Proc. Assoc. Advmt. Anim. Breed. Genet. 23:79-82

VALIDATION OF SINGLE STEP GENOMIC BEST LINEAR UNBIASED PREDICTION IN BEEF CATTLE

M.G. Jeyaruban, P.M. Gurman, D.J. Johnston, A.A. Swan, R.G. Banks and C.J. Girard

Animal Genetics & Breeding Unit*, University of New England, Armidale, NSW, 2351 Australia

SUMMARY

This study investigated the accuracy of predicting future phenotypes of young Angus and Hereford cattle using Single-step Genomic BLUP (SSGBLUP) compared to the traditional pedigree-based BLUP evaluation (NRMBLUP). Forward cross-validation, using two comparison methods, was used to quantify the predictability of the two evaluations. For each breed, two data sets named 'full' and 'partial' were generated. The 'full' data set included all relationships, all genotypes and phenotypes of animals born up to November 2018. For 'partial' data sets, phenotypes of animals born after December 2014 were removed and the data for animals removed after December 2014 were used as the 'validation data set'. SSGBLUP and NRMBLUP analyses were performed separately for the full and partial data sets and EBVs were predicted for animals in the validation data set. In Method 1, R squared values (R^2), regression coefficients (REG) and adjusted correlation (ACOR), between pre-corrected phenotypes and predicted EBVs were compared. In Method 2, correlation ratios between EBVs from full and partial evaluations were estimated to calculate the increase in predictability between the SSGBLUP and NRMBLUP. The estimated R^2, REG and ACOR using SSGBLUP were higher than those from NRMBLUP. A similar pattern was observed for correlation ratios from Method 2. The increase in ability to predict future phenotypes using Method 1 ranged from 30 to 50% and 10 to 36% for genotyped and 2 to 4% and 1 to 2 % for non-genotyped Angus and Hereford cattle, respectively. Using Method 2, the ability to predict future phenotypes ranged from 22 to 40% and 6 to 28% for genotyped and 1 to 2% and 0.5 to 1 % for non-genotyped Angus and Hereford cattle in the validation set, respectively. This study showed that there was an increase in the accuracy to predict future performance from SSGBLUP compared to NRMBLUP in Angus and Hereford cattle. The increase in predictive ability varied according to the heritability of a trait, the number of phenotypes and genotypes included in the evaluation and whether the animals were genotyped or not in the evaluation.

INTRODUCTION

BREEDPLAN analytical software developed by the Animal Genetics and Breeding Unit (AGBU) is used for genetic evaluation of beef cattle using best linear unbiased prediction (BLUP) (Graser *et al*. 2005). Prior to 2012, EBVs were predicted using pedigree based BLUP models (NRMBLUP). Since 2012, the BREEDPLAN software has been upgraded to include a range of DNA marker-based predictions. With the development of 50K micro arrays in 2008, genome wide SNP based prediction called Molecular Breeding Values 'MBVs' were included using a post-BLUP blending method. This meant that genotype information did not influence EBVs of pedigree-only animals. Furthermore, blending of MBVs into existing EBVs is sensitive to various biases which can be complicated to eliminate. These biases are mostly overcome by implementing Single-step Genomic BLUP (SSGBLUP). In SSGBLUP, information from pedigree, phenotypes and genotypes are jointly used. SSGBLUP combines the genomic relationship matrix (G) for genotyped animals with the pedigree-based relationship (A) for non-genotyped animals (Christensen and Lund 2010). Therefore, SSGBLUP is expect to produce more accurate EBVs for animals with genotypes than NRMBLUP.

Since 2017, SSGBLUP has been implemented for the genetic evaluation and use in Angus, Brahman,

Hereford and Wagyu breeds in Australia (Johnston *et al*. 2018). An important implementation step is to quantify the extent of increase in predictability of SSGBLUP over NRMBLUP. A forward cross validation method proposed by Legarra and Reverter (2018) was used in this study to compare the predictability of SSGBLUP and NRMBLUP. Predictability is defined as how well the EBVs predict observed performance.

MATERIALS AND METHODS

Data used in this study were submitted by Angus and Hereford breeders and their breed societies for use in the November 2018 BREEDPLAN evaluation. Data included 600 day weight (FWT), scan eye muscle area in heifers (HEMA) and bulls (BEMA), and scrotal circumference (SC). Univariate analyses were performed for each trait using models described by Graser *et al*. (2005). Table 1 summarises the number of animals with phenotypes and genotypes for each trait across the two breeds.

Forward cross-validation described by Legarra and Reverter (2018) was used to compare the predictability of SSGBLUP and NRMBLUP. For each breed, two data sets named 'full' and 'partial' were generated. The Full data set included all relationships, genotypes and phenotypes of animals born up to November 2018. For the 'partial' data set, phenotypes of animals born after December 2014 were removed and the data for animals removed were used as the 'validation data set'. The SSGBLUP and NRMBLUP analysis were performed separately for the full and partial data sets and EBVs were predicted for animals in the validation data set. A strict criteria was implemented to ensure good convergence.

Two approaches were used to assess the ability to predict the future phenotypes in the validation data set using EBVs estimated from the partial data. In approach 1, adjusted phenotypes in the 'validation set' were regressed against the EBVs from partial analyses of SSGBLUP (SEBVp) and NRMBLUP (NEBVp) within their respective contemporary group. R-squared values (R^2) and regression coefficients (REG) were estimated. Accuracy of prediction was calculated as a correlation between adjusted phenotypes and SEBVp or NEBVp and the correlations were adjusted for by dividing by the square root of the heritability (ACOR). The increase in ability to predict future genotypes (PRED1) of young Angus and Hereford cattle was assessed as a ratio between ACOR of SSGBLUP and NRMBLUP.

In approach 2, the Pearson correlations between EBVs using full ($\hat{U}_f$) and partial ($\hat{U}_p$) for animals in the validation data set were computed as per the formula given below from Legarra and Reverter (2018),

$$\hat{\rho}_{f,p} = \frac{\frac{1}{n}\left(\hat{U}_p - \bar{\hat{U}}_p\right)'\left(\hat{U}_f - \bar{\hat{U}}_f\right)}{\sqrt{\frac{1}{n}\left(\hat{U}_f - \bar{\hat{U}}_f\right)'\left(\hat{U}_f - \bar{\hat{U}}_f\right)\frac{1}{n}\left(\hat{U}_p - \bar{\hat{U}}_p\right)'\left(\hat{U}_p - \bar{\hat{U}}_p\right)}}$$

Where n is the number of animals in validation set, $\hat{U}_f$ are the full EBVs, $\bar{\hat{U}}_f$ the mean of the full EBVs, $\hat{U}_p$ are the partial EBVs, $\bar{\hat{U}}_p$ the mean of the partial EBVs. Legarra and Reverter (2018) showed that $\hat{\rho}_{f,p}$ was equal to the ratio of accuracy of partial (acc_p) and accuracy of full (acc_f) of SSGBLUP or NRMBLUP. This was modified to get the increase in predictive ability (PRED2) of SSGBLUP by calculating the ratio between acc_p of SSGBLUP and acc_p of NRMBLUP as per the equation given below,

$$PRED2 = ((corr(SEBV_p, SEBV_f) / corr(NEBV_p, SEBV_f)) - 1)*100$$

RESULTS AND DISCUSSION

The data used in the methods is summarised in Table 1. In addition to the number of records given in Table 1, Angus and Hereford had 55999 and 10,971 genotyped animals, respectively, in the full and partial analyses. The number of animals with phenotypes and genotypes in the validation data for each trait ranged from 11,455 to 14,162 for Angus and 1,507 and 3,908 for Hereford. Heritabilities used in the prediction for FWT, HEMA, BEMA and SC for Angus were 0.38, 0.26, 0.24 and 0.39, respectively and for Hereford were 0.31, 0.24, 0.23 and 0.44, respectively.

Table 1. Summary of data used in the prediction

Trait	Angus				Hereford			
	Number of records		[1] Number in validation set		Number of records		Number in validation set	
	Full	Partial	Geno	Non	Full	Partial	Geno	Non
FWT	801,991	673,969	14,162	100,076	514,345	464,703	3,569	40,959
HEMA	368,832	289,344	11,455	68,033	128,810	104,557	1,507	22,746
BEMA	406,378	316,707	13,546	76,125	177,311	148,585	3,908	24,818
SC	335,437	256,152	12,404	66,881	133,276	108,026	3,432	21,818

[1] 'Geno': genotyped animals; 'Non': non-genotyped animals.

Genotyped animals. For genotyped animals in the validation set, estimated R^2, REG, ACOR and PRED1 from Method 1 and the PRED2 from Method 2 are given in Table 2. Using Method 1 for Angus, estimated R^2 values ranged from 0.11 to 0.22 for SSGBLUP and from 0.06 to 0.12 for NRMBLUP. Estimated R^2 values were higher for SSGBLUP than NRMBLUP for all traits. The estimated REG using SSGBLUP were also higher than those using NRMBLUP. However, the estimated REG was higher than 1 for SSGBLUP indicating that EBVs were under-predicted for SSGBLUP. The ACOR ranged from 0.67 to 0.79 for SSGBLUP and 0.48 to 0.59 using NRMBLUP. Adjusted correlations were higher for SSGBLUP than for NRMBLUP for all traits. The PRED1 ranged from 30 to 53%.

For Hereford, estimated R^2 values ranged from 0.08 to 0.17 for SSGBLUP and from 0.05 to 0.13 for NRMBLUP. As observed for Angus, estimated R^2 values were higher for SSGBLUP than NRMBLUP for all traits. Estimated REG using SSGBLUP were also higher than those using NRMBLUP. The ACOR ranged from 0.56 to 0.62 for SSGBLUP and 0.41 to 0.54 using NRMBLUP. The ACOR were higher for SSGBLUP than for NRMBLUP for all traits. PRED1 ranged from 10 to 36%.

Using Method 2 for Angus, PRED2 ranged from 23 to 50%, respectively. For Hereford PRED2 ranged from 6 and 28%, respectively.

Non-genotyped animals. Using Method 1 for Angus, changes in the estimated R^2, REG, ACOR and PRED1 between SSGBLUP and NRMBLUP were similar to those observed for genotyped animals. However, increases were lower than the values observed for genotyped animals, with results for PRED1 ranging from 3 to 6%. A similar pattern was observed for Hereford where PRED1 ranged from 1to 3%.

Using Method 2 for Angus, similar to genotyped animals, the predictability of SSGBLUP was higher than for NRMBLUP for all traits. The PRED2 ranged from 2 to 5%. For Hereford, PRED2 ranged from 1 to 2%.

Results for both procedures showed higher predictability for SSGBLUP as compared to NRMBLUP. However, estimated regression slopes greater than one indicate that cross-validation using Method 1 may be biased due to errors in adjusting the fixed effects, selection and the heritability used in the evaluation (Legarra and Reverter 2018). As expected, the advantage in predictability of both procedures using SSGBLUP (compared to NRMBLUP) was higher for genotyped animals than non-genotyped

animals. Furthermore, Angus, with a higher number of phenotypes and genotypes animals gave higher PRED1 and PRED2 for all traits than in Hereford. When the genotyped and non-genotyped animals were combined, the increase in predictability estimated for SSGBLUP in this study was lower than the range (25 to 36%) published by Lourenco et al (2018) for Angus cattle in USA. Lourenco et al (2018) had more animals with records and genotypes than the numbers available in this study.

Table 2. Estimated R squared (R^2), regression coefficient (REG) and adjusted correlations (ACOR) from Method 1and increase in predictability from Method 1 (PRED 1 %) and Method 2 (PRED 2 %) for SSGBLUP over NRMBLUP for genotyped animals

Trait	Method 1						Methods	
	SSGBLUP			NRMBLUP			1	2
	R^2	REG	ACOR	R^2	REG	ACOR	PRED1	PRED2
				Angus				
FWT	0.22	1.16±0.02	0.79	0.12	1.07±0.02	0.58	36	28
HEMA	0.15	1.07±0.02	0.77	0.09	1.08±0.03	0.59	30	23
BEMA	0.11	1.04±0.03	0.67	0.06	0.93±0.03	0.48	39	26
SC	0.22	1.22±0.02	0.75	0.09	1.09±0.03	0.49	53	50
				Hereford				
FWT	0.10	1.11±0.05	0.56	0.05	0.93±0.06	0.41	36	28
HEMA	0.08	0.99±0.09	0.56	0.06	0.92±0.09	0.51	10	6
BEMA	0.08	1.17±0.06	0.61	0.06	1.05±0.06	0.51	19	13
SC	0.17	1.07±0.04	0.62	0.13	0.99±0.04	0.54	15	12

CONCLUSIONS

Ability to predict the future phenotypes of both genotyped and non-genotyped animals was higher for SSGBLUP compared to NRMBLUP. Both methods of comparisons yielded very similar results. Furthermore, ability to predict the future phenotypes was influenced by the number of genotyped animals in the evaluation and the heritability of the trait used. Higher numbers of genotyped animals and higher heritability resulted in increased predictability for SSGBLUP.

ACKNOWLEDGEMENT

The authors would like to thank the Meat and Livestock Australia (MLA) for their financial support through L.GEN.0174 and Angus Australia, Hereford Australia and Agricultural Business Research Institute, UNE for providing data for this study.

REFERENCES

Christensen O.F. and Lund M.S. (2010). *Genet. Sel. Evol*. **42:** 2.

Graser H-U., Tier B., Johnston D.J. and Barwick S.A. (2005) *Aust. J. Exp. Agric*. **45**:913.

Johnston D.J, Ferdosi M.H, Connors N.K, Boerner V, Cook J, Girard C.J, Swan A.A, Tier B. (2018). In *Proc. 11th World Cong. Genet. Appl. Livest. Prod.* **11:**.269.

Legarra A. and Reverter A. (2017). *Proc. Assoc. Advmt. Anim. Breed. Genet.* **22**:73

Legarra A. and Reverter A. (2018) *Genet. Sel. Evol*. **50:** 53**.**

Lourenco D.A.L., Tsuruta S., Fragomeni B.O., Masuda Y., Aguilar I., Legarra A., Millier S., Moser D. and Misztal I. (2018). In *Proc. 11th World Cong. Genet. Appl. Livest. Prod.* **11:** 495.