AB.26

Accuracy of genotype imputation in Canchim cattle using FImpute and Beagle software

Tatiane C S Chud; Universidade Estadual Paulista Julio de Mesquita Filho, Jaboticabal, SP, Brazil
Ricardo V Ventura; Centre for Genetic Improvement of Livestock, University of Guelph, Guelph, ON, Canada
Flavio S Schenkel; Centre for Genetic Improvement of Livestock, University of Guelph, Guelph, ON, Canada
Ismael Urbinati; Universidade Estadual Paulista Julio de Mesquita Filho, Jaboticabal, SP, Brazil
Roberto Carvalheiro; Universidade Estadual Paulista Jul io de Mesquita Filho, Jaboticabal, SP, Brazil
Luciana C A Regitano; Embrapa Pecuaria Sudeste, Sao Carlos, SP, Brazil
Cintia R Marcondes; Embrapa Pecuaria Sudeste, Sao Carlos, SP, Brazil
Danisio P Munari; Universidade Estadual Paulista Julio de Mesquita Filho, Jaboticabal, SP, Brazil
Corresponding author e-mail: tatischud@gmail.com

Abstract:

Introduction: Imputation methods have been used to increase genomic information, allow more animals in genome-wide analyses and reduce genotyping costs. Imputation allows making use of variants that have not been directly genotyped with a low density (LD) panel by inferring missing marker genotypes using information from a reference population genotyped with a higher density (HD) panel. Objective: Population imputation utilizing linkage disequilibrium among markers was implemented using the software BEAGLE v3.3.2 and FImpute v2.2 in a synthetic breed (Zebu x Charolais) beef cattle population (Canchim cattle) genotyped with the Illumina 777K SNP panel. Imputation using different sets of low density SNP panels were investigated (6K, 50K, 20K and 90K) and, for each scenario, three minor allele frequency (MAF) thresholds were applied. Methods: Data set contained 395 Canchim bulls and cows and 1 ancestor Charolais bull (born between 1999 and 2005) genotyped with the Illumina BovineHD panel (~786,000 SNPS). Animals were sorted by birth year; the reference population was composed of older animals while the imputation set contained younger animals that had their 777K genotypes masked to different LD SNP panels. Reference animals were born up until 2004 (253 animals) and validation animals were born in 2005 (143 animals). Multiple scenarios combining different MAF thresholds were tested: no MAF filter applied (S1) and SNPs with MAF < 0.0025 (S2) or MAF < 0.10 (S3) were excluded. Final data set comprised 616,175 autosomal SNPs in S1, 614,727 in S2 and 531,429 in the last scenario (S3). The 6K, 50K, 20K and 90K panels were created by masking SNPs originally in the Illumina BovineHD SNP chip by selecting the markers in common with the Illumina BovineLD (6K), BovineSNP50, GeneSeek Genomic Profiler (GGP) LDi 20K, a nd GGP HDi 90K. Population imputation was implemented using the software FImpute and BEAGLE. Imputation accuracy was measure as genotype concordance rate. Results: The imputation accuracy ranged from 86% to 97% using FImpute and from 79% to 95% when using BEAGLE. Higher imputation accuracies were obtained when the LD panel was composed with 90K markers. FImpute achieved an overall average increase of 6% in imputation accuracy from 6K and 20K to HD and 2% from 50K and 90K to HD, while reducing run-time by 20 to 100, compared to BEAGLE. The difference in imputation accuracy and run-time between BEAGLE and FImpute was lower from 50K and 90K panels to HD. Small reduction in the imputation accuracy was observed in S3, when compared with scenarios S1 and S2. Conclusions: Accurate imputation was observed when imputing from the Illumina BovineSNP50 (~0.95) and GGP HDi 90k (~0.97) to the HD SNP panel using FImpute. BEAGLE was less accurate and much less computationally efficient than FImpute over all the scenarios. Filtering by MAF did not improve the imputation accuracy and, on the contrary, reduced it when a substantial MAF was imposed (>0.10).

"Fellowship Sao Paulo Research Foundation (FAPESP), n.2012/21891-8" e "Fellowship Sao Paulo Research Foundation (FAPESP), n.2013/09050-0", para os seguintes autores, Tatiane C S Chud e Ismael Urbinati