

## ACCURACY OF GENOMIC PREDICTIONS IN NELLORE BEEF CATTLE

R. Carvalheiro<sup>1</sup>, J. C. McEwan<sup>2\*</sup>, H.H.R. Neves<sup>1</sup>, Y.T. Utsunomiya<sup>1</sup>, A.M. Pérez O'Brien<sup>3</sup>, S. A. Boison<sup>3</sup>, J. Sölkner<sup>3</sup>, F.S. Schenkel<sup>4</sup>, C.P. Van Tassell<sup>5</sup>, T.S. Sonstegard<sup>5</sup> and J.F. Garcia<sup>6</sup>

<sup>1</sup>UNESP - Univ Estadual Paulista. Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal, São Paulo, Brazil. <sup>2</sup>AgResearch, Invermay, New Zealand. \*Email: john.mcewan@agresearch.co.nz. <sup>3</sup>BOKU - University of Natural Resources and Life Sciences, Vienna, Austria. <sup>4</sup>Centre for Genetic Improvement of Livestock, University of Guelph, Canada. <sup>5</sup>ARS-USDA - Agricultural Research Service - United States Department of Agriculture. <sup>6</sup>UNESP - Univ Estadual Paulista. Faculdade de Medicina Veterinária de Araçatuba, São Paulo, Brazil.

### SUMMARY

This study was carried out to assess the quality of genomic predictions in a Nellore beef cattle population, for 14 growth, carcass composition and reproduction traits, evaluated either at weaning or yearling. A forward prediction scheme was applied, so that information on a set of older animals (bulls and cows with accurate proofs in 2007) was employed to derive genomic prediction equations, while information on younger bulls (2012 proofs) was considered for validation purposes. The validation accuracies of genomic predictions averaged 0.47, consistent with the expectation for such statistics. Accuracies for two selection indexes including either weaning traits (WI) or both weaning and yearling traits (FI) were 0.44 and 0.58, respectively. For younger animals with own performance, genomic predictions increased by 10% (on average) the individual accuracies for both WI and FI.

### INTRODUCTION

Since Meuwissen *et al.* (2001) showed that accurate predictions of the genetic merit of selection candidates could be obtained by using information from dense marker panels, the potential for incorporation of genomic selection (GS) in breeding programs has been an active research topic. It has been argued that GS schemes could increase the rates of genetic gain substantially, via increased accuracy of estimated breeding values and reduced generation intervals. It could also reduce costs for breeding organizations (Schaeffer 2006).

While cost-effectiveness of GS was evident for dairy cattle, justifying its current application in several countries, GS is still not employed by most beef cattle breeding programs. This is explained in part by the fact that GS methodology provides more modest benefit in beef cattle, especially for traits routinely recorded in early life, due to the differences in population structure, the smaller amount of information available to derive accurate and cost-effective prediction equations and the shorter generation intervals often found in beef cattle.

Despite these constraints, some potential for application of GS in *Bos indicus* (Nellore) beef cattle is envisaged. This breed plays an important role in beef production in Brazil. The large number of animals routinely recorded by breeding organizations and the fact that bulls are progeny-tested relatively late in life could justify GS, especially for important traits in which the progress achieved by conventional selection is currently limited. The aim of this study was to assess the quality of genomic predictions in a Nellore beef cattle population.

### METHODS

**Data.** Phenotypic information consisted of EBVs obtained from routine evaluations of Conexão Delta G, a commercial beef cattle breeding program kept by an alliance of producers distributed across 12 states of Brazil. EBVs were based on records of 1,168,792 animals, collected between 1983 and 2012. Fourteen traits were considered, including weight traits, scrotal

circumference and carcass traits evaluated through visual scores (Table 1), as well as two selection indexes, including either weaning traits (WI) or both weaning and yearling traits (FI).

A total of 2,275 animals (influential bulls and cows) were genotyped with the Illumina® Bovine HD panel. Quality control of genotypes was carried out through an iterative process using the following SNP selection criteria: call rate > 0.98, minor allele frequency > 0.02, p-value for Hardy-Weinberg equilibrium test >  $10^{-5}$ . SNPs meeting these criteria were further screened to interrogate their correlation with other syntenic SNPs located within a 100 marker window, allowing only one marker from each pair of highly correlated SNPs ( $r^2 > 0.995$ ) to remain in the SNP dataset. Finally, samples with call rates < 0.90 were excluded from the analysis. The process was repeated until neither SNP nor samples were excluded, resulting in a final dataset of 995 bulls and 1,266 cows with 311,359 SNP called.

**Study design.** A forward prediction scheme was applied to compute genomic predictions. For each trait, the training population was composed of all genotyped animals with EBV's accuracy  $\geq 0.50$  in 2007, while younger bulls were included in the validation set (average accuracy of EBV of 0.88 in 2012). EBVs obtained in 2007 were considered as the response variable in model training, while EBVs obtained in the 2012 evaluation were used in the validation step, so information on validation animals (and their descendants) did not contribute to marker effect estimation.

Genomic predictions (DGV) were obtained using genomic BLUP (GBLUP), implemented using *gebv* software (Sargolzaei *et al.*, 2009). GBLUP equations included a modified genomic relationship matrix ( $G^*$ ), obtained as  $G^* = 0.8G + 0.2A$ , where  $G$  is the genomic relationship matrix computed similarly as in VanRaden *et al.* (2009), using observed allele frequencies.  $A$  is the regular numerator relationship matrix. The weights on  $G$  and  $A$  matrices were chosen based on previous analyses, which indicated better predictive ability with the adopted weights. A weighted analysis was conducted to account for differences in accuracies of EBVs of the training set. A diagonal matrix  $R$  was included in the GBLUP equations, whose entries were equal to  $R_{ii} = (1/Rel_i) - 1$ , where  $Rel_i$  is the reliability of the  $i^{\text{th}}$  EBV.

In order to explore the importance of relatedness in the predictive ability, an alternative version of GBLUP was fitted. A modified genomic relationship matrix was computed as  $G^{**} = 0.999G_1 + 0.001A$ , where  $G_1$  is a genomic relationship matrix built considering exclusively the markers in BTA1, and  $A$  as described previously. The reasoning of this strategy is that information in one chromosome is expected to capture relationships, albeit this is subject to the assumption that LD is consistent across the genome, but is unlikely to contain all QTL (Daetwyler *et al.*, 2012). The relative influence of relationships on the accuracy of genomic predictions were assessed by contrasting the accuracies obtained using either  $G^*$  or  $G^{**}$  in GBLUP equations.

**Analysis of results.** The Pearson's correlation between DGV and EBV of the validation animals ( $r_{EBV,DGV}$ ) was considered as a proxy for the accuracy of prediction. In order to evaluate the amount of uncertainty about such correlations, 95% confidence intervals for  $r_{EBV,DGV}$  were obtained using Fisher's Z transformation. The intercept ( $b_0$ ) and the slope ( $b_1$ ) of the regression of EBV on DGV were evaluated to describe bias and inflation of the predictions, respectively. The expected values for accuracy of prediction  $cor(g,\hat{g})$  were obtained using a deterministic formula (Daetwyler *et al.*, 2010), by assuming a value of 120 for the effective population size and that the average reliability of the EBVs would be equal to the heritability of the pseudo-phenotypes available for model training.

## RESULTS AND DISCUSSION

The accuracies of the genomic predictions varied from 0.24 (birth weight and gestation length) to 0.68 (finishing precocity at yearling), with an average of 0.47 (Table 1), although the amount of uncertainty about such estimates was usually large, as a consequence of the relatively small

validation sets in this study. The expected values for these statistics, calculated using the formula proposed by Daetwyler *et al.* (2010), ranged from 0.48 (scrotal circumference) to 0.58 (weight gain from birth to weaning). In previous analyses using just sire genotypes, the use of deregressed EBVs (Garrick *et al.*, 2009) as pseudo phenotypes generated very similar results as using EBVs, while different statistical methods (e.g. BayesC and Bayesian LASSO) generated results similar to those of GBLUP (data not shown).

Table 1. Summary statistics<sup>2</sup> of genomic predictions in Nellore beef cattle

Trait <sup>1</sup>	Ntrain	accT	Nval	accV	rEBV,DGV (95% CI)	rPEV	cor(g,ĝ)	BTA1	b <sub>0</sub> (%)	b <sub>1</sub>
WG	1,757	0.70	179	0.90	0.38 (0.25,0.50)	0.53	0.58	0.33	-23.2	0.80
CW	1,688	0.69	179	0.90	0.40 (0.27,0.51)	0.53	0.56	0.38	-9.3	1.12
PW	1,689	0.69	179	0.90	0.53 (0.41,0.62)	0.53	0.56	0.50	-21.0	0.99
MW	1,688	0.69	179	0.90	0.54 (0.42,0.63)	0.53	0.57	0.51	-20.5	1.02
NW	1,682	0.69	179	0.90	0.47 (0.35,0.58)	0.53	0.56	0.42	26.6	0.92
PWG	1,655	0.69	108	0.88	0.53 (0.38,0.65)	0.53	0.56	0.45	-19.6	0.95
CY	1,628	0.68	108	0.88	0.45 (0.29,0.59)	0.53	0.55	0.39	8.5	1.05
PY	1,628	0.68	108	0.88	0.68 (0.56,0.77)	0.53	0.55	0.63	-17.6	1.12
MY	1,628	0.68	108	0.88	0.64 (0.51,0.74)	0.53	0.55	0.60	-16.3	1.11
NY	1,625	0.68	108	0.88	0.47 (0.31,0.61)	0.53	0.55	0.43	24.7	0.99
SC1	1,089	0.70	56	0.83	0.46 (0.23,0.65)	0.50	0.49	0.39	-43.0	0.72
SC2	1,078	0.70	61	0.83	0.49 (0.28,0.66)	0.50	0.48	0.45	-33.9	0.55
BW	1,686	0.67	124	0.88	0.24 (0.07,0.40)	0.51	0.55	0.21	-3.4	0.38
GL	1,339	0.67	76	0.90	0.24 (0.01,0.44)	0.47	0.51	0.17	-15.5	0.67
WI	1,692	0.70	179	0.90	0.44 (0.31,0.55)	0.53	0.57	0.40	-25.4	0.90
FI	1,413	0.70	140	0.88	0.58 (0.46,0.68)	0.52	0.53	0.52	-21.0	0.92
Overall	1,560	0.69	129	0.88	0.47(0.32,0.60)	0.52	0.55	0.44	-13.1	0.89

<sup>1</sup>WG: weight gain from birth to weaning (about 205 days of age); CW, PW, MW, NW: visual scores taken at weaning for carcass conformation, finishing precocity, muscling and navel, respectively; PWG: weight gain from weaning to yearling (about 550 days of age); CY, PY, MY, NY: visual scores taken at yearling for carcass conformation, finishing precocity, muscling and navel, respectively; SC1 and SC2: scrotal circumference adjusted for age and for age and weight, respectively. BW: birth weight; GL: gestation length; WI: weaning index, composed by traits evaluated at weaning; FI: final index, composed by traits evaluated at weaning and yearling (FI).

<sup>2</sup>Ntrain (Nval): number of animals in the training (validation) set; accT (accV): average accuracy of EBVs of training (validation) animals; rEBV,DGV: validation accuracy, i.e. Pearson's correlation between EBV and genomic prediction (DGV); (IC95%): limits of the 95% confidence interval for rEBV,DGV; rPEV: average of the individual accuracies of DGV (obtained using elements of the inverse of the coefficient matrix); cor(g,ĝ): expected accuracy of prediction according to a deterministic formula (Daetwyler *et al.*, 2010). BTA1: validation accuracy obtained using only marker information from BTA1; b<sub>0</sub> and b<sub>1</sub>: intercept and slope of the regression of EBV on DGV, respectively (b<sub>0</sub> is expressed relatively to the standard deviation of the EBVs for each trait, in %).

As a general rule, the validation accuracies matched well with the expected values (Table 1), although substantial departure from expectation was observed for birth weight and gestation length (accuracies about 50% lower than expected). Also, higher than expected accuracies were found for finishing precocity (PY) and muscling (MY) scores at yearling. This result seems to confirm the findings of Carvalho *et al.* (2012), who also obtained higher than expected accuracies for these traits, after fitting GBLUP with a smaller training set, and suggested that these traits were affected by the presence of genotype stratification associated to differences of the within group EBV means. Results of a principal component analysis based on the genomic relationship matrix

evidenced the existence of two subgroups of the sampled population. Further inspection corroborate the hypothesis that PY and MY are both affected by genotype stratification associated to a large difference of EBV subgroup means for these traits (data not shown).

The accuracies for two selection indexes, including just weaning traits (WI) or both weaning and yearling traits (FI) were about 0.44 and 0.58, respectively. Overall, the DGVs were slightly inflated (slope  $\sim 0.89$ ) and overestimated (intercept  $< 0$ ), although there was some variation in this pattern across traits. The most inflated estimates were verified for BW, SC2 and GL, while the predictions for SC1 and SC2 were those for which the overall mean of DGVs departed most from that of EBVs.

The averages of the individual accuracies of DGV (computed using the estimated Prediction Error Variance - rPEV) were more consistent across traits (average  $\sim 0.52$ ) than rEBV, DGV and also matched more closely to the expected values for accuracy of prediction (Table 1). For younger animals with own performance recorded, the genomic predictions increased by 10% (on average) the individual accuracies for both WI and FI, when compared with the accuracies of traditional EBVs (data not shown).

This increase in accuracy was found to be associated to the relatedness of each young animal to the training set, so that individual accuracies of DGV for animals with a sire in the training set were increased by 20% (on average), compared to traditional evaluations. Similar association was also verified by Clark *et al.* (2012), who found strong correlations between different measures of relatedness to the training set and rPEV.

When averaged across traits, about 90% of the accuracy obtained when considering information from all chromosomes was recovered using only the information from BTA1, which reinforces the importance of relatedness (population structure) contribution to the accuracy of genomic predictions in this population, as is unlikely that all the QTLs affecting the studied traits are located on BTA1.

While the present study focused on routinely recorded traits, larger benefit is expected for traits in which the genetic progress achieved through conventional selection is limited (e.g. sexual precocity, productive longevity and meat quality), and this should be the subject of future research.

## CONCLUSIONS

The current genomic predictions matched reasonably well the expectations and increased by 10% (on average) the individual accuracies of younger animals with own performance for two selection indexes, including either weaning traits or both weaning and yearling traits.

## REFERENCES

- Carvalho R., Perez O'Brien A.M., Neves H.H.R., Sölkner J., Utsunomiya Y.T., and Garcia J.F. (2012) *Proc of 63rd EAAP Annual Meeting*, EAAP, Bratislava, Slovakia.
- Clark S.A., Hickey J.M., Daetwyler H.D., and van der Werf J.H.J. (2012) *Gen. Sel. Evol.* **44**:4.
- Daetwyler H.D., Kemper K.E., van der Werf J.H.J., and Hayes B.J. (2012) *J. Anim. Sci.* **90**:3375.
- Daetwyler H.D., Pong-Wong R., Villanueva B., and Woolliams J.A. (2010) *Genetics* **185**:1021.
- Garrick D., Taylor J.F., and Fernando R.L. (2009) *Gen. Sel. Evol.* **41**:55.
- Meuwissen T.H.E., Hayes B.J., and Goddard M.E. (2001) *Genetics* **157**:1819.
- Sargolzaei M., Schenkel F.S., and VanRaden P.M. (2009) In 'Technical report to the Dairy Cattle Breeding and Genetics Committee', University of Guelph, Guelph. Available at <http://cgil.uoguelph.ca/dcbgc/Agenda0910/GEBV.pdf>
- Schaeffer L.R. (2006) *J. Anim. Breed. Genet.* **123**:218.
- VanRaden P.M., Van Tassel C.P., Wiggans G.R., Sonstegard T.S., Schnabel R.D., Taylor J.F., and Schenkel F.S. (2009) *J. Dairy Sci.* **92**:16.