I-INTERNATIONAL MEETING OF ANIMAL SCIENCE IN SEMI-ARID REGIONS
Universidade Federal do Agreste de Pernambuco – UFAPE
July 03rd to 05th, 2024, Garanhuns-PE

Field of Work: Animal Genetics and Breeding

Imputation accuracy for rare alleles in Canchim cattle

Roney Teixeira da Silva*[1], Donagh Pearse Berry[2], Luciana Correia de Almeida Regitano[3],
Danísio Prado Munari[4], Marcos Eli Buzanskas[5]
[1]Federal University of Paraíba, Areia/PB, Brazil; [2]Animal & Grassland Research and Innovation Center, Cork, Ireland; [3]Embrapa Southeast Livestock, São Carlos/SP, Brazil; [4]São Paulo State University, Jaboticabal/SP, Brazil; [5]São Paulo State University, Botucatu/SP, Brazil. *roney.teixeira@unesp.br

Genotype imputation is used to predict unobserved genotypes due to different densities of markers available in the market and to improve the quality and reliability of genotypic data to contribute to the results of genomic studies. In this context, populations of composite breeds, such as Canchim beef cattle (CA), derived from crosses between Nelore (NE) and Charolais (CH) breeds, can benefit from the incorporation of genome information from the founder breeds for various purposes. The aim of this study was to evaluate the imputation accuracy for rare alleles in Canchim animals with the incorporation of genotypes from the founder breeds. For the target population, genotypes from 285 CA animals were used, while the reference population consisted of 814 NE animals and 897 CH animals. All animals were genotyped with the high-density BovineHD BeadChip panel, containing 777,962 markers. However, the genotypes of CA animals were reduced to the density of the BovineSNP50 BeadChip panel, with 54,609 markers. Three imputation scenarios were tested, where the breeds with one asterisk represent the reference population and the breeds with two asterisks represent the target population (C1 = NE* x CA**, C2 = CH* x CA** and C3 = NE+CH* x CA**). Quality control of the genotypes was performed using the PLINK program, where SNPs or samples with call rates below 90%, SNPs with minor allele frequency (MAF) below 0.001, and SNPs deviating from Hardy-Weinberg equilibrium below 10^{-4} were excluded. Imputation was performed using the FImpute program. Imputation accuracy was determined by the concordance rate, which represents the percentage of correctly imputed alleles when comparing the original and imputed CA genotypes. Imputation accuracy for rare alleles was verified based on MAF with frequencies between 0.001 to 0.01, 0.01 to 0.02, 0.02 to 0.03, 0.03 to 0.04, and 0.04 to 0.05. In scenario C1, accuracies were 92.84%, 90.04%, 88.74%, 86.07%, and 84.74% for the described MAFs. For C2, accuracies were 95.32%, 93.13%, 92.10%, 90.73%, and 89.54%, respectively. Higher percentages were observed for C3, being 98.38%, 97.06%, 95.97%, 94.98%, and 94.14%, respectively. Important results for imputation accuracies of genotypes in MAF from 0.001 to 0.01 were observed, making it possible to impute rare genotypes with high confidence in CA animals with the inclusion of NE and CH breeds in the reference population. Scenario C3 showed the highest imputation accuracy for all studied MAFs compared to C1 and C2. Higher accuracies were observed in scenarios C1 and C2 when considering CH animals as the reference population, which is due to the closer genetic proximity between these breeds and the higher likelihood of them sharing similar haplotype blocks. It is concluded that there are benefits in imputation accuracy for rare alleles in CA cattle with the addition of genotypes from the founder breeds (CH and NE) and the use of multi-breed reference panels is suggested.

Keywords: beef cattle, genetic composition, genomic data, molecular marker.