

# Genotype imputation for genomic selection

31/01/2020

Genotype imputation as a cost-effective strategy to increase genotype data for genomic selection in South African beef cattle

Industry Sector: Cattle and Small Stock

Research Focus Area: Livestock Production with Global Competitiveness: Breeding, Physiology and Management

Year of completion: 2020

Primary Researcher: Prof Mahlako Makgahlela

Secondary Researcher: Prof A Maiwashe

The Research Team

Title	Name	Surname	Highest Qualification	Research Institution
Prof	FWC	Neser	PhD	University of Free State
Dr	MD	MacNeil	PhD	Delta Genetics
Prof	MM	Scholtz	PhD	Agricultural Research Council
Ms	S	Mdyogolo	MSc	Agricultural Research Council
Dr	AA	Zwane	PhD	Agricultural Research Council

## EXECUTIVE SUMMARY

### Introduction

The discovery of DNA polymorphisms (single nucleotide polymorphisms (SNP) or simply genomic data) and their cost-effective genotyping platforms have provided breeders and scientists in animal breeding additional tools to select young animals without performance records with much higher accuracies. Breeding programmes incorporating genomic information have achieved substantial increase in genetic improvement for cattle populations around the world. As a start, genotyping strategies are determined to identify individuals that are genotyped to increase the accuracy of predictions, and estimate relationships between candidates more reliably. Meanwhile, accuracy of genome-based schemes is a function of the reference population from which prediction equations for estimating genomic breeding values (GEBV) are developed. Setting up a sizable reference population is costly and remains a challenge for the uptake of genomic selection in South Africa.

### Objective Statement

The aim of this research was to assess the accuracy of genotype imputation from low-density (7 931 SNPs or 7K) or medium density (150 000 SNPs or 150K) to high density (777 962 SNPs or 777K) panels using the reference population defined as influential animals explaining substantial genetic variation in the Afrikaner (AFR), Brahman (BRA) and Brangus (BNG).

## Project Aims

1. Develop a cost-effective genotyping strategy for setting up the reference population using pedigree data
2. Evaluate the accuracy of genotype imputation for SNP marker panels of different densities

## Results

In identifying influential animals for the AFR, BRA, BNG and the other beef breeds (i.e., Limousine, Santa Gertrudis, Simbra and Simmentaler), it was found that only 100 ancestors explained approximately 50% of the genetic variation, and about 90% of the genetic variation was explained by 200 ancestors for all breeds. Genetic variation explained by top 1000 important ancestors was 95, 96 and 84% for AFR, BRA and BNG, respectively. Imputation accuracies within breed reference population, measured as the concordance rate, were 96.60, 91.39 and 89.91 using Beagle and 95.3, 92.8 and 96 using FImpute for AFR, BRA and BNG, respectively. Accuracies in multi-breed were lower ( $\pm 80\%$ ) than within-breed reference populations. Furthermore, results demonstrated that accuracy tends to be greater when imputing from low-density 7K to medium-density 150K than bypassing the latter and impute from low-density 7K to high-density 777K. Higher accuracies were observed using Fimpute (93-97 %) versus Beagle (89-95%) and Impute (91-94%).

## Conclusion

This study investigated the accuracies of imputation within breed and across breeds by masking actual genotypes in the Afrikaner, Brahman and Brangus breeds, and through genotype imputation from low density 7K or medium density 150K to high-density 777K in the Brahman cattle breed. The reference populations for imputation for all breeds were defined as influential animals with high marginal genetic contributions to young animals who were born in the last decade of the pedigree data, which were found to be few relative to the pedigreed population. Genotyped animals used in this study were few but promising accuracies were observed in within breed reference populations for AFR and BRA. Thus, imputation workflow established in this research could be integrated within the framework for implementation of genomic evaluations of GEBV and genomic selection in the Brahman and Afrikaner cattle breeds.

## Popular Article

### Title for Popular Article

#### **Genotype imputation: An essential, promising and cheap tool for assembling the reference population for genomic selection in the Afrikaner and Brahman cattle breeds of South Africa**

The world will need to produce 100% more food in the next 40 years than currently produced (UNFAO, 2002). Accordingly, the demand for beef products, being the top valuable livestock product, will continue to increase significantly. Meanwhile, competition for resources will intensify, dictating that livestock systems must increase both productivity and efficiency. More than 60% of the additional food must come through technological innovations. Genomics is among technologies that will play a pivotal role in meeting the increasing demand while safeguarding natural resources and preventing environmental degradation. Genomic selection (GS) is the selection of genetically superior breeding animals based on genomic breeding values (GEBV) calculated from thousands of DNA markers or single nucleotide polymorphisms (SNP). Breeding programmes using GEBV have achieved substantial increase in genetic improvement for cattle populations around the world. The infrastructure for

implementing GS within breed is a sufficient reference population of phenotyped (measured economic traits) and genotyped (SNP genotypes) animals. Setting up a sizable reference population requires substantial capital investments, and remains a challenge for the uptake of genomic selection in South Africa. There are several SNP genotyping panels of low-, medium- and high densities in terms of the number of SNP markers. Imputation is a method used to fill missing SNP on the low-density panel using medium- or high-density panel as reference population, without paying for the extra information. It provides an opportunity to achieve a sizable reference population for timely uptake of GS.

The aim of this research was to assess the accuracy of genotype imputation from low-density (7 931 SNPs or 7K) to medium density (150 000 SNPs or 150K) or high density (777 962 SNPs or 777K) panels using the reference population defined as influential animals explaining the genetic diversity in the Afrikaner (AFR), Brahman (BRA) and Brangus (BNG). In identifying influential animals for the AFR, BRA, BNG and the Limousine, Santa Gertrudis, Simbra and Simmentaler, it was found that only 100 ancestors explained approximately 50% of the genetic diversity, and about 90% of the genetic diversity was explained by 200 ancestors for all breeds. Genetic diversity explained by top 1000 important ancestors was 95, 96 and 84% for AFR, BRA and BNG, respectively. Imputation accuracies within breed reference population, measured as correctly imputed SNP, were 96.60, 91.39 and 89.91 using Beagle and 95.3, 92.8 and 96 using Fimpute for AFR, BRA and BNG, respectively. Accuracies in multi-breed were lower ( $\pm 80$ ) than within-breed reference populations. Furthermore, results demonstrated that accuracy tends to be greater when imputing from low-density 7K to medium-density 150K than bypassing the latter and impute from low-density 7K to high-density 777K. Higher accuracies were observed using Fimpute (93-97 %) versus Beagle (89-95) and Impute (91-94). Genotyped animals used in this study were few but promising accuracies were observed in within breed reference population for AFR and BRA. Thus, imputation workflow established in this research could be integrated within the framework for implementation of genomic evaluations of GEBV and genomic selection in the Brahman and Afrikaner cattle breeds.

### **Authors**

Dr M.L. Makgahlela, Ms S. Mdyogolo, Prof F.W.C. Naser, Dr M. D. MacNeil, Prof M. M. Scholtz & Prof A. Maiwashe

Please contact the Primary Researcher if you need a copy of the comprehensive report of this project at: [mmakgahlela@arc.agric.za](mailto:mmakgahlela@arc.agric.za)