*Proc. Assoc. Advmt. Anim. Breed. Genet. 22:309-312*

# WEIGHTING FACTORS FOR GENOMIC INFORMATION USED IN SINGLE-STEP GENOMIC SELECTION IN AUSTRALIAN BEEF

**Y. D. Zhang, A. Swan, D.J. Johnston and C. J. Girard**

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

## ABSTRACT

Single-step genomic evaluation utilises all phenotypes, pedigree and genotypes and could significantly enhance beef cattle genetic evaluation. An appropriate weighting factor for genomic and pedigree information is required to predict single-step estimated breeding values (EBVs). This study assessed the optimal weighing factor lambda (λ, ranging between 0 and 1 for none to 100% weighing on genomic information) for a series of beef traits using an empirical approach. The optimal value of λ was identified from the maximum accuracies of genomic predictions by internal cross-validation. The estimated genomic accuracies for Brahman cattle ranged from 0.23 to 0.70 for traits with adequate numbers of genotypes and phenotypes. The accuracy of genomic predictions generally increased as the λ weighting factor increased for a range of traits and typically approached an asymptote towards the optimal λ. For traits with adequate numbers of records, the optimal λ values ranged from 0.4 to 0.8.

## INTRODUCTION

Application of genomic selection in livestock enables more accurate selection of animals at younger ages, and for hard to measure and sex-limited traits. Ultimately, the use of genomic selection can increase genetic gain. Best Linear Unbiased Prediction (BLUP) is a traditional and reliable tool to estimate breeding values and it has served animal breeders well. Genomic BLUP (GBLUP) works in the similar way to BLUP, but substitutes the pedigree based relationship matrix A with the genomic relationship matrix *G*. The recently developed single step genomic BLUP (ssGBLUP) by Legarra *et al*. (2009) and Christensen and Lund (2010) makes use of genotypes, all phenotypes and pedigree information, aiming to streamline the application and enhance the accuracies of EBV. The variance matrix of EBVs for ssGBLUP combines *A* and *G*, and the inverse matrix ($H^{-1}$) required to solve the ssGBLUP equations has a simple form as shown by Aguilar et al. (2010): $H^{-1} = A^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & G^{-1} - A_{22}^{-1} \end{bmatrix}$. An appropriate weighting of pedigree and genomic information when constructing *G* is required because SNP marker panels do not explain all of the additive genetic variation (e.g. Goddard *et al*. 2011). A modified genomic relationship matrix is typically used, as $G = \lambda G_m + (1 - \lambda)A_{22}$, where $\lambda$ is the fraction of the additive genetic variance explained by markers, ranging between 0 and 1. This study assessed the optimal weighing factor λ using an empirical approach.

## MATERIALS AND METHOD

**Data.** Phenotypes, pedigree and genotypes for this study were from the BREEDPLAN database for Brahman cattle. Traits in this analysis included growth (5 traits), ultrasonic scanning body composition (6 traits), carcase characteristics (6 traits), flight time, scrotal circumference and days to calving (DTC). Table 1 summarises the pedigree, records, number of genotypes available for each trait. DTC was measured repeatedly. On average, every animal had 2.5 DTC records and 4.6 for genotyped animals. In total there were 7166 animals genotyped and included in the $H^{-1}$ matrix.

*AGBU is a joint venture of NSW Department of Primary Industries and the University of New England

**Statistical models.** Single step GBLUP (ssGBLUP) analyses were performed in Wombat (Meyer 2007), using all available records for each of these traits. The data were obtained from the BREEDPLAN database, with phenotypes pre-adjusted for all fixed effects but contemporary group. The model fitted contemporary group as the sole fixed effect, and the additive genetic breeding value as a random effect. In addition, for birth weight, weaning weight and yearling weight maternal genetic effects were also fitted, and DTC was analysed using a repeatability model. Analyses were performed with the $H^{-1}$ matrix calculated for a range of λ values between 0 and 1 with an increment of 0.1. The optimal value of λ was identified by the highest accuracies of ssGBLUP EBV via five-fold cross-validation. Animals with both genotypes and phenotypes were split into five groups, based on half-sib family structure, with no progeny within half-sib families allocated to more than one group. In each of the five analyses, four groups were used as the genomic reference to predict EBVs of the fifth (test) group. Phenotypes for animals in the test group were omitted from the training data, but their pedigree and genotype data were included in the $H^{-1}$ matrix in order to obtain their EBVs. This cross-validation was performed for the range of λ values from 0 to 1.0.

**Table 1. Summary of data for each trait, numbers of animals (N) and number of sires (Sires) for phenotypes and genotypes, and average size of test set in 5-fold cross-validation (ncv)**

| Trait | Phenotypes | | Genotypes | | |
|---|---|---|---|---|---|
| | N | Sire | N | Sire | ncv |
| Birth weight (BWD) | 19567 | 1145 | 2072 | 219 | 357 |
| Weaning weight: 200 day (WWD) | 198250 | 5249 | 5677 | 653 | 796 |
| Yearling weight: 400 day (YWD) | 101415 | 4382 | 4607 | 510 | 701 |
| Final weight: 600 day (FWD) | 102490 | 4370 | 4295 | 506 | 647 |
| Mature cow weight (MCW) | 8433 | 930 | 1241 | 155 | 203 |
| Heifer scan eye muscle area (HEA) | 10562 | 714 | 1814 | 84 | 341 |
| Bull scan eye muscle area (BEA) | 10852 | 1013 | 1459 | 140 | 260 |
| Bull scan rib fat (BRF) | 9921 | 963 | 839 | 121 | 141 |
| Bull scan p8 fat depth (BP8) | 10128 | 971 | 854 | 127 | 141 |
| Carcase weight (CWT) | 2982 | 178 | 933 | 89 | 171 |
| Carcase P8 depth (CP8) | 2675 | 146 | 911 | 89 | 167 |
| Carcase rib fat (CRF) | 2569 | 146 | 859 | 88 | 156 |
| Carcase intramuscular fat (CIM) | 2703 | 154 | 926 | 89 | 170 |
| Shear Force (SHF) | 2584 | 146 | 898 | 89 | 163 |
| Flight Time (FLT) | 7756 | 280 | 1195 | 81 | 227 |
| Scrotal size (SS) | 27709 | 2049 | 1686 | 261 | 263 |
| Days to calving (DTC) | 18178 | 1349 | 1130 | 139 | 178 |
| Pfizer MBV Tenderness (MPT) | 6909 | 1158 | 1920 | 173 | 342 |

The accuracies of genomic predictions were calculated as the correlation between EBVs and adjusted phenotypes, scaled by the square root of the heritability of the trait, which was estimated using all records and pedigree. The means of the five scaled correlation coefficients are presented as the accuracy. For repeated records (DTC), adjusted phenotypes were calculated as the average residual from a repeatability model fitting contemporary group, then weighted according to Garrick *et al.* (2009). The heritability used to calculate the accuracy for DTC was also adjusted

according to $h^2_{adj} = h^2/(t + \frac{1-t}{n})$, where t is the repeatability and n is the average records per animal within each test set. EBVs for three λ values (0, 0.5 and 1.00) were compared in five classes where animals were phenotyped or genotyped or both.

**RESULTS AND DISCUSSION**

Results are summarised in Table 2. There was a wide variation in the value of λ at the highest accuracies from 0.1 to 1.0. The highest accuracies of EBV ranged from 0.15 for CIM to 0.70 for SS. For traits with reasonable number of records (BWD, WWD, YWD, FWD, MCW, HEA, BEA, SS, FLT), the λ values ranged from 0.4 to 0.8, and the corresponding accuracies of EBV ranged from 0.23 to 0.70. The λ value at the highest accuracy for CRF (0.1) differed markedly from most traits, possible due to the quality of phenotypes for this trait (carcase might be trimmed prior to measurement). As the maximum was approached, accuracy was relatively insensitive over a large range in λ values. This was observed in most traits as the response surface generally approached an asymptote.

**Table 2. Results of Brahman cross-validation tests for a range values of λ, with estimated heritability ($h^2$), maximum accuracy (r_max), λ_max (λ at r_max), and range in λ where accuracy varied by -0.01 around r_max (λ_low to λ_high)**

| Trait | $h^2$ | λ_max | r_max | λ_low | λ_high |
|---|---|---|---|---|---|
| BWD | 0.45 | 0.60 | 0.53 | 0.30 | 1.00 |
| WWD | 0.32 | 0.40 | 0.45 | 0.20 | 0.70 |
| YWD | 0.38 | 0.60 | 0.33 | 0.30 | 0.90 |
| FWD | 0.43 | 0.70 | 0.53 | 0.30 | 1.00 |
| MCW | 0.60 | 0.80 | 0.40 | 0.50 | 1.00 |
| HEA | 0.30 | 0.90 | 0.23 | 0.50 | 1.00 |
| BEA | 0.29 | 0.50 | 0.37 | 0.30 | 0.70 |
| BRF | 0.28 | 0.90 | 0.26 | 0.60 | 1.00 |
| BP8 | 0.42 | 0.70 | 0.28 | 0.40 | 1.00 |
| CWT | 0.51 | 0.30 | 0.47 | 0.20 | 0.60 |
| CP8 | 0.30 | 1.00 | 0.27 | 0.50 | 1.00 |
| CRF | 0.26 | 0.10 | 0.24 | 0.00 | 0.30 |
| CIM | 0.25 | 0.40 | 0.15 | 0.10 | 0.80 |
| SHF | 0.27 | 0.80 | 0.41 | 0.40 | 1.00 |
| FLT | 0.28 | 0.50 | 0.51 | 0.20 | 0.80 |
| SS | 0.43 | 0.70 | 0.70 | 0.40 | 1.00 |
| DTC | 0.05 | 0.80 | 0.34 | 0.60 | 1.00 |
| MPT | 0.72 | 1.00 | 0.50 | 0.70 | 1.00 |

Table 3 shows impacts of three values of λ on EBVs for CP8, DTC, MCW and WWD. The variation in EBVs increased from λ=0 to 0.5 for phenotyped or phenotyped but not genotyped classes of animals, but less in moving from λ=0.5 to 1.0. In contrast, the variation in EBVs increased with from λ=0 to 1.0 for genotyped and genotyped but not phenotyped classes of animals, and so for most traits in both genotyped and phenotyped class. Correlations of EBVs across three values of λ were consistently high for P and P-G classes. Understandably, this was due to the impact of direct phenotypic information. For G or G-P animals, EBVs were predicted

through a combination of pedigree and genomic relationships, and correlations between λ=0.5 and λ=1.0 were always very high (0.93 to 0.97); lower correlations were observed for EBVs between λ=0 and λ=0.5 (0.88 to 0.95). The correlation the λ=0 and λ=1.0 further decreased (0.70 to 0.85).

**Table 3. Comparison of EBV for animals in phenotyped (P), phenotyped but not genotyped (P-G), genotyped (G), genotyped but not phenotyped (G-P), and both phenotyped and genotyped (P+G) classes over three values of λ (0, 0.5 and 1.0). EBV standard deviations for three values of λ (0 = sd0, 0.5 = sd50, 1.0 = sd100), and correlations between EBVs (e.g. r50_100 for correlation of EBV between λ values 0.5 and 1.0)**

| Trait | Group | N | sd0 | sd50 | sd100 | r0_50 | r0_100 | r50_100 |
|---|---|---|---|---|---|---|---|---|
| CP8 | P | 2675 | 1.30 | 1.41 | 1.36 | 0.99 | 0.97 | 0.99 |
| CP8 | P-G | 1764 | 1.30 | 1.40 | 1.33 | 1.00 | 0.99 | 1.00 |
| CP8 | P+G | 911 | 1.30 | 1.44 | 1.42 | 0.98 | 0.94 | 0.99 |
| CP8 | G | 7166 | 0.74 | 0.87 | 0.98 | 0.91 | 0.77 | 0.96 |
| CP8 | G-P | 6255 | 0.61 | 0.75 | 0.89 | 0.87 | 0.70 | 0.96 |
| DTC | P | 18178 | 4.52 | 4.77 | 4.82 | 0.99 | 0.97 | 0.99 |
| DTC | P-G | 17048 | 4.50 | 4.73 | 4.76 | 0.99 | 0.97 | 0.99 |
| DTC | P+G | 1130 | 4.86 | 5.38 | 5.72 | 0.97 | 0.90 | 0.98 |
| DTC | G | 7166 | 3.31 | 3.92 | 4.51 | 0.91 | 0.79 | 0.97 |
| DTC | G-P | 6036 | 2.93 | 3.58 | 4.24 | 0.89 | 0.75 | 0.97 |
| MCW | P | 8433 | 26.96 | 27.46 | 26.71 | 1.00 | 0.99 | 1.00 |
| MCW | P-G | 7192 | 26.88 | 27.24 | 26.45 | 1.00 | 1.00 | 1.00 |
| MCW | P+G | 1241 | 27.41 | 28.66 | 28.08 | 0.99 | 0.96 | 0.99 |
| MCW | G | 7166 | 17.07 | 19.28 | 21.15 | 0.92 | 0.81 | 0.97 |
| MCW | G-P | 5925 | 13.96 | 16.61 | 19.32 | 0.88 | 0.73 | 0.96 |
| WWD | P | 198250 | 8.81 | 8.92 | 8.80 | 1.00 | 0.99 | 1.00 |
| WWD | P-G | 192573 | 8.85 | 8.95 | 8.81 | 1.00 | 1.00 | 1.00 |
| WWD | P+G | 5677 | 7.51 | 8.09 | 8.56 | 0.95 | 0.84 | 0.96 |
| WWD | G | 7166 | 7.40 | 8.01 | 8.49 | 0.94 | 0.83 | 0.96 |
| WWD | G-P | 1489 | 6.69 | 7.33 | 7.91 | 0.92 | 0.79 | 0.95 |

In view of these results, a value of λ=0.5 has been adopted in preliminary ssGBLUP analyses for BREEDPLAN, but this may change as more experience is gained with the method. In future, a high weighting factor, for example, λ=0.7, could be considered, as shown in Table 2, a high λ will be beneficial to most traits, but adversely affect WWD, CWT and CIM. The current data structure for the Brahman Breedplan analysis may have impacts on findings. Further study using data from other breeds, e.g. Angus, is required to validate these results.

**REFERENCES**

Aguilar I., Misztal I., Johnson D.L., Legarra A., Tsuruta S., *et al*. (2010) *J. of Dairy Sci.* **93**:743.
Christensen O.F. and Lund M.S. (2010) *Genet. Sel. Evol.* **42**: 2.
Legarra A., Aguilar I. and Misztal I. (2009) *J. of Dairy Sci.* **92**: 4656.
Meyer K. (2007) *J. of Zhejiang Uni. Sci. B* **8**: 815.
Goddard M., Hayes B. and Meuwissen T. (2011) *J. of Anim. Breed. Genet.* **128**: 409.
Garrick D.J., Taylor J.F. and Fernando R.L (2009). *Genet. Sel. Evol.* **41**:55.