomic breeding value estimation. Each phenotype was corrected for the estimated fixed effects obtained from the linear model as described below and the corrected phenotype records were used for further analysis.

$$y_{ijkl} = \mu + L_i + S_j + H_k + e_{ijkl}$$

where y_{ijkl} is the 305 day standard lactation milk yield of l^{th} animal, μ is overall mean, L_i is the fixed effect of i^{th} lactation, S_j is the fixed effect of j^{th} season and H_k is the fixed effect of k^{th} herd (Herd refers to the Block, where the animal is located) and e_{ijkl} is the residual effect unaccounted by the model for individual animal record.

Genotypes: Of the 9,108 recorded animals, 2,571 animals were genotyped using INDUSCHIP developed by National Dairy Development Board (Saha *et al.* 2020) having around 53 K SNPs covering the whole genome. Other 804 animals that were from the same herds or villages were genotyped using INDUSCHIP.

Additionally, 117 animals that included 23 bulls used for Artificial Insemination in the area and cows from prominent breeders were genotyped by BovineHD Beadchip (Illumina, San Diego, CA, USA). INDUSCHIP genotypes were extracted from HD genotypes for these animals and were combined with the other genotype data obtained using INDUSCHIP.

Standard quality checks were done using PLINK software (Purcell and Chang, www.cog-genomics.org/plink/1.9/). The genotype data were removed for loci that were having more than 10% missing genotypes, the animals having less than 90% SNPs genotyped and data for SNPs with less than 1% minor allele frequencies.

After employing the quality filters, genotype data for 45,875 SNPs and 3,514 animals were available for further analysis.

GBLUP: Not having pedigree data for all animals, corrected phenotype records of 2,571 animals with genotype data were used. Genetic analysis was done by GBLUP

model using genotypes of 3,514 animals. The following model was employed for genomic analysis:

$$y = \mu + Xa + e$$

where y is a vector of corrected phenotypes, μ is the intercept, a is the vector of individual Genomic Breeding Values (GBV), and X is an incidence matrix relating GBV to corresponding corrected phenotypes and e is random error. Genetic effects were assumed to follow a normal distribution $a \sim N(0, G\sigma_{ak}^2)$, where σ_{ak}^2 is Genetic Variance and G is the genomic relationship matrix and σ^2e is error variance. Genomic relationship matrix was calculated by following the VanRaden method 1 (VanRaden 2008) using the Gmatrix programme (<http://dmu.agrsci.dk/Gmatrix/>).

Variance components for milk production trait were estimated using AIREML (Average Information Restricted Maximum Likelihood) using DMU software (Madsen *et al.* 2014). Breeding values of the animals were estimated by the GBLUP model using the variance components obtained from the above analysis using DMU software.

Evaluating predictability of Genomic Breeding Values: As the pedigree records were not available and the records did not comprise the known sire and daughter relationships, a classical validation could not be carried out. Hence, the predictive ability of genomic breeding values was assessed in the form of Pearson's correlation between the GBVs (estimated when the phenotype of animals is masked) and the corrected phenotype of the animals. For cross-validation, five validation datasets were prepared by each time-masking milk yield records of randomly selected 20% animals (i.e. validation animals). GBVs were estimated for the validation animals and were compared to the corrected yields of these animals. Thus, it was mimicked to the actual selection situation wherein, at the time of selection of animal, performance record or daughter based EBV of heifer or bull calves is not available. Only dam's performance record is available for young heifers and bull calf. For each validation set, records from 514 animals were masked and 2,057 animals' records were used for genomic breeding value estimation except for one validation dataset where records for 515 animals were masked.

A hypothetical correlation of 0.1 was assumed between the corrected phenotype of validation animals and their BV calculated using pedigree information based on past experiences in crossbred cattle in smallholder conditions (Nayee *et al.* 2018, Gajjar *et al.* 2018). The extent of benefit of GBV was compared against 0.1 correlation for assessing the value of GBV against the existing selection scheme in Gir cattle.

Reliabilities of GBVs were obtained as $r^2 = 1 - (PEV / Var_g)$ where r^2 is reliability of breeding value, PEV is Predicted Error Variance of breeding value and Var_g is Additive Genetic Variance based on the formula $PEV = (1 - r^2)\sigma_a^2$ (Mrode and Thompson 2005_a).

As described by Mrode and Thompson (2005_b), the reliability of breeding value of an animal based on pedigree records will be as given below:

$$r_0 = \left(\frac{1}{2} \sqrt{rs^2 + rd^2} \right)$$

where r_0 is the accuracy of offspring breeding value, rs^2 and rd^2 are the reliabilities of breeding values of sire and dam respectively. Hence, the reliabilities obtained for GBV of animals whose records were masked were compared with the assumed reliability to estimate an increase in reliability of GBV.

Detecting biasness in GBVs: Tsuruta *et al.* (2019) in a simulation study demonstrated that when males had no daughters and no siblings with phenotypes, if the regression coefficient of True Breeding Value (TBV) on GBV was less than 0.9, it represented inflation in estimation of GBVs. This they termed as 'Bias' in the estimation of GBV. To test the bias, a regression slope of TBV on GBV was examined. If the GBVs are biased upward or downward, the slope will be considerably deviate from 1.0.

In the present study, since either TBV or BV based on daughter records was not available, the bias of GBV was estimated by calculating a regression coefficient of corrected yield on GBVs taking animals having records. The data were plotted using R software (R Core Team 2019).

RESULTS AND DISCUSSION

Standard lactation milk yield by parity in Gir cows used under the study are provided in Table 1.

Estimates for various effects obtained by the fixed effect model were significant. Here different animals were recorded in different parity (instead of the same animal recorded in different parity) hence the correction represents age correction instead of actual parity correction. The lactation records were corrected for fixed effects. The corrected yields of animals having genotype information were further used for genomic evaluation.

Variance components and heritability: The variance components obtained for 305 DMY for Gir cows are given in Table 2.

The heritability estimate obtained for Gir cattle in the present study (0.46) is on the higher side than the reports available in the literature for indigenous cattle using pedigree information such as 0.17 to 0.32 in Guzerat cattle (Peixoto *et al.* 2006), 0.16 to 0.24 in Gyr breed (Ledic *et al.* 2002), 0.14 to 0.34 in Gir (Herrera *et al.* 2008), 0.14 to 0.24 in Guzerat breed (Santos *et al.* 2013), and 0.05 to 0.32 in Sahiwal cattle (Dongre and Gandhi 2014). Probably, the

Table 1. Parity wise 305 day milk yield (305 DMY) in Gir cows under the study

Parity number	No. of observations	Mean 305 DMY (kg)	SD
1	2478	2051	756
2	2438	2055	780
3	2458	1964	731
4 and above	1733	1991	729
Grand total	9107	2017	752

use of a marker-based G matrix instead of a pedigree-based A matrix has inflated the heritability estimates in the current study.

Correlations between corrected phenotype and GBV: The correlation of GBVs of animals whose records were dropped with the corrected phenotypic record are provided in Table 3.

As shown in Table 3, the correlation between GBVs and corrected yields of animals was 45% for all validation animals against 89% when using complete data.

Genomic breeding values and reliabilities: Genomic Breeding Values obtained for bulls under the present study based on their genotypes using the full data set are given in Table 4.

As evident from Table 4, the GBV for bulls ranged from -186.55 to 1017.42. The average reliability of bulls GBV was 41% and it varied from 5% to 82% depending upon the genomic relationship of the bull with other animals in the reference population.

According to current selection practices in India, all these bulls were selected solely based on their dam's record. As mentioned in the material and methods, the reliability of a progeny selected bulls only based on one parent's performance will be 1/4th of the heritability, i.e. 11.5% in the current study. Thus, the average reliability of bulls has increased by about 4 folds in the present study. The increase in reliability will accelerate genetic progress if bulls are selected based on their Genomic Breeding Values (GBV) instead of the current practice of using their dam's performance record. The correlation of GBVs with

Table 2. Phenotypic and genetic parameters for Gir cows

Particular	Value
No. of observations	2571
Average 305 day lactation milk yield (kg)	2162±13.80
Standard deviation of milk yield	670
Residual variance	230396
Genetic (marker) variance	198702
Phenotypic variance	429098
Heritability	0.46±0.039

Table 3. Correlation between corrected phenotype and GBV in Gir cows

Particular	No. of animals whose record was set missing	No. of records used	Correlation of GBV with corrected yield
Overall no missing data	0	2571	0.891
Validation set 1	515	2056	0.433
Validation set 2	514	2057	0.463
Validation set 3	514	2057	0.463
Validation set 4	514	2057	0.482
Validation set 5	514	2057	0.426
Overall pooled across validation sets	2571	—	0.452

corrected phenotype indicates that GBV can be better predictor of performance of the genotyped animals compared to their pedigree information.

Bias in estimation of Genomic Breeding Value: As seen in Fig. 1, the regression coefficient of GBV on corrected phenotype was 1.0451 (SE=0.0406). Since the regression coefficient is around 1.0, the GBVs could be considered not biased as proposed by Tsuruta *et al.* (2019). With the expansion of the reference population and the use of more number of records for better estimates of corrected yields, GBV estimation is expected to be more accurate as compared to what is observed in the present study.

The present study demonstrates that it is feasible to implement genomic selection in Gir cattle in Indian conditions using the female reference population and there is a considerable advantage in terms of increased reliabilities. The correlation coefficient of GBV and corrected phenotype is large enough to consider selecting Gir bulls for frozen semen production based on GBV instead

of the present method of dam's yield based selection. Based on average 41% reliabilities of GBV obtained for bulls, it is expected that the bulls can be selected with around 4 times more accuracy than the current method, accelerating genetic progress significantly in the Gir population that is covered under Artificial Insemination. The present study provides a piece of evidence in favor of selecting Gir bulls based on genomic breeding values using GBLUP.

ACKNOWLEDGEMENT

The authors humbly thank NDDDB management and Department of Animal Husbandry and Dairying, GoI, for funding and supporting liberally for this project. The efforts and cooperation extended by the Field staff and officials of Sabarmati Ashram Gaushala, Gir PT project for collection of accurate phenotype data and samples for the present study is highly appreciable. The authors also acknowledge the services extended by the staff of Genomics lab, NDDDB for accurate genotyping of the samples.

REFERENCES

- Campos G S, Reimann F A, Cardoso L L, Ferreira C E R, Junqueira V S, Schmidt P I, Neto J B, Yokoo M J I, Sollero B P, Boligon A A and Cardoso F F. 2018. Genomic prediction using different estimation methodology, blending and cross-validation techniques for growth traits and visual scores in Hereford and Braford cattle. *Journal of Animal Science* **96**(7): 2579–95.
- Dongre V B and Gandhi R S. 2014. Genetic and phenotypic parameters of fortnightly test day and first lactation 305-day or less milk yield in Sahiwal cattle. *International Journal of Livestock Research* **4**(3): 17–20.
- Gajjar S, Guldbrandtsen B, Su G, Nayee N G, Sahana G, Trivedi K and Lund M S. 2018. Breed-of-origin specific genomic relationship matrix improves genomic prediction accuracy in crossbred Holstein Friesian cattle in India. *Proceedings of the World Congress on Genetics Applied to Livestock Production, Volume Electronic Poster Session—Theory to Application—I, Bind 11* 2018. 11.754.
- Garrick D J, Taylor J F and Fernando R L. 2009. De-regressing estimated breeding values and weighting information for genomic regression analyses. *Genetics Selection Evolution* **41**: 55.
- Herrera L G G, El Faro L, de Albuquerque L G, Tonhati H and Machado C H C. 2008. Genetic parameters of test-day and accumulated 305 day milk yields in first lactation of Gyr cattle. *Revista Brasileira de Zootecnia* **37**: 1774–80.
- ICAR Guidelines. 2020. International Committee on Animal recording. <https://www.icar.org/index.php/icar-recording-guidelines/>
- Ledic I L, Tonhati H, Verneque R S, ElFaro L, Martinez M L, Costa C N, Pereira J C C, Fernandes L O and Albuquerque L G. 2002. Estimativas de parâmetros genéticos, fenotípicos e ambientes para as produções de leite no dia do controle e em 305 dias de lactação de vacas da raça Gir. *Revista Brasileira de Zootecnia* **31**: 1953–63.
- Lund M S, De Roos A P W, De Vries A G, Druet T, Ducrocq V, Fritz S, Guillaume F, Guldbrandtsen B, Liu Z, Reents R, Schrooten C, Seefried F and Su G. 2011. A common reference population from four European Holstein populations increases reliability of genomic predictions. *Genetics Selection*

Table 4. GBV for 305 DMY for bulls with only genotypes

Bull code	GBV	Reliability	Bull code	GBV	Reliability (%)
132	1017.42	26%	107	651.77	56
131	965.76	29%	116	640.68	26
115	918.03	21%	108	637.98	57
110	785.56	34%	130	590.28	23
112	781.34	29%	102	525.80	82
118	772.44	25%	106	509.00	72
133	769.75	25%	129	426.86	20
103	751.50	72%	104	401.19	68
117	739.01	25%	111	91.94	46
113	707.50	51%	109	3.13	52
101	702.12	69%	105	-186.55	5
114	654.76	22%	Mean	602.50	41
for bulls					

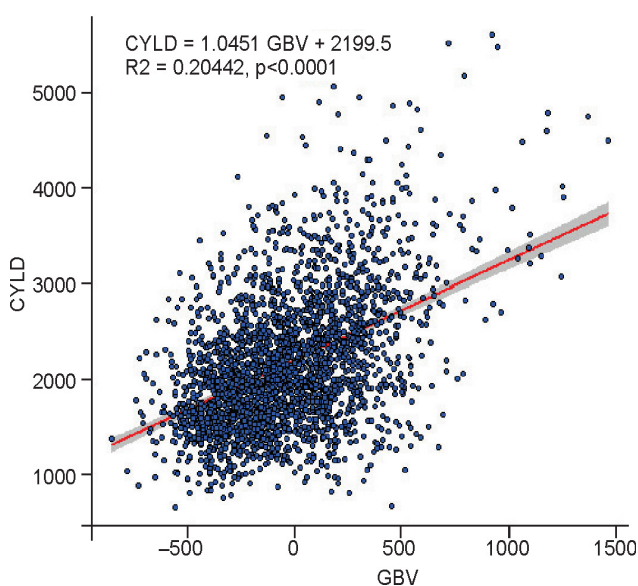


Fig. 1. Regression of CYLD on GBV. CYLD, corrected Yield; GBV, Genomic Breeding Value and R^2 = Model R^2 value.

- Evolution* **43**: 43.
- Ma P, Lund M S, Nielsen U S, Aamand G P and Su G. 2015. Single-step genomic model improved reliability and reduced the bias of genomic predictions in Danish Jersey. *Journal of Dairy Science* **98**: 9026–34.
- 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. *Proc. 10th World Congress of Genetics Applied to Livestock Production*.
- Mrode R A and Thompson R. 2005a. Linear models for the prediction of animal breeding values. CABI Publishers, Oxford, UK. 2nd ed. p. 51.
- Mrode R A and Thompson R. 2005b. Linear models for the prediction of animal breeding values. CABI Publishers, Oxford, UK. 2nd ed. p. 10.
- Mrode R, Ojango J M K, Okeyo A M and Mwacharo J M. 2019. Genomic selection and use of molecular tools in breeding programs for indigenous and crossbred cattle in developing countries: current status and future prospects. *Frontiers in Genetics* **9**: 694.
- 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* **68**(9): 84–88.
- Nayee N, Su G, Gajjar S, Sahana G, Saha S, Trivedi K, Guldbrandsen B and Lund M. 2018. Genomic prediction by single-step genomic BLUP using cow reference population in Holstein crossbred cattle in India. *Proceedings of the World Congress on Genetics Applied to Livestock Production, Volume Electronic Poster Session—Theory to Application-1*, Bind **11** 2018. 11.411.
- Peixoto M G C D, Verneque R S, Teodoro R L, Penna V M and Martinez M L. 2014. Genetic trend for milk yield in Guzerat herds participating in progeny testing and MOET nucleus schemes. *Genetics and Molecular Research* **5**(3): 454–65.
- Pryce J E, Hayes B J and Goddard M E. 2012. Genotyping dairy females can improve the reliability of genomic selection for young bulls and heifers and provide farmers with new management tools.
- R Core Team. 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. (<https://www.R-project.org/>)
- Rath D, Trivedi K R, Siddiqui M U and Nayee N G. 2017. Genomic selection of cattle and buffaloes. *Indian Dairyman* **69**(3): 60–64.
- Saha S, Nayee N, Shah H, Gajjar S, Kishore G, Gupta R O and Trivedi K R. 2020. Effect of composition and size of the reference population in genotype imputation efficiency of INDUSCHIP in HF Crossbred cattle. *Indian Journal of Dairy Science* **73**(3): 250–55.
- Santos D J A, Peixoto M G C D, Aspilcueta R R, Verneque R S, Panetto J C C and Tonhati H. 2013. Comparison of random regression models to estimate genetic parameters for milk production in Guzerat (*Bos indicus*) cows. *Genetics and Molecular Research* **12**(1): 143–53.
- Tsuruta S, Lourenco D A L, Masuda Y, Misztal I and Lawlor T J. 2019. Controlling bias in genomic breeding values for young genotyped bulls. *Journal of Dairy Science* **102**(11): 9956–70.
- VanRaden P M. 2008. Efficient methods to compute genomic predictions. *Journal of Dairy Science* **91**: 4414–23.
- VanRaden P M, Van Tassell C P, Wiggans G R, Sonstegard T S, Schnabel R D, Taylor J F and Schenkel F S. 2009. Reliability of genomic predictions for North American Holstein bulls. *Journal of Dairy Science* **92**: 16–24.