290. A new metric to assess reference populations for genomic selection in Australian beef breeds

K.L. Moore*, M.H. Ferdosi, C.G. Girard, S.F. Walkom and D.J. Johnston

AGBU, a joint venture of NSW Department of Primary Industries and University of New England, 2351, Armidale, Australia; kmoore7@une.edu.au

https://www.wageningenacademic.com/doi/pdf/10.3920/978-90-8686-940-4_290 - Sunday, July 30, 2023 11:18:38 PM - IP Address:129.180.167.125

Abstract

Research has shown that to maximise benefits (i.e. EBV accuracy and spread and decreased generation interval) of genomic selection, reference populations of genotyped and phenotyped animals need to be large, diverse in relationships and closely related to the selection generation. In practice, applying these design principles can be challenging as metrics to aid decision making are often not available. This paper applies a metric that objectively describes reference populations and their impact on accuracy for seven Australian beef populations. All populations for live weight and abattoir carcase had different levels of phenotyping, genotyping and overall size. Carcase traits had the lowest reference population size, and assessment of average relatedness in almost all the breeds showed there were herds that were not currently represented in the reference population. Generally, increasing reference size resulted in increased accuracy, but there were some exceptions. The metrics described in this paper are easy to apply and can assist in the construction of reference populations.

Introduction

Single-step GBLUP has been implemented into Australian BREEDPLAN genetic evaluations (Johnston *et al.* 2018). Reference population data to underpin genomics are generated by a number of projects, and are run by either breed societies or research organisations. Reference population design is critical and impacts the benefit obtained from genomics. The accuracy of genomic selection is directly impacted by reference population size, trait heritability, and effective population size (Goddard and Hayes, 2009), as well as relatedness amongst the reference animals, and relatedness to selection animals (Pszczola *et al.* 2011). As the relationship between the reference and selection population increases, smaller reference population sizes are required to achieve the same level of accuracy (Lee *et al.* 2017). Therefore, multiple design principals need to be balanced when constructing reference populations. Applying these design principals can be challenging for many practical reasons, including that there are limited quantitative metrics and tools available to make informed objective decisions on sire selections for building reference populations. The aim of this study was to develop a metric that describes and assesses reference populations for suitability for genomic selection on a breed, herd or individual level, and to explore the relationship between reference size and the realised accuracy observed in industry datasets from single-step genetic evaluations.

Materials & methods

Seven Australian beef populations represented a wide range of population structures and will be anonymised. Breeds A, B, D and G have established BREEDPLAN single-step genetic evaluations with breed F recently implementing single-step. Breed C and E are yet to implement single-step into their genetic evaluations. All metrics described in this paper were applied to animals born 2010 or later. Live weight and carcase reference populations were considered, with reference animals being those involved in structured projects and also qualifying industry animals. The reference population was defined as animals with both phenotype and genotype available, as well as genotyped but un-phenotyped sires with five or more progeny phenotypes not already included in the reference. For inclusion in the reference population, animals required at least one live weight record (200, 400 or 600 day) for the live weight reference population, and at least one of abattoir

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_290, © K.L. Moore *et al.* 2022

carcase records (carcase weight, P8 rump fat, rib fat, eye muscle area, retail beef yield or intramuscular fat) for the carcase reference population.

Numerator relationship matrices were constructed with unpublished AGBU nrmblock software written by Ferdosi based on algorithms by Aguilar *et al.* (2011) and Sargolzaei *et al.* (2005). For each animal, the average relatedness with all animals and then reference only animals for each trait was calculated based on the off-diagonal elements. To assess how trait specific reference populations represented the wider population, the average relatedness to reference animals (y-axis) was plotted against the average relatedness to all animals (x-axis) (Figure 1). This visual metric was quantified by the regression slope for all animals and for individual herds. If all animals were phenotyped and genotyped the figure (and slope) is a 1:1 line with no deviations. A regression slope close to 1 was considered optimum, <1 under and >1 over-represented in the reference population.

To explore the relationship between reference size and selection accuracy, single trait BLUP approximate accuracies from the BREEDPLAN genetic evaluations were used. The average live weight (200, 400 or 600 day) accuracy for genotyped but un-phenotyped animals was plotted against the reference size for each breed. To control against influencing factors (i.e. different effective population sizes) and considering only the larger reference populations, a series of phenotype knockouts were undertaken for breed B and D 400 day yearling weights. Phenotypes of whole contemporary groups were randomly removed from the analysis to achieve reference sizes between 1000 and 20,000 animals. The accuracy of genotyped but un-phenotyped animals was plotted against the reference size.

Results

Each beef population was different sized with varying levels of phenotyping and genotyping (Table 1). The percentage of genotyped animals were 4, 6, 8, 12, 16, 50 and 80% respectively for Breeds C, F, D, B, A, G and E. Reference sizes were greatest for live weight (3.5 to 68.3% of all animals) and smallest for abattoir carcase (0.0 to 3.7% of all animals) traits where recording tended to be limited to research projects.

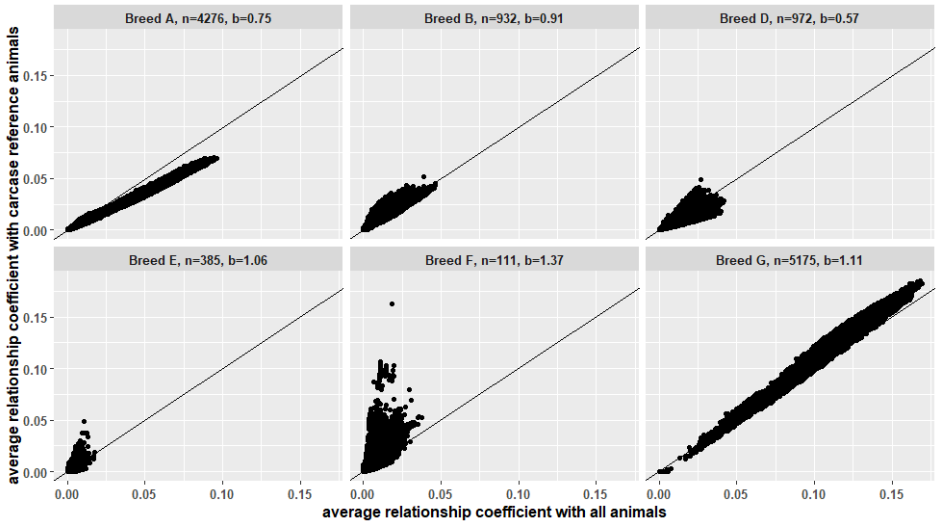

Figure 1. The average relatedness to all animals (x axis) compared with animals in the carcase reference population (y axis) for six Australian beef populations.

https://www.wageningenacademic.com/doi/pdf/10.3920/978-90-8686-940-4_290 - Sunday, July 30, 2023 11:18:38 PM - IP Address:129.180.167.125

https://www.wageningenacademic.com/doi/pdf/10.3920/978-90-8686-940-4_290 - Sunday, July 30, 2023 11:18:38 PM - IP Address:129.180.167.125

Table 1. Description of population and reference size and regression slope from average relationship plots for live weight (LWT) and abattoir carcase (CAR) traits.

	Population size				Regression slope			
	N animals	Reference size (herds)		N GT[1]	LWT		CAR	
Breed	2010+	LWT	CAR	herds	All[2]	Herds[3]	All[2]	Herds[3]
A	1,043,859	136,637 (662)	4,276 (17)	891	1.25	0.71	0.75	0.60
B	165,714	12,896 (73)	932 (5)	273	0.78	0.74	0.91	0.82
C	106,222	3,791 (125)	0 (0)	236	1.29	0.76	-	-
D	410,167	31,379 (390)	972 (12)	483	1.16	0.03	0.57	0.03
E	10,435	7,127 (7)	385 (7)	6	1.25	1.20	1.06	0.79
F	80,085	4,619 (24)	111 (2)	38	1.24	0.90	1.37	0.77
G	175,423	26,923 (133)	5,175 (40)	347	1.20	1.14	1.11	1.03

[1] GT=genotyped.

[2] All = slope for all animals.

[3] Herds= Minimum slope for an individual genotyped herd.

Slopes were calculated for all animals and individual herds, Table 1 reports the overall slope and the minimum herd slope. Larger reference groups were more representative of the general population. For breed G, the overall slope (1.11) and minimum herd slope (1.03) indicated a relevant reference population for genotyping herds. However, the minimum herd slope for un-genotyped herds (results not shown) were as low as 0.02. Breed D had the lowest slope both on a breed (0.57) and herd level (0.03) and along with Breed C showed the most variation in herd slopes (results not shown).

The size and structure of the reference population both impacted accuracy of genotyped un-phenotyped animals. Figure 2 shows that as reference size increased, EBV accuracy also increased. The trajectory was similar for all breeds, as well as when knock-out results were considered for Breeds B and D. Breed G accuracy was higher than expected given reference size, but Figure 1 showed the breed to have higher average relationship values. Breeds B, E and F all had similar accuracy yet reference sizes ranged from 4,619 to 12,896.

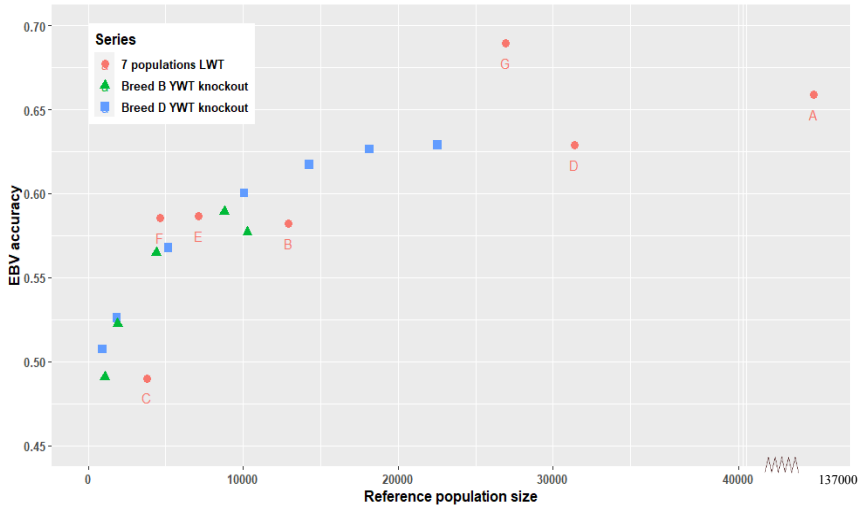

Figure 2. The relationship between reference size and average accuracy of genotyped only animals for live weight (LWT) from seven populations and yearling live weight (YWT) for Breed B and D knockout evaluations.

Discussion

The regression slope metric describes how well reference populations covered the general populat-ion. Larger refer-ence populations generally represent-ed the general population better. This may be more of a reflection on the reference size relative to the wider population size.

Breed B was much closer to 1. Breed D has a much larger population with more herds genotyping but overall fewer animals genotyped. This was also evident with Breed A, who had lower slopes despite having one of the numerically largest carcase reference populations, but when expressed as a proportion of the wider population, the carcase reference was one of the smallest. Results for Breed A show the population to be generally consistent with no herds identified with low slopes. The shape of the average relatedness figure was also informative with the ideal shape being a narrow cloud following the 1:1 line. Breed F had a large number of dots above the line indicating animals already heavily represented in the reference. This plot shape is more likely with a very small reference like we see with Breed F. Herds that genotype benefit from genomic selection due to their linkage to the reference population. The regression slope metric can identify herds with low slopes so that their genetics can either be included into reference data projects or they can use reference animals to build genetic links with their herd and ultimately increase the overall accuracy of genomic selection.

Breed G accuracy was higher compared to other breeds with similar sized reference populations, but Figure 1 shows the breed to have higher average relatedness values. Breeds B, E and F all had similar accuracy despite different reference numbers and proportions. A generally consistent relationship between size and EBV accuracy was observed. Compared to theoretical values (for similar heritability) reported by Goddard and Hayes (2009), the accuracy values we observed from industry datasets were higher when the reference was small (≤2,000) but lower with larger reference populations (≥10,000). For yearling weight, after 10,000 reference animals the observed accuracy tended to plateau rather than continuing to increase as theoretical equations have predicted. Further research is required to understand these differences.

These metrics are an effective and easy way of describing reference populations in the context of the general population and at individual herd level. They provide important information to assist in enhancing the design of the reference population. Breeds can utilise these metrics to identify animals/herds that are not connected to the reference and respond appropriately. These metrics do not consider data quality, so that would need to be considered separately. This simple quantitative metric can help maximise the benefits of the reference population and may be the first objective measurement available to breed societies for managing their reference populations. The next stage of the research will be to investigate the relationship between the relatedness to reference metric with EBV accuracy.

Acknowledgements

This work was supported by Meat and Livestock Australia (L.GEN.1904) and thanks to Angus Australia, Australian Brahman Breeders Association, Australian Wagyu Association, Charolais Society of Australia, Herefords Australia Ltd, North Australian Pastoral Company and Santa Gertrudis Breeders Association and their members for access to their data.

References

Aguilar I., Misztal I., Legarra A. and Tsuruta S. (2011) J. Anim. Breed. Genet. 128(6):422–428. https://doi.org/10.1111/j.1439-0388.2010.00912.x

Goddard M.E. and Hayes B.J. (2009) Nat. Rev. Genet. 10(6):381–391. https://doi.org/10.1038/nrg2575

Johnston D.J., Ferdosi M.H., Connors N.K., Boerner V., Cook J., *et al.* (2018) Proc. of the 11th WCGALP, Auckland, New Zealand

Lee S.H., Clark S. and Van Der Werf J.H.J. (2017) PLoS ONE 12(12):1–22. https://doi.org/10.1371/journal.pone.0189775

Pszczola M., Strabel T., Mulder, H.A. and Calus, M.P.L. (2012) J. Dairy Sci. 95(1):389–400. https://doi.org/10.3168/jds.2011-4338

Sargolzaei M., Iwaisaki H. and Colleau J. J. (2005) J. Anim. Breed. Genet. 122(5):325–331. https://doi.org/10.1111/j.1439-0388.2005.00538.x

https://www.wageningenacademic.com/doi/pdf/10.3920/978-90-8686-940-4_290 - Sunday, July 30, 2023 11:18:38 PM - IP Address:129.180.167.125