

Investigating the relationship between imputation accuracies and relatedness.

Natalie K. Connors and Mohammad H. Ferdosi
Animal Genetics & Breeding Unit, University of New England, Armidale, NSW, 2351 Australia.

Introduction

- Australian beef industry's BREEDPLAN single-step genetic evaluation uses genomic information for the prediction of Estimated Breeding Values (EBVs).
- Imputation of missing SNPs & imputation of low density (LD) to high density (HD) genotypes is essential to combine various SNP densities.
- Target population's imputation accuracy is dependent on relatedness to the reference population.
- Genomic prediction accuracy can be used as a 'score' for an animal's relatedness to the reference population.
- The objective of this study was to identify how well the relatedness can predict imputation accuracies.

Methods

- Genotypes simulated using QMSim Version 1.1 for 10 generations.
- Generations 4 to 10 were used, each containing 40 males, mated with 800 females, litter size of 2.
- Genotypes contained 20 chromosomes with 2000 SNPs in each chromosome.
- SNPs were masked and missing SNPs were imputed using Fimpute version 2.2 (Sargolzaei *et al.*, 2014).
- Genomics Relationship Matrices (GRM) were built using the Dodds method (Dodds *et al.*, 2015) after masking SNPs.
- Reference populations included each generation up to those generations being imputed, including 1600, 2400, 3200, 4000, 4800, 6400, 8000, and 9600 individuals.
- Predictions performed on SNP densities of 500, 700, 1000, 1500, 2000, 2500, 3000, 4000, 5000, 7000, 10000, 15000, 20000, and 30000 SNPs.
- Concordance between imputed SNPs and original SNPs determined imputation accuracy.

Results

- SNP density $\geq 10,000$ has highest correlation between imputation & relatedness, regardless of reference population size.
- Positive non-linear relationship between imputation accuracy & relatedness for individuals across multiple generations with >1000 SNPs used.

- Figure 1 shows highest imputation accuracy for generation 5, reducing for each generation as they become less related to the reference population.

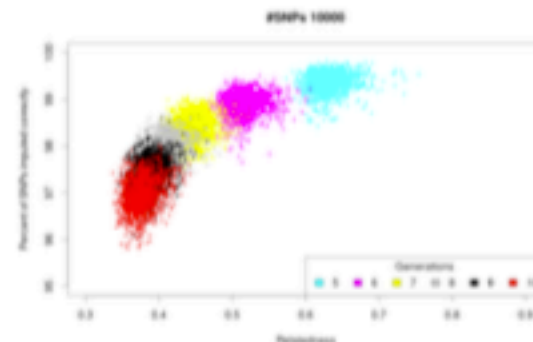


Figure 1. Imputation accuracy of 10,000 SNPs based on reference population of 1600 individuals.

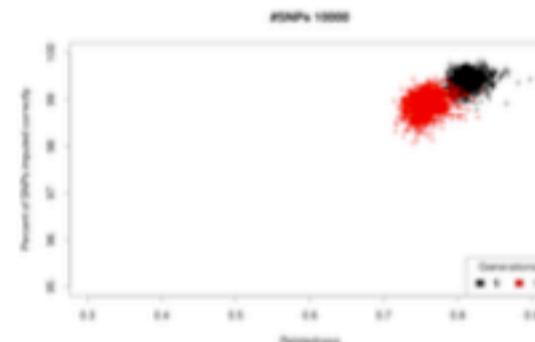


Figure 2. Imputation accuracy of 10,000 SNPs based on reference population of 8000 individuals.

- Figure 2 shows $>98\%$ imputation accuracy for generations 9 and 10, reflecting their relatedness with reference population comprised of generations 4-8.

Conclusions

- Multiple generations needed to achieve higher correlations between imputation accuracy and relatedness.
- Imputation accuracies may be predicted using relatedness.
- Relatedness may explain low EBV accuracies and EBV instability within BREEDPLAN data, due to low imputation accuracies.

References

1. Dodds, K.G., McEwan, J.C., Brauning, R. et al. Construction of relatedness matrices using genotyping-by-sequencing data. *BMC Genomics* 16, 1047 (2015).
2. Sargolzaei, M., Chesnais, J.P. & Schenkel, F.S. A new approach for efficient genotype imputation using information from relatives. *BMC Genomics* 15, 478 (2014).