

ACCURACY OF GENOMIC PREDICTIONS FOR MILK PRODUCTION TRAITS IN PHILIPPINE DAIRY BUFFALOES

J.R.V. Herrera^{1,3}, E.B. Flores², N. Duijvesteijn³, N. Moghaddar³ and J.H.J. van der Werf³

¹Philippine Carabao Center-University of the Philippines Los Banos, College, Laguna, 4031 Philippines

²Philippine Carabao Center National Headquarters, Muñoz Nueva Ecija, 3119 Philippines

³School of Environmental and Rural Science, University of New England, Armidale, NSW, 2351 Australia

SUMMARY

A three-fold cross-validation scheme was used to compare the accuracies of genomic prediction for milk yield (MY), fat yield (FY) and protein yield (PY) from Philippine dairy buffaloes using Genomic Best Linear Unbiased Prediction (GBLUP) and single-step GBLUP (ssGBLUP) with the accuracies based on pedigree BLUP (pBLUP). To also assess the bias of the prediction, the regression coefficient (slope) of the adjusted phenotypes on the predicted breeding values was also calculated. Results showed that genomic methods (GBLUP and ssGBLUP) provide more accurate predictions compared to pBLUP. Average GBLUP and ssGBLUP accuracies were 0.24 and 0.29, respectively, while average pBLUP accuracies (for GENO and ALL data) were 0.21 and 0.22, respectively. Slopes of the two genomic methods were also closer to one, indicating lesser bias, compared to pBLUP. Average GBLUP and ssGBLUP slopes were 0.89 and 0.84, respectively while average pBLUP (for GENO and ALL data) slopes were 0.80 and 0.54, respectively. With the higher accuracy of prediction and lesser bias, ssGBLUP will be used by PCC to identify replacement breeding bulls at a younger age.

INTRODUCTION

The Philippine Carabao Center (PCC) has put in place a genetic improvement program that includes a system of evaluating genetically superior individual animals for milk and milk component traits; maintenance of nucleus herds of dairy buffaloes (mostly Bulgarian Murrahs) as source of breeding animals and provision of frozen semen from the best riverine buffalo germplasm for artificial insemination. However, present constraints are: only ~ 1200 cows are in the nucleus herds, currently testing only 12 bulls per year, accuracies of progeny test bulls are low due to small number of daughters with lactation records, and generation interval is long for AI sires, ~ 8 yrs (Flores 2014).

The availability of the Affymetrix 90K Buffalo Genotyping Array (Affymetrix, Inc., Sta. Clara, CA) in 2013 makes it now possible to do genomic studies in the bubaline species. When the trait of interest cannot be recorded on the selection candidate, genomic selection schemes are very attractive even when the number of phenotypic records is limited, because traditional breeding requires progeny testing schemes with long generation intervals (Schaeffer 2006).

Genomic prediction studies in dairy buffaloes are very limited and were based on small data sets. Tonhati *et al.* (2016) used single-step GBLUP (ssGBLUP) to estimate the Predicted Transmitting Ability (PTA) accuracies for seven milk traits on 452 Brazilian buffaloes. Using a 5-fold cross validation, Liu *et al.* (2018) evaluated the reliability of genomic Estimated Breeding Values (GEBVs) and their correlation with EBVs for six milk production traits from 412 Italian Mediterranean buffaloes. The objective of this study was to determine the accuracy of genomic prediction and bias for milk yield (MY), fat yield (FY) and protein yield (PY) from Philippine dairy buffaloes using Genomic Best Linear Unbiased Prediction (GBLUP) and ssGBLUP compared to prediction accuracy and bias based on pedigree BLUP (pBLUP).

MATERIALS AND METHODS

The 305-day-MY,-FY,-PY data and blood samples used in this study were obtained from the PCC. All animals are housed in nucleus/institutional farms and cooperatives managed by PCC. Data collection and storage is managed by the Center's Animal Breeding and Genomics Section (ABGS). Descriptive statistics of the phenotypic data are presented in Tables 1 and 2. Number of animals with one, two and three lactation records are shown in Tables 3 and 4.

Two data sets were analysed. One only contains female buffaloes that have both phenotypes and genotypes (hereby referred to as GENO) (Table 1). Analyses done on this data were pBLUP and GBLUP. The other data set (hereby referred to as ALL) (Table 2) contains all the above animals, plus additional females with phenotype but are not genotyped. Analyses done on this data were pBLUP and ssGBLUP. A pedigree file containing 2642 animals was used for pBLUP and ssGBLUP.

Table 1. Descriptive statistics of GENO data to be used for pBLUP and GBLUP analyses

Trait*	No. of animals	No. of records	No. genotyped	Mean (kg)	Min (kg)	Max (kg)	SD (kg)
MY	904	1773	904	1573.2	103.1	3054.5	505.9
FY	856	1384	856	119.0	30.2	206.9	27.7
PY	856	1384	856	70.7	22.5	127.9	16.0

*MY-milk yield, FY-fat yield, PY- protein yield

Table 2. Descriptive statistics of ALL data to be used for pBLUP and ssGBLUP analyses

Trait*	No. of animals	No. of records	No. genotyped	Mean (kg)	Min (kg)	Max (kg)	SD (kg)
MY	1975	3821	904	1466.3	103.1	3150.9	518.0
FY	1918	3405	856	111.9	29.3	210.1	29.1
PY	1918	3405	856	66.3	19.9	128.8	17.3

*MY-milk yield, FY-fat yield, PY- protein yield

Animals were genotyped using the Axiom 90k Buffalo genotyping array. Polymorphic markers were identified using the Axiom Analysis Suite set on default settings. Additional quality control measures applied include: a SNP was removed if its minor allele frequency (MAF) is less than 0.05, is out of Hardy-Weinberg equilibrium (P-value less than 1×10^{-15}), has no genome location and is not found in the autosomes. After quality control, only 60,827 SNPs in 29 autosomes were used.

A 3-fold cross-validation scheme was used to compare accuracy of prediction and bias using GBLUP and ssGBLUP with those of pBLUP. Animals were assigned to one of 3 test sets: with 1 lactation record, 2 lactation records, and 3 lactation records (Tables 3 and 4). Phenotypes of animals in the test sets were masked, and breeding values were then estimated for each set either by pBLUP and GBLUP for the GENO data or pBLUP and ssGBLUP for ALL data using univariate repeatability animal models. All models included breed, lactation and herd-year-season (HYS) as fixed effects; and animal and permanent environmental effects as random effects. pBLUP uses a numerator relationship matrix (also known as A-matrix) based on the pedigree. The creation of the genomic relationship matrix (also known as a G-matrix) used in GBLUP and ssGBLUP is based on Van Raden (2008). The ssGBLUP (Legarra *et al.* 2014) uses a H-matrix where the G-matrix replaced the A_{22} matrix (weight of 1 on the G-matrix). Breeding values (BVs) were generated using ASReml 4.1 (Gilmour *et al.* 2008).

Accuracy of prediction was calculated as the correlation between the predicted BVs of the test set and adjusted phenotypes, which were corrected for fixed effects, divided by the square root

of the heritability of the trait, corrected for the number of lactations used in the test set. To assess the bias of the prediction, the regression coefficient (slope) of the adjusted phenotypes on the predicted breeding values was also calculated, with slopes of approximately one showing zero bias. Slopes greater than or lesser than one indicates underestimation and overestimation, respectively, of BVs.

Table 3. Number of animals (number of records) for test and training sets for MY

test set	training set	
	GENO	ALL
329 ^a (329)	575 (1444)	1646 (3492)
281 ^b (562)	623 (1211)	1694 (3259)
294 ^c (882)	610 (891)	1681 (2939)

*a,b,c-number of animals with 1,2,3 lactation records, respectively

Table 4. Number of animals (number of records) for test and training sets for FY and PY

test set	training set	
	GENO	ALL
441 ^a (441)	415 (943)	1477 (2964)
302 ^b (604)	554 (780)	1616 (2801)
113 ^c (339)	743 (1045)	1805 (3066)

*a,b,c-number of animals with 1,2,3 lactation records, respectively

RESULTS AND DISCUSSION

Results showed that genomic methods (GBLUP and ssGBLUP) provide more accurate predictions compared to pBLUP (Table 5). Average pBLUP (for GENO and ALL data) accuracies for the three traits were 0.21 and 0.22, respectively, while average GBLUP and ssGBLUP accuracies were 0.24 and 0.29, respectively. GBLUP and ssGBLUP accuracies were on average 0.03 and 0.07 higher, respectively, compared to pBLUP accuracies.

In the case of prediction bias, slopes for all methods were below one, indicating overestimation of breeding values (Table 6). However, slopes of the two genomic methods are closer to one, indicating lesser bias, compared to pBLUP slopes. Average pBLUP (for GENO and ALL data) slopes for the three traits were 0.80 and 0.54, respectively, while GBLUP and ssGBLUP slopes were 0.89 and 0.84, respectively.

CONCLUSIONS

These results are preliminary and additional animals will be genotyped to increase the sample size to have a more reliable conclusion. As of the moment, results showed that GBLUP and ssGBLUP provide more accurate predictions compared to pBLUP. Moreover, prediction bias of the two genomic methods are lesser compared to pBLUP. With the higher accuracy of prediction and lesser bias, ssGBLUP will be used by PCC to identify replacement breeding bulls at a younger age.

Table 5. Average prediction accuracy for pBLUP, GBLUP and ssGBLUP estimated from 3-fold cross-validation scheme

Trait	GENO			ALL		
	pBLUP	GBLUP	Increase in accuracy	pBLUP	ssGBLUP	Increase in accuracy
MY	0.20 ±0.04	0.28 ±0.06	0.08	0.17 ±0.02	0.30 ±0.04	0.13
FY	0.23 ±0.04	0.24 ±0.05	0.01	0.26 ±0.14	0.30 ±0.01	0.04
PY	0.20 ±0.05	0.20 ±0.05	0	0.23 ±0.14	0.26 ±0.02	0.03
Average	0.21	0.24	0.03	0.22	0.29	0.07

*MY-milk yield, FY-fat yield, PY- protein yield

Table 6. Estimated slopes calculated from breeding values from pBLUP, GBLUP and ssGBLUP

Trait	GENO		ALL	
	pBLUP	GBLUP	pBLUP	ssGBLUP
MY	0.69 ±0.39	0.85 ±0.28	0.42±0.07	0.85±0.16
FY	0.94 ±0.17	0.99 ±0.22	0.62±0.36	0.88±0.04
PY	0.76 ±0.11	0.83 ±0.34	0.57±0.38	0.79±0.10
Average	0.80	0.89	0.54	0.84

*MY-milk yield, FY-fat yield, PY- protein yield

ACKNOWLEDGEMENTS

The authors would like to acknowledge: i) PCAARRD-DOST for funding the genotyping project, ii) various coordinators, herd managers and directors of various PCC centers for making data and samples available, iii) staff of PCC's ABGS for preparing the blood samples for genotyping. JRVH would like to acknowledge PCC-SEARCA for his PhD scholarship.

REFERENCES

- Flores EB (2014) PhD Thesis. University of New England.
 Gilmour A.R., Gogel B.J., Cullis B.R., Welham S.J. and Thompson R. (2009) ASReml User Guide Release 4.0. 2015. Hemel Hempstead, UK: VSN International Ltd.
 Legarra, A., Christensen O.F., Aguilar I. and Misztal I. (2014) *Livest. Sci.* **166**: 54.
 Liu J.J., Liang A.X., Campanile G., Plastow G., Zhang C., Wang Z., Salzano A., Gasparri B., Cassandro M. and Yang L.G. (2018) *J. Dairy Sci.* **101**: 433.
 Tonhati H., Cardoso D.F., Jordan D. and Santos A. (2016). 11th World Buffalo Congress paper presentation.
 Schaeffer L. R. (2006). *J Anim. Breed. Genet.* **123**:218.
 VanRaden P.M. 2008. *J Dairy Sci* **91**:4414.