



Uso de raças fundadoras beneficiam a acurácia de genótipos imputados em bovinos da raça Canchim

Roney Teixeira da Silva^{1*}, Renalia Silva Militão¹, Giovanna Maria dos Santos Câmara¹, Maria Victória Henrique Genuíno¹, Ayrton Fernandes de Oliveira Bessa¹, Ana Carolina de Jesus Oliveira², Donagh Pearse Berry³, Luciana Correia de Almeida Regitano⁴, Danísio Prado Munari², Marcos Eli Buzanskas¹

¹ Departamento de Zootecnia, Centro de Ciências Agrárias, Universidade Federal da Paraíba (UFPB), Areia, Paraíba, Brasil.

² Departamento de Engenharia e Ciências Exatas, Faculdade de Ciências Agrárias e Veterinárias (FCAV, UNESP), Jaboticabal, São Paulo, Brasil.

³ Teagasc, Animal & Grassland Research and Innovation Centre, Moorepark, Fermoy Co. Cork., Irlanda.

⁴ Embrapa Pecuária Sudeste, São Carlos, São Paulo, Brasil.

*Autor correspondente: roneyteixeira.dasilva@gmail.com

Resumo: Objetivou-se neste estudo avaliar melhorias na acurácia de imputação em animais Canchim e do grupo genético MA com a incorporação de genótipos de animais Nelore e Charolês. Foram testados oito cenários de imputação, em que os genótipos de alta densidade foram reduzidos para a densidade do painel BovineSNP50 BeadChip (50K) com capacidade de 54.609 SNPs apenas para a população alvo (CA e MA). A população de referência (Nelore e Charolês) permaneceu com genótipos em alta densidade. Foi determinada a acurácia de imputação por meio da taxa de concordância e quadrado de correlação alélica. Variações foram observadas quanto ao número de marcadores após o controle de qualidade dos genótipos em cada cenário. Os cenários C5 e C7 mostraram a maior acurácia pela contribuição do maior parentesco entre populações de animais CA e do grupo genético MA, e o C2 baixa acurácia de imputação devido ao tamanho inferior da população referência. Concluiu-se que a utilização de genótipos das raças fundadoras promove aumento na acurácia para os animais CA e MA, podendo reduzir os custos de genotipagem em raças compostas. A inclusão da raça Charolês como parte da população referência, em conjunto com animais CA, resulta nos maiores ganhos em acurácia.

Palavras-chave: bovinocultura de corte; composição genética; genômica; marcador molecular.

Using founder breeds may benefit the accuracy of imputed genotypes in Canchim cattle

Abstract: The objective of this study was to evaluate improvements in the imputation accuracy in Canchim animals and MA genetic group by adding genotypes from Nelore and Charolais animals. Eight imputation scenarios were built, in which high-density genotypes were reduced to the density of the BovineSNP50 BeadChip panel (50K) with a capacity of 54,609 SNPs for the target population (CA and MA). The reference population (Nelore and Charolais) remained with high-density genotypes. The imputation accuracy was determined using the concordance rate and allelic correlation square. Variations were observed regarding the number of markers, when performing the genotype quality control of each scenario. Scenarios C5 and C7 presented the highest accuracy due to a greater kinship between CA and MA animals, and C2 presented low imputation accuracy due to the smaller reference population size. It was concluded that using founder breeds promotes an increase in accuracy for CA and MA animals, and may reduce genotyping costs in composite breeds. The inclusion of Charolais breed as part of the reference population, together with CA animals, results in the greatest gains in accuracy.

Keywords: beef cattle; genetic composition; genomics; molecular marker.

Introdução

A pecuária de corte brasileira emprega diferentes estratégias de cruzamento para a formação de raças compostas, como o Canchim, que é proveniente da alternância de cruzamentos entre raças taurinas (*Bos taurus taurus*) e zebuínas (*Bos taurus indicus*), resultando em indivíduos com composição genética aproximada de 62,50% de Charolês e 37,50% de Nelore (ALENCAR, 1988). Ferramentas da genômica têm promovido avanços na produção animal, sendo os marcadores moleculares do tipo SNP (*single nucleotide polymorphisms*) os mais utilizados nesta área. Neste sentido, métodos de imputação permitem a inferência de genótipos entre painéis de diferentes densidades, em que são consideradas populações de referência, com genótipos em alta densidade, e populações de imputação, com genótipos em menores

densidades (SARGOLZAEI; CHESNAIS; SCHENKEL, 2014). A incorporação informações familiares (pedigree) (HAYES *et al.*, 2012) e a combinação de dados de diferentes raças (LARMER; SARGOLZAEI; SCHENKEL, 2014) permite melhores acurácias de imputação.

Para raças compostas, a adição de informações das raças fundadoras na imputação possibilita o aumento do número de animais com informações de painéis de alta densidade e proporciona avaliações genéticas mais confiáveis e com menor custo (VENTURA *et al.*, 2014). Desta forma, o objetivo desse trabalho foi avaliar a melhoria da acurácia de imputação em animais Canchim e do grupo genético MA com a incorporação de genótipos das raças fundadoras Nelore e Charolês.

Material e Métodos

Utilizou-se genótipos de 285 animais da raça Canchim e 114 animais do grupo genético MA, cedidos pela Embrapa Pecuária Sudeste, localizada em São Carlos, SP. Como raças fundadoras, foram utilizados genótipos de 897 animais da raça Charolês e 809 animais da raça Nelore. Os animais foram genotipados com painel de alta densidade BovineHD BeadChip (HD), que apresentam 777.962 marcadores. Considerou-se apenas os cromossomos autossômicos e em posições conhecidas de acordo com o mapa ARS-UCD 1.2 (ROSEN *et al.*, 2020). Por meio do programa computacional PLINK v.1.9, foi realizado o controle de qualidade dos dados e foram excluídos SNPs e amostras com taxas de leitura inferior a 90%, SNPs com *minor allele frequency* inferior a 0,10% e SNPs com desvio do equilíbrio de Hardy-Weinberg inferiores a 10^{-4} .

Construíram-se oito cenários de imputação, em que os genótipos de alta densidade foram reduzidos para a densidade do painel BovineSNP50 BeadChip (50K) com capacidade de 54.609 SNPs, apenas para a população alvo (Canchim e MA). As populações de referência permaneceram com genótipos em alta densidade (HD), alternando-se as raças (Tabela 1). Para a imputação, utilizou-se o algoritmo disponibilizado no programa FIMPUTE.

Tabela 1. Cenários de imputação considerando populações de referência com genótipos em alta densidade e populações alvo com genótipos de baixa densidade.

Cenário de imputação	População de referência	População alvo	
		CA	MA
C1	CA	-	x
C2	MA	x	-
C3	NE + CA	-	x
C4	NE + MA	x	-
C5	CH + CA	-	x
C6	CH + MA	x	-
C7	NE + CH + CA	-	x
C8	NE + CH + MA	x	-

C1 a C8 = cenários; CA = Canchim; CH = Charolês; NE = Nelore; MA = grupo genético MA.

Determinou-se a acurácia de imputação por meio da taxa de concordância (TC) e quadrado de correlação (r^2 alélico). Foram comparados os marcadores imputados com aqueles verdadeiros presentes no painel HD original e, em seguida, calculada a proporção de genótipos imputados corretamente, erroneamente e os não imputados, em que a taxa de concordância corresponde a proporção de genótipos imputados corretamente e o r^2 alélico determinado pelo quadrado de correlação entre a contagem de alelos imputados e a contagem de alelos do genótipo original.

Resultados e Discussão

Variações foram observadas quanto ao número de marcadores após o controle de qualidade de genótipos de cada cenário estudado (Tabela 2). Observou-se que nos cenários avaliados a medida de r^2 alélico variou de 0,8764 (C2) a 0,9332 (C7) e a TC de 86,26% (C2) a 92,17% (C5). Essas variações caracterizam o C2 com menor acurácia em ambas medias testadas, e o C5 e C7 valores de acurácia mais elevados, mesmo com diferença mínima de 0,40% para a TC e 0,0015 para o r^2 alélico.

Tabela 2. Número de animais das populações de referência (NPR), e de imputação (NPI) após o controle de qualidade de genótipos, taxa de concordância (TC) e r^2 alélico (r^2).

Cenário	NPR	NPI	SNPs HD*	SNPs 50K**	TC (%)	r ²
C1	283	112	705272	39490	89,93	0,9080
C2	112	283	658480	39490	86,26	0,8764
C3	809	395	654258	35185	90,10	0,9148
C4	921	283	607043	35233	86,66	0,8812
C5	1180	112	635757	37969	92,17	0,9317
C6	1009	283	705567	40935	90,74	0,9192
C7	1989	112	306859	19748	91,77	0,9332
C8	1818	283	310726	20514	88,19	0,9021

C1 a C8 = cenários; C1 = CA x MA; C2 = MA x CA; C3 = NE + CA x MA; C4 = NE + MA x CA; C5 = CH + CA x MA; C6 = CH + MA x CA; C7 = NE + CH + CA x MA; C8 = NE + CH + MA x CA; CA = Canchim; CH = Charolês; NE = Nelore; MA = grupo genético MA; *SNPs HD = *single nucleotide polymorphisms* em painel de alta densidade; ** SNPs 50K = *single nucleotide polymorphisms* em painel de média densidade.

Nos C5 e C7, verificou-se a maior acurácia em ambas as medidas testadas. Este resultado pode ser atribuído ao maior parentesco entre as populações de animais Canchim e do grupo genético MA. Tanto na formação da raça Canchim quanto no grupo MA, observou-se maior predominância das contribuições genéticas provenientes da raça Charolês. Buzanskas *et al.* (2017) estimaram a introgressão racial e constaram que em animais do grupo MA, houve uma contribuição genética de 24% de zebuínos e 76% de animais Charolês. Da mesma forma, em animais do grupo CA, foi observado contribuição de 27% de zebuínos e 73% de animais Charolês.

O C2 é caracterizado por apresentar uma população referência de menor tamanho em comparação aos demais cenários. Essa condição limita a capacidade de capturar toda a variação genética presente nos animais da raça Canchim na população imputada, resultando em baixa acurácia de imputação. Heidaritabar *et al.* (2016) afirmaram que a baixa acurácia pode ser atribuída às diferenças genéticas entre as raças envolvidas, desequilíbrio de ligação entre marcadores e *loci* de interesse, bem como ao tamanho inferior da população referência em relação a população imputada.

Conclusão

A utilização de genótipos das raças fundadoras promove aumento na acurácia de imputação para os animais Canchim e MA. A inclusão da raça Charolês como parte da população referência, em conjunto com animais Canchim, resulta nos maiores ganhos de acurácia observados na imputação. A utilização de populações de referências multirraciais na imputação de genótipos pode reduzir os custos de genotipagem em raças compostas.

Literatura citada

- BUZANSKAS, M. E. *et al.* Study on the introgression of beef breeds in Canchim cattle using single nucleotide polymorphism markers. PLOS ONE, v. 12, n. 2, p. e0171660, 2017.
- HAYES, B. J. *et al.* Accuracy of genotype imputation in sheep breeds. Animal Genetics, v. 43, n. 1, p. 323-330, 2007.
- HEIDARITABAR, M. *et al.* Accuracy of genomic prediction using imputed whole-genome sequence data in white layers. Journal of Animal Breeding and Genetics, v. 133, n. 3, p. 167-179, 2016.
- LARMER, S. G.; SARGOLZAEI, M.; SCHENKEL, F. S. Extent of linkage disequilibrium, consistency of gametic phase, and imputation accuracy within and across Canadian dairy breeds. Journal of Dairy Science, v. 97, n. 5, p. 3128-3141, 2014.
- ROSEN, B. D. *et al.* De novo assembly of the cattle reference genome with single-molecule sequencing. GigaScience, v. 9, n. 3, p. 1-9, 2020.
- SARGOLZAEI, M.; CHESNAIS, J. P.; SCHENKEL, F. S. A new approach for efficient genotype imputation using information from relatives. BMC Genomics, v. 15, n. 1, p. 478, 2014.
- VENTURA, R. V. *et al.* Impact of reference population on accuracy of imputation from 6K to 50K single nucleotide polymorphism chips in purebred and crossbreed beef cattle. Journal of Animal Science, v. 92, n. 4, p. 1433-1444, 2014.