Application of ssGBLUP in Murrah buffaloes reared under small holder conditions in India

N. Nayee, S. Saha, A. Sudhakar, S. Gajjar, A. Mahajan and K. Trivedi

National Dairy Development Board, Anand, India

Introduction

Buffaloes contribute more than 50% of milk in India. Among 13 recognized buffalo breeds, Murrah is the most popular buffalo breed in India. Application of Genomic Selection (Meuwissen *et al.,* 2001) is expected to double the rate genetic progress (Schaeffer, 2006) in dairy cattle. Similarly, Genomic Selection can help increasing rate of genetic progress in Murrah buffaloes covered under AI.

Female reference populations are now gaining in popularity, especially for novel traits, such as feed efficiency, methane emissions, and detailed reproductive measures, the traits which are expensive to measure and are therefore available for only some animals in the population (Pryce et. al. 2012). As the population of buffaloes is restricted to mostly South East Asia, and these buffaloes are mostly kept under small holder conditions, large scale progeny testing programs are not implemented for buffaloes. Use of genomic selection based on mixed reference population, where both AI bulls and recorded females are genotyped, can be a practical solution in such areas.

Considering above, a study was conducted to demonstrate practicability of implementing Genomic Selection for young bull selection and its superiority over traditional methods like dam yield based selection as well as selection based on pedigree based breeding values. A single step GBLUP model was used to estimate Genomic Breeding Values.

Material and methods

Phenotype records

National Dairy Development Board, Anand, India is involved in implementing progeny testing projects for various cattle and buffalo breeds. The projects collect performance records of cattle and buffaloes reared by mostly small and marginal farmers under small holder conditions. Each farmer rears one to three milking animals. Data used in present study are collected by milk recorders engaged by milk cooperative unions of Gujarat state of India. Village level inseminators engaged by cooperatives provide AI services to farmers in villages and record pedigree information of all the daughters born by test insemination. The daughters when calve for 1st lactation, their yield is recorded for 10 monthly test days both morning and evening. All the information is recorded in Information Network for Animal Productivity and Health (INAPH- (Nayee et.al. 2016)).

68,808 first lactation test day milk yield records of daughters born to Murrah Sires were extracted from INAPH. Records from animals that were recorded for first time within 5 to 35 days post calving, having sire IDs against them and have minimum 3 test day records were used for present study. Daily milk yields between 1 to 25 Kg was only used as valid records. Total 61321 test day records belonging to 6811 daughters of 190 Murrah sires were used for final analysis.

Genotypes

A custom made SNP genotyping array on Illumina platform, BUFFCHIP, was used to genotype 3087 buffaloes and 138 bulls. 40748 SNPs that were having Minor allele Frequency (MAF) of >0.1 and genotyping rate of >90% were used for present study.

$$y_{thijkl} = A_h + HYMR_i + YS_j + OWN_k + \sum_{m=0}^{3} \emptyset_{ltm}\beta_m + \sum_{m=0}^{3} \emptyset_{ltm}u_{lm} + \sum_{m=0}^{3} \emptyset_{ltm}pe_{lm} + e_{thijkl}$$

Statistical Models for breeding value estimation

Similarly genotype information for all 3225 individuals that have >90% SNPs genotyped out of the QC passed SNPs was used for present study.

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

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

where y_{thijkl} is the test-day milk yield of cow l produced - within the hth A (age at calving class, fixed effect), the ith HYMR (herd x year of recording x month of recording, random effect) and j^{th} YS (year x season of calving, fixed effect); with the kth OWN (Owner, random effect) β_m is the fixed regression coefficient of the tth test day record on the mth order of Legendre polynomial; u_{lm} and pe_{lm} are random regression coefficients of the test day record on the mth order of Legendre polynomial for animal additive genetic and permanent environmental effects for animal k; ϕ_{ltm} is the mth Legendre polynomial of the tth 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. Owner of the animal was considered as random effect. 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.

Variance components obtained by AIREML using pedigree information were used for both BLUP and ssGBLUP breeding value estimation. The variance components and breeding values obtained under random regression model were represented in terms of 305 day EBV as per the procedure described by Mrode and Thompson (2005).

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 (**H**) instead of pedigree-based relationship matrix (**A**). The inverse of the **H** matrix is

$$H^{-1} = \begin{bmatrix} G_w^{-1} - A_1^{-1} & 0 \\ 0 & 0 \end{bmatrix} + A^{-1}$$

Here G_w was the adjusted **G** matrix, $G_w=0.8*G + 0.2*A$, in which **G** was the genomic relationship matrix and **A** was pedigree based relationship matrix. To get **G** in the same scale as **A**, the scale of **G** was adjusted so that the average of diagonals and average of off-diagonals were the same as those in **A** (Christensen *et al.,* 2012).

Purpose of genomic evaluation is to predict breeding value of individual especially bulls with high accuracy. In other words, we want to test how breeding value (or selection criteria) of a bull is correlated with average of sire's daughter's corrected yield. To evaluate predictive ability of selection criteria viz. bull's dam yield (DY), traditional BLUP model based on pedigree and ssGBLUP model additionally using genotype information, a 5 fold cross-validation method was adopted.

Estimation of predictive ability of various selection criteria for bull selection

In each validation dataset, records of all the daughters of 20% randomly selected bulls (paternal half-sib groups) were set missing and then EBF or GBV for the bull were estimated using traditional BLUP or ssGBLUP. The GBVs were compared with average of corrected phenotype of the daughters of the bulls in validation set. Only sires with more than 15 daughter records were considered for them to be included in validation process.

The corrected phenotype for a buffalo (Yc) was obtained by correcting observed phenotype for all other effects in the model except for animal genetics and permanent environment effects obtained by traditional BLUP model using full data.

The EBV or GBV obtained only based on pedigree or genotype for a bull were compared for their ability to predict corrected daughter average. This ability was calculated by comparing Pearson correlation coefficients of EBV/GBV with that of average Yc of daughters for a bull.

Of the total 190 sires having daughter records, 118 sires had >=15 daughter records. Thus in each validation data set, at random around 23 sire's daughter observations were dropped. However care was taken to keep similar number of observations in each validation data set to avoid bias.

The regression coefficient of EBV with GEBV were calculated to see bias in estimation of Genomic Breeding Values.

Table 1. Predictive ability of selection criteria.

| Validation set No.	No. of bulls	Correlation with Daughter's corrected yield			% increase in correlation for GBV	
		DY	EBV	GBV	Over DY	Over EBV
1	22	-0.03	-0.02	0.49	1733%	2550%
2	22	0.23	0.35	0.42	83%	20%
3	24	0.16	0.33	0.42	163%	27%
4	25	0.42	0.25	0.17	-60%	-32%
5	25	-0.08	0.17	0.43	638%	153%
Overall	118	0.15	0.21	0.37	147%	76%

Results and discussions

Average first lactation test day yield of the daughters of Murrah bulls used for AI in the project area was 5.85 Kg per day which translates to 1785 Kg per lactation (305 day std. lactation yield). The heritability estimate for milk yield was 0.199, total phenotypic variance was 158788 and total genetic variance was 31556 Kg^2 in this population.

The correlation between GBV and Average of daughter's corrected Yield (Yc) of individual animal is compared with the correlations obtained between DY and Yc and pedigree based EBV and Yc in table 1 for each validation data set and overall for all validation bulls.

As seen in the table, overall, the correlation between DY and Yc over all validation datasets considered together is 0.15, the correlation between EBV and Yc is 0.21 whereas correlation between GBV and Yc is 0.37. Thus there is 0.16 (147%) rise in correlation coefficient if GBV is considered for selection in place of DY. Also, the GBVs are 76% correlated to daughter production compared to pedigree based EBVs.

The comparison mimics real world scenario. Here we are interested in selecting bulls that should increase average yield of their daughters. When we select bulls, we will not have their daughter's production information. However we will be knowing bull's pedigree information and also information on their parent's performance (sire EBV and dam's Milk Yield as well as EBV). The results here indicates that the bulls can be selected more accurately if we add genomic information and use ssGBLUP over and above the traditional information while selecting a bull for semen production.

Figure 1. Regression of GEBV and EBV – 305D Milk Yield.
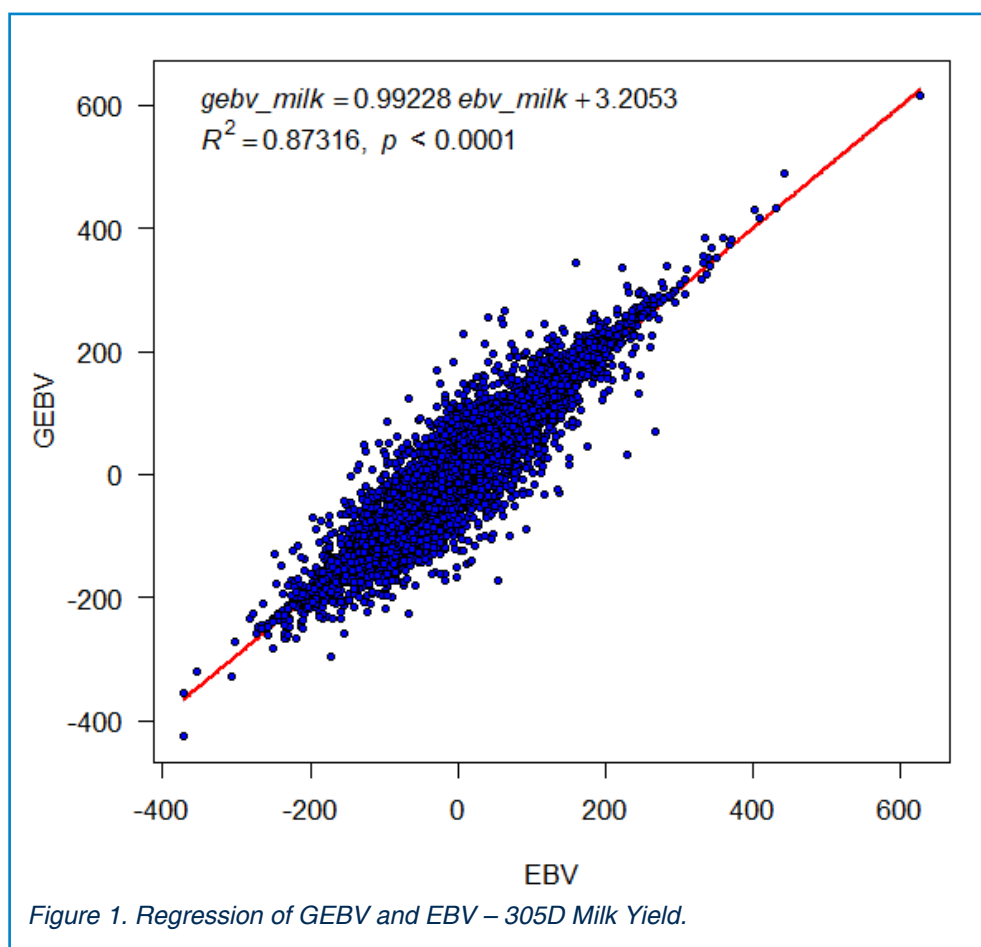
Tsuruta et.al (2019) in a simulation study demonstrated that, when males had no daughters and no siblings with phenotypes, regression coefficient of TBV with GBV was < 0.9 as GEBV for males represented inflation. This is termed as "Bias" in estimation of GEBV. To see whether there is any bias while calculating genomic breeding values compared to traditional breeding values, the GEBV were plotted against EBV and regression coefficient was calculated. Figure 1 below shows the scatter of EBV against GEBV and regression equation for the same.

Above figure clearly shows that there is no bias while estimation GEBVs through ssGBLUP as compared to traditional breeding values (EBV). Vitezica et.al. (2011) in a simulation study concluded that predictions by the multiple-step method were biased. Similarly Ma et.al. (2015) has showed that the use of ssGBLUP has reduced the bias in genomic prediction in Danish Holstein population. Current results also shows that ssGBLUP estimates are unbiased.

Regression of EBV on GEBV

Conclusion

Use of genomic information for estimation of breeding value with test day milk yield records was attempted for Murrah buffaloes performing under small holder conditions of India. Average daughter corrected phenotype was considered as reference point while comparing accuracy of DY, EBV or GBV. The GBVs obtained for validation animal with missing record were having higher correlation with corrected phenotype of the individual. The correlations were 147% higher for bulls for GBV compared to DY which was traditionally used for selection of bulls in India. The present study encourages bull selection based on GBVs obtained through ssGBLUP method to get higher genetic progress in small holder conditions.

Acknowledgement