Proc. Assoc. Advmt. Anim. Breed. Genet. 20:211-214

ACCURACY OF IGENITY DIRECT GENOMIC VALUES IN AUSTRALIAN ANGUS

V. Boerner and D. J. Johnston

Animal Genetics and Breeding Unit[1], University of New England
Armidale, 2351, NSW, Australia

SUMMARY

The quality of Igenity[2] direct genomic values (GEBVs) derived by two different prediction procedures for 12 traits of 1032 Angus bulls was estimated as the genetic correlation to their phenotypic target traits. In addition, the effect of a decreasing genetic relationship between validation and training population was inferred by subdividing the set of 1032 GEBVs accordingly. Genetic correlations estimated were medium to high even when all training individuals were excluded from the analysis, and well in line with those already published. Thus blending Australian Angus breeding values with Igenity GEBVs can be beneficial for breeders.

INTRODUCTION

GEBVs, calculated by applying previously derived prediction equations to known SNP genotypes, are available for Australian Angus beef cattle from at least two commericial suppliers (www.pfizer.com,www.igenity.com). The value of this additional information to breeders depends on the genetic correlation (accuracy, r_g) to their phenotypic target traits. An analysis of GEBVs from both providers by the American Angus Association found such correlations between 0.65 and 0.29 depending on the trait (Northcutt 2011). Evaluations of Pfizer Molecular Value Predictions done in the Australian Angus population resulted in r_gs between 0.45 and 0.2 (Johnston *et al.* 2010). For Igenity molecular breeding values r_gs of 0.8 for scan intra-muscular fat content of yearling bulls and 0.38 for carcase marbling score were found in American Angus (MacNeil *et al.* 2010).

This paper presents results of a correlation analysis of Angus GEBVs supplied by Ingenity for 12 different traits of which phenotypic target traits are also recognised in the usual breeding value estimation for this breed. As the training individuals were part of the GEBV set, and for each trait GEBVs from two different prediction procedures were supplied, we have also analysed the effect of an increasing genetic distance between training and validation population on r_gs and how differently derived prediction equations affect these correlations.

METHODS

The accuracy of GEBVs was determined as the genetic correlation between GEBVs (modelled as traits) and the corresponding phenotypic target traits estimated using REML or Gibbs sampling in a bi-variate approach.

Direct genomic values of 1032 Angus bulls for birth weight (d.BWD), 200 day weight direct (d.WWD), 200 day weight maternal (d.WWM), 400 day weight (d.YWD), mature cow weight (d.MCW), scrotal circumference (d.SC), carcase weight (d.CWT), carcase intra-muscular fat content (d.CIM), carcase ribeye area (d.CEA), direct calving ease (d.CED), maternal calving ease (d.CEM) and docility (d.DOC), predicted by two different procedures (50K3 and 50KGB), were supplied by Igenity (http://www.igenity.com/). For both

[1]A joint venture of the NSW Department of Primary Industry and the University of New England
[2]Igenity is a registered trademark of Neogen Corporation.

prediction procedures, the underlying genotype was obtained from an Illumina 50K Bead Chip, but 50K3 GEBVs were calculated from prediction equations derived on 392 SNP individually chosen for each trait, whereas 50KGB GEBVs were calculated from prediction equations derived in a GBLUP approach. Across prediction procedures, GEBVs were supplied in two sets, A: 736 GEBVs of American Angus individuals genotyped in the US and used in the Igenity training set, and B: 355 GEBVs of Australian Angus individuals genotyped in Australia. To analyse the effect of an increasing genetic distance (decreasing genetic relationship) between training and validation population, sets A and B were united and then subdivided as follows: FULL: all genotyped individuals of set A and B (n=1032). AU: only set B individuals (n=345). AUS: as AU, but direct progeny of individuals in set A were excluded (n=188).

Phenotypic traits included in the analysis were birth weight (p.BWD, n=248562), 200 day weight (p.WW, n=234087), 400 day weight (p.YWD, n=156893), mature cow weight (p.MCW, n=90795), carcase weight (p.CWT, n=4535), carcase intramuscular fat percentage (p.CIM, n=3434), carcase eye muscle area (p.CEA, n=2732), scrotal circumference (p.SC, n=159171), calving ease (p.CE, n=161172) and docility (p.DOC, n=13050). Records were obtained from the Australian Angus Society database. Note that in Australian Angus phenotypic calving ease and docility are recorded as calving difficulty and wildness, respectively, so negative correlations were expected for these traits.

The linear model was $y = Xb + Z_d u_d + Z_m u_m + Z_q p_q + Z_r p_r + e$, where y is a vector phenotypes, b is a vector of fixed effects, u_d is a vector of random direct genetic effects, u_m is a vector of random maternal genetic effects, p_q is a vector of random maternal environmental effects, p_r is a vector of random permanent environmental effects and e is a vector of random residual effects. X, Z_d, Z_m, Z_q and Z_r are incidence matrices linking the effects to their respective phenotypes. Note that for GEBVs, X is a vector of ones. It was assumed that traits $\sim N(Xb, Z_d A Z_d' \sigma_d^2 + Z_m A Z_m' \sigma_m^2 + Z_d A Z_m' \sigma_{d,m} + Z_q I Z_q' \sigma_q^2 + Z_r I Z_r' \sigma_r^2 + I\sigma_e^2)$, where A is the numerator relationship matrix built from a pedigree such that every individual with an observation had at least, if available, three generations of ancestors and I is an identity matrix. u_m and p_q were modelled only for p.BWD, p.WW, p.YWD and p.CE, and p_r only for p.MCW.

The software used to estimate parameters of continuously distributed phenotypic traits and their related GEBVs was WOMBAT(Meyer 2007). Parameters of categorically distributed phenotypic traits and their related GEBVs (p.CE, p.DOC, d.CED, d.CEM, and d.DOC) were estimated using a Gibbs sampling approach for threshold traits (Albert and Chib 1993), implemented in the thrgibbsf90 software (Tsuruta and Misztal 2006).

RESULTS

Table 1 summarises r_gs between GEBVs and their phenotypic target traits. Note that for d.WWM and d.CEM the correlation to the maternal genetic component of p.WW and p.CE, respectively, is given. In general, r_gs of 50KGB and 50K3 GEBVs were very similar and showed the same trend in response to changes of the GEBV set. For FULL sets, highest r_g of 0.69 was found for $\text{d.SC}^{\text{Full}}_{\text{50K3}}$, followed by 0.67 for $\text{d.BWD}^{\text{Full}}_{\text{50K3}}$. The exclusion of US training individuals (FULL→AU) led to a decrease in r_g of more than 0.1 only for d.BWD, d.WWD, d.YWD, d.MCW, d.SC and $\text{d.WWM}_{\text{50K3}}$. For all other GEBVs a decrease < 0.1 or even an increase (e.g. d.CIM, $\text{d.CWT}_{\text{50K3}}$, d.CEM) was be observed. Thus, r_gs of continuous reproductive and growth traits were affected most by this exclusion, whereas carcase and categorical traits were unaffected. When excluding additionally the progeny of training individuals (AU →AUS), r_gs of growth and reproductive traits decreased further (except

Table 1: Genetic correlation (accuracy)|*standard error* between GEBVs and their phenotypic target traits by estimation procedures and GEBV subsets

GEBV	phenotypic trait	50KGB[1]			50K3[2]		
		FULL[3]	AU[4]	AUS[5]	FULL	AU	AUS
d.BWD	p.BWD	0.65|*0.03*	0.45|*0.06*	0.46|*0.08*	0.67|*0.03*	0.44|*0.07*	0.35|*0.09*
d.WWD	p.WW	0.64|*0.03*	0.42|*0.06*	0.35|*0.09*	0.60|*0.03*	0.44|*0.06*	0.32|*0.10*
d.YWD	p.YWD	0.61|*0.03*	0.37|*0.06*	0.28|*0.10*	0.53|*0.04*	0.31|*0.07*	0.15|*0.10*
d.MCW	p.MCW	0.48|*0.05*	0.26|*0.08*	0.12|*0.11*	0.47|*0.05*	0.29|*0.08*	0.16|*0.12*
d.SC	p.SC	0.61|*0.03*	0.42|*0.07*	0.41|*0.10*	0.69|*0.03*	0.53|*0.07*	0.49|*0.10*
d.CWT	p.CWT	0.50|*0.12*	0.47|*0.14*	0.49|*0.18*	0.55|*0.12*	0.57|*0.15*	0.78|*0.16*
d.CIM	p.CIM	0.40|*0.13*	0.46|*0.15*	0.59|*0.17*	0.54|*0.14*	0.75|*0.14*	0.91|*0.16*
d.CEA	p.CEA	0.47|*0.13*	0.45|*0.16*	0.50|*0.20*	0.40|*0.16*	0.30|*0.19*	0.45|*0.26*
d.WWM	p.WW	0.35|*0.06*	0.30|*0.08*	0.26|*0.12*	0.36|*0.06*	0.24|*0.10*	0.20|*0.14*
d.CED	p.CE	-0.21|*0.11*	-0.18|*0.07*	0.04|*0.09*	-0.15|*0.11*	-0.11|*0.07*	0.17|*0.11*
d.CEM	p.CE	-0.24|*0.09*	-0.41|*0.06*	-0.38|*0.07*	-0.25|*0.10*	-0.47|*0.05*	-0.39|*0.09*
d.DOC	p.DOC	-0.23|*0.08*	-0.25|*0.09*	-0.13|*0.11*	-0.25|*0.09*	-0.27|*0.11*	-0.13|*0.11*

1: GEBV estimated by a GBLUP approach from trait-independent SNP genotypes obtained from an Illumina 50K Bead Chip, 2: GEBV estimated from 392 SNP individually chosen for each trait where genotypes were obtained from an Illumina 50K Bead Chip, 3: all genotype individuals, 4: individuals of Australian origin only, 5: individuals of Australian origin but no direct sons of US bulls.

$d.BWD_{50KGB}$), whereas r_gs of carcase traits increased (e.g. d.CIM, d.CEA and d.CWT).

Independently of the GEBV set size the vast majority of REML estimates of GEBV heritabilities (h^2) was one, and their standard errors increased as set size decreased (results not shown). Gibbs sampling h^2 estimates were never one, even for the FULL set regardless of the estimation procedure, and generally decreased with decreasing set size (from FULL to AUS). For continuously distributed traits the variance of the direct additive genetic effect (σ_a^2) was much larger for the phenotypic trait than for the related GEBV(e.g. 11.2 for p.CEA and 0.03 for $d.CEA^{Full}_{50K3}$, results not shown). In contrast, σ_a^2 of p.CE and p.DOC were generally smaller than those of their related GEBVs. Comparing both the estimation procedures, σ_a^2 of 50K3 GEBVs were always larger than those of 50KGB GEBVs (results not shown).

DISCUSSION

Using the AU set as a reference, the results given here (0.24 to 0.75 for continuous traits) are well in line with those already published (MacNeil *et al.* 2010; Northcutt 2011; Johnston *et al.* 2010). Blending Pfizer GEBVs of similar accuracies into Australian Angus BREEDPLAN estimated breeding values resulted in an increased overall accuracy of 1.4 % to 7.5 % dependent on the trait (Johnston *et al.* 2012). Thus, similar results can be achieved when blending Australian Angus estimated breeding values with Igenity GEBVs.

Results also show that a selection of 392 SNP individually chosen for each trait out of those present on the Illumina 50K Bovine Bead Chip performs as well as a GBLUP approach using all available SNP. Moreover, trends in r_gs and their standard errors of both prediction procedures are similar, and, apart from statistical significance, for the majority of traits r_gs from the 50K3 approach were slightly higher than from the 50KGB approach. Thus, if these 392 SNP track large haplotypes, it raises questions about the additional benefit of using 800K or full genome sequencing for accuracies of GEBVs. For growth and reproductive traits r_gs

decreased from FULL to AUS, which is in line with the theoretical expectation. Contrarily, especially for carcase traits r_gs did not generally decrease with an increasing genetic distance between the training and the validation set. This is especially the case for d.CIM_{50K3}, where r_g increased from 0.54 (FULL) to 0.91 (AUS), for d.CWT_{50K3} (0.55→0.78), and for d.CIM_{50KGB} (0.40→0.59). A possible reason for this observation is the decrease in subset sizes of GEBVs (AU (345) and AUS (188)) which possibly offset the effect of a decreasing relationship by sampling. However, a decrease in subset size occurred across GEBVs, thus also in those where r_gs decreased as expected. Compared to growth and reproductive traits, carcase traits are characterised by a generally low number of phenotypic observations. Excluding US animals when moving from FULL to AUS possibly increased the average genetic relationship between individuals with GEBVs and individuals with phenotypic observations for the 3434 p.CIM records much more than for the 234087 p.WW records. Thus, a possible positive effect of this increased relationship on r_gs might have superposed negative effects of a decreased GEBV subset size and increased genetic distance between training and validation set. However, as the average genetic relationship between GEBV sets and phenotypic trait sets was not analysed, further research in this area is necessary. Since sample sizes of GEBVs and also of phenotypic carcase traits are still limited, results need to be verified by larger number of phenotypic records and more individuals with both phenotypes and genotypes. Nevertheless, results indicate that blending Australian Angus estimated breeding values with Igenity GEBVs can improve overall accuracy especially for difficult to measure traits.

ACKNOWLEDGEMENTS

The authors thank Meat and Livestock Australia for financially supporting this work (Project B.BFG.0050). We acknowledge Igenity, Lincoln, NB, USA, for providing GEBVs and the Angus Australia Society for providing access to their data.

REFERENCES

Albert J. H. and Chib S. (1993) *J Am Stat Assoc* **88**:669 .

Johnston D. J., Jeyaruban M. G. and Graser H.-U. (2010) Evaluation of Pfizer Animal Genetics HD 50K MVP Calibration.

Johnston D. J., Tier B. and Graser H.-U. (2012) *Anim Prod Sci* **55**:100.

MacNeil M. D., Nkrumah J. D., Woodward B. W. and Northcutt S. L. (2010) *J Anim Sci* **88**(2):517.

Meyer K. (2007) *J Zhejiang Univ Sci B* **8**(11):815.

Northcutt S. L. (2011) Genomic Choices.

Tsuruta S. and Misztal I. (2006) in Proc. 8th. WCGALP, Belo Horizonte, Brazil, August, 2006, pp. 27 – 31.