IMPUTATION ACCURACY FOR MISSING ALLELES IN CROSSBRED BEEF CATTLE

P.K. Wahinya and M.H. Ferdosi

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

SUMMARY

Establishing a consistent set of markers through the imputation of various marker densities is crucial in building the genomic relationship matrix for genomic prediction. However, imputation in crossbred populations presents challenges. This paper investigated the imputation accuracy of purebred and crossbred Brangus, Simmental and Wagyu cattle. The imputation was carried out independently within each breed using a population-based approach. A reference population of 3,000 randomly selected purebred and crossbred animals with medium-density markers was used in each population to impute the target population with low-density markers (10,000 markers), and this process was repeated five times. On average, imputation accuracies higher than 0.9 were estimated for all three populations. However, the accuracy decreased as the relationship to the reference population within each breed decreased.

INTRODUCTION

Genomic information from Single Nucleotide Polymorphism (SNP) panels is incorporated into BREEDPLAN (Australian beef cattle genetic evaluation system; (Johnston *et al.* 2018)) to predict estimated breeding values. One of the main challenges of including genomic information is that individual animals are genotyped on different SNP panels. Often, different panels contain different SNP densities depending on the size of the panel. Therefore, imputation of un-genotyped SNPs is a standard procedure before building the genomic relationship matrix. This ensures a common consensus SNP panel from different density SNP panels enabling genotypes to be analysed together. The imputation accuracy of the missing SNPs is essential because it influences downstream analyses, such as genome-wide association studies and genomic prediction. Several factors like the relatedness between individuals in the reference and target populations, marker density, reference genome assembly and population structure influence the imputation accuracy (Ferdosi *et al.* 2021a). BREEDPLAN genetic evaluations are currently performed within breeds, however, a substantial proportion of the animals registered within breeds are crossbred at different levels. Currently, crossbred animals with low relationship (less than 80% relationship to the genomic population) to the reference population are not included in BREEDPLAN, partly because imputation in crossbred populations is challenging. Thus, this study aimed to investigate the imputation accuracy in Wagyu, Brangus and Simmental purebred cattle and commercial crossbred animals within each breed.

MATERIALS AND METHODS

Genotypes. Genotypes of 12,058 Wagyu, 8,103 Brangus and 4,778 Simmental cattle including their crosses, were extracted from BREEDPLAN data using quality controls within the BREEDPLAN genomic pipeline (Connors *et al.* 2017). Within the genomic pipeline, an estimated relationship to the reference population approximates how similar animals are to a known population, which can be inferred as breed proportion, and will be referred to as breed proportion for the remainder of this study (Boerner and Wittenburg 2018). The reference population refers to the deviation of individual allele frequencies from the mean allele frequencies in the entire population, which is considered pure based on pedigree. The total number of SNPs considered after

---

* A joint venture of NSW Department of Primary Industries and the University of New England

quality control were, 31,608, 33,302 and 33,126 for Wagyu, Brangus and Simmental breeds, respectively. Individuals and SNPs were removed if they had more than 20% of SNPs missing, while SNPs were only considered if their minor allele frequency was greater than or equal to 5%. Only autosomal SNPs were used for the analysis.

Reference and target populations. Imputation was performed independently within each breed 5 times. In each analysis, 3,000 animals were randomly selected and used as the reference while the remaining formed the target population with 10,000 randomly selected markers (the remaining SNPs were masked to missing genotypes). This is an optimal strategy to improve imputation accuracy (Ferdosi *et al.* 2021a).

Imputation. FImpute v3 software (Sargolzaei *et al.* 2014) with default parameter settings was used to impute the genotypes that were masked as missing. The imputation was population-based, and the pedigree was not utilized. To investigate the imputation accuracy, genotypes of the animals in the target population were imputed back within each breed and were compared with the true genotypes.

Imputation accuracy. The average Pearson's correlation between imputed and true SNPs was estimated to determine the imputation accuracy.

RESULTS AND DISCUSSION

Figure 1 shows the Pearson correlation coefficients between the true and imputed genotypes for Brangus, Simmental and Wagyu cattle. A relatively high imputation accuracy was observed. The average imputation accuracies were 0.96 (0.59 – 1), 0.97 (0.72 – 1) and 0.93 (0.59 – 0.99) for Brangus, Simmental and Wagyu, respectively. Overall the imputation accuracies were higher for animals with a breed proportion greater than 80% within each breed. The imputation accuracy decreased for most animals as their breed proportions decreased. Lower imputation accuracies were estimated for Wagyu crosses (shown as outliers in Figure 1) compared to Brangus and Simmental, especially for animals with less than 50% breed proportion. These results were expected and similar to what is reported elsewhere (Ventura *et al.* 2014; Aliloo and Clark 2021). Incorporating animals with a more substantial relationship to the target population enhances the detection of extended haplotypes which can improve missing SNP imputations. The reference populations in this study were a mixture of pure and crossbred commercial animals within each breed, selected at random. This strategy is important to introduce haplotypes from the other breeds present in the crossbreds. However, all the breeds involved in the crosses were not included in the reference set. The inclusion of these breeds is likely to improve the imputation accuracy, again due to enhanced detection of extended haplotypes.

The differences in the imputation accuracies between the three breeds could be attributed to differences in their population structure, effective population size and the number of crosses. The Simmental population had the least number of commercial crosses while Wagyu had the highest number of commercial crosses with multiple breed combinations. Overall, these results show that randomly missing SNPs in pure breed and crosses of Brangus and Simmental were imputed accurately. Further strategies need to be explored to improve the imputation accuracy for crosses, particularly for Wagyu, where accuracies were largely lower than 0.9 for individuals with less than 50% breed proportion. Wagyu displays a low effective population size and low haplotype diversity (Ferdosi *et al.* 2021b), which could contribute to the low imputation accuracy for animals with low Wagyu content when the parent breeds are missing in the reference population.

This study utilized a population-based imputation method and excluded pedigree information to determine the lowest possible imputation accuracy, since all individuals do not have a pedigree. Combining population and pedigree imputation methods could potentially increase accuracy, but such gains are expected to be minimal with 3,000 individuals in the reference population and a medium-density target population (more than 10k SNPs). Nevertheless, incorporating purebred individuals from which crossbreds were derived may enhance the imputation accuracy for crossbreds.

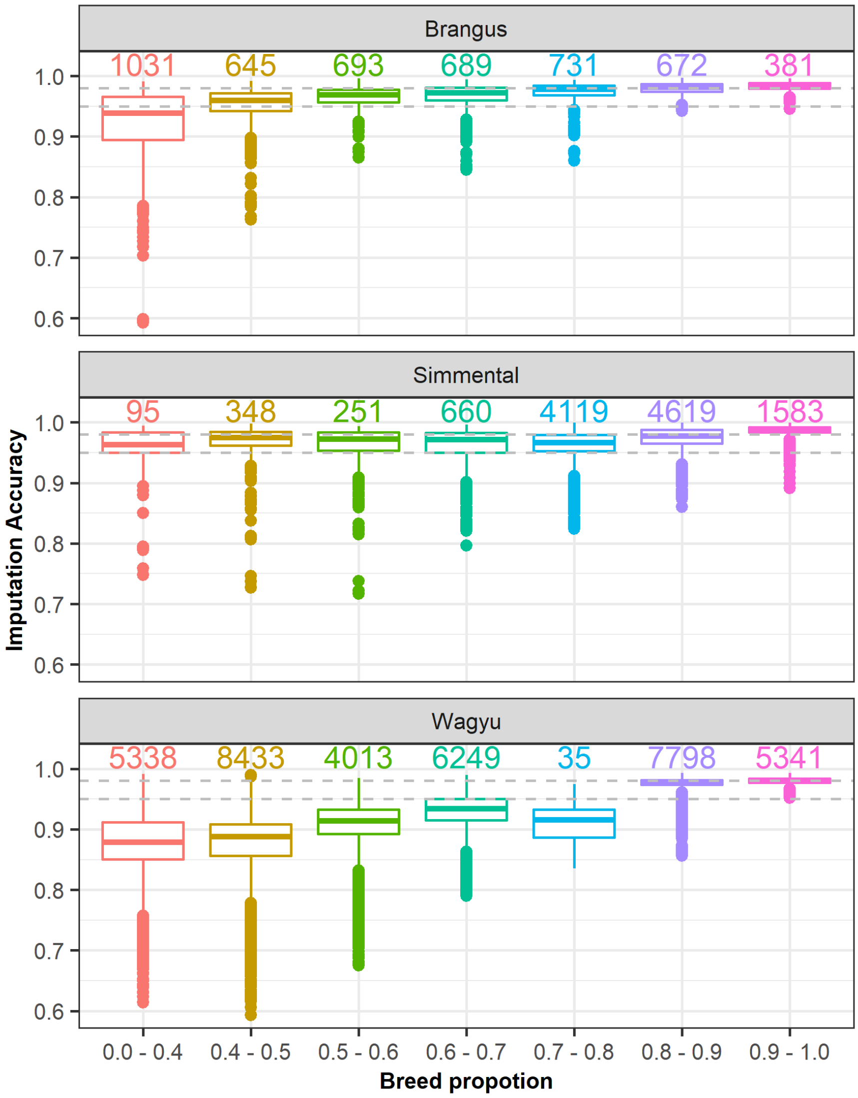

Figure 1. Boxplots showing imputation accuracy across pure-bred and crossbred Brangus, Simmental and Wagyu cattle. The number of animals analysed as target population for different breed proportion ranges combined for the five analyses are shown above each boxplot. Dotted lines represent 0.98 and 0.95 imputation accuracies

CONCLUSION

In this study we investigate the imputation accuracy for missing SNP markers in pure-bred and crossbreed Brangus, Simmental and Wagyu cattle. The results show that the imputation accuracy was on average higher for animals that are more closely related to the reference population.

However, there are differences in the imputation accuracy between the populations, and further studies should evaluate strategies to improve the imputation accuracies for individuals with lower relationships to the reference population when some of the parent breeds are missing. The increase in imputation accuracy would result in a more precise determination of relationships among individuals and an improvement in genomic prediction accuracy, particularly for crossbred individuals.

ACKNOWLEDGEMENTS

This research was partially funded by Meat and Livestock Australia project L.GEN.1704. The authors acknowledge the Agricultural Business Research Institute for the provision of and access to BREEDPLAN data, the Australian Wagyu Association, Simmental Australia and Brangus cattle association, and their breeder members who participate in BREEDPLAN.

REFERENCES

Aliloo H. and Clark S.A. (2021) *Anim. Prod. Sci.* **61:** 1958.

Boerner V. and Wittenburg D. (2018) *Front. Genet.* **9:**

Connors N., Cook J., Girard C., Tier B., Gore K., Johnston D. and Ferdosi M. (2017) *Proc. Assoc Advmt Anim Breed Genet.* **22:** 317.

Ferdosi M., Connors N. andKhansefid M. (2021a) *Proc. Assoc. Advmt. Anim. Breed. Genet.* **24:** 114.

Ferdosi M., Connors N. and Khansefid M. (2021b) *Proc. Assoc. Advmt. Anim. Breed. Genet.* **24:** 118.

Johnston D.J., Ferdosi M., Connors N., Boerner V., Cook J., Girard C., Swan A. and Tier B. (2018) *Proc. World Congr. Genet. Appl. Livest. Prod.* **2018:** 269.

Sargolzaei M., Chesnais J.P. and Schenkel F.S. (2014) *BMC Genomics* **15:** 478.

Ventura R.V., Lu D., Schenkel F.S., Wang Z., Li C. and Miller S.P. (2014) *J Anim. Sci.* **92:** 1433.