

### Imputação utilizando painéis personalizados em bovinos da raça Canchim

Gabriela Leoni<sup>1\*</sup>, Marcos Eli Buzanskas<sup>2</sup>, Cintia Righetti Marcondes<sup>3</sup>, Luciana Correia de Almeida Regitano<sup>3</sup>, Danísio Prado Munari<sup>4</sup>, Priscila Arrigucci Bernardes<sup>1,5</sup>

<sup>1</sup>Departamento de Zootecnia e Desenvolvimento Rural, UFSC, Florianópolis, Santa Catarina, Brasil.

<sup>2</sup>Departamento de Zootecnia, Universidade Federal da Paraíba, Areia, Paraíba, Brasil.

<sup>3</sup>Embrapa Pecuária Sudeste, São Carlos, São Paulo, Brasil.

<sup>4</sup>Departamento de Engenharia e Ciências Exatas, Universidade Estadual Paulista, Jaboticabal, São Paulo, Brasil.

<sup>5</sup>CQUniversity Institute for Future Farming Systems, Rockhampton, Queensland, Australia.

\*Autor correspondente: gabi2015leonii@gmail.com

**Resumo:** Uma alternativa para reduzir o custo da aplicação da seleção genômica é a realização da imputação dos genótipos. No entanto, para raças em que a população é pequena, há indisponibilidade de grande quantidade de animais genotipados com alta densidade, como observado para raça Canchim. A utilização de raças fundadoras pode colaborar com a imputação, porém algumas vezes resulta em baixa acurácia. Assim, o objetivo deste estudo foi comparar a acurácia de imputação de painéis comerciais com painéis personalizados de baixa densidade imputados para um painel de alta densidade em bovinos da raça Canchim, utilizando população referência formada por animais da raça Nelore. Foram utilizados 400 animais Nelore na população referência e 396 animais Canchim e do grupo genético MA na população de imputação. Os painéis comerciais de baixa densidade considerados continham 20 mil e 50 mil SNPs (polimorfismos de nucleotídeo único) e 3 tipos de personalizações foram aplicadas com o objetivo de obter as mesmas densidades dos painéis comerciais, formando assim 8 diferentes cenários. A acurácia de imputação apresentou valores superiores nos painéis personalizados quando comparados com os comerciais e também, conforme a densidade dos painéis eram maiores, as acurácias tiveram seu valor aumentado. Houve pequena diferença entre a primeira forma de personalizar o painel comparado com a segunda e terceira, indicando que a maneira como o painel é construído interfere nas acurácias de imputação.

**Palavras-chave:** bovinos de corte, genômica, marcadores moleculares

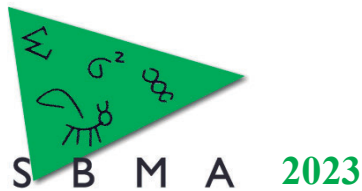
### Imputation using custom panels in Canchim cattle

**Abstract:** An alternative to reduce the cost of applying genomic selection is the imputation of genotypes. However, for breeds in which the population is small, there is unavailability of a large number of genotyped animals with high density, as observed for the Canchim breed. The use of founding breeds or of close origin in the reference population can collaborate with the imputation, but sometimes results in low accuracy. Thus, the objective of this study was to compare the accuracy of imputation of commercial panels with customized low density panels imputed to a high density panel in Canchim cattle, using a reference population formed by Nelore animals. 400 Nelore animals were used in the reference population and 396 Canchim and MA animals in the imputation population. The low-density commercial panels considered contained 20,000 and 50,000 SNPs and 3 types of customizations were applied in order to obtain the same densities as the commercial panels, thus forming 8 different scenarios. The accuracy of imputation showed higher values in the customized panels when compared to the commercial ones and also, as the density of the panels were higher, the accuracies had their value increased. There was a small difference between the first way to customize the panel compared to the second and third, indicating that the way the panel is built interferes with the accuracies of imputation.

**Keywords:** beef cattle, genetic markers, genomics

### Introdução

As tecnologias fundamentadas em informações genômicas favoreceram o desenvolvimento científico na produção animal, principalmente no âmbito da genética e melhoramento. Dentre os diferentes marcadores moleculares, os mais utilizados para essa prática são os polimorfismos de nucleotídeo único (SNP), que estão distribuídos por todo o genoma. Estes marcadores são utilizados para a aplicação da



seleção genômica, sendo que para se obter adequada resposta a essa seleção, torna-se necessário obter grande quantidade de animais genotipados com uma alta quantidade de marcadores SNP. No entanto, a utilização de todos animais da população genotipados com painéis de alta densidade de marcadores ainda possui custo elevado. Dessa maneira, uma alternativa para baratear o custo da aplicação da seleção genômica seria realizar a imputação dos genótipos.

A imputação consiste em utilizar informações da população referência, genotipada em maior densidade, para imputar os genótipos de marcadores ausentes de indivíduos genotipados com painéis de menor densidade. Neste contexto, a aplicação da imputação em pequenas populações pode apresentar dificuldades para formar uma população referência adequada, uma vez que há indisponibilidade de grande quantidade de animais genotipados com alta densidade, como pode ser observado para a raça Canchim. O bovino de corte Canchim foi desenvolvido no Brasil com o propósito de reunir as características de rusticidade presentes em raças zebuínas (*Bos taurus indicus*) com as características de produtividade de animais taurinos (*Bos taurus taurus*).

Uma alternativa para raças de formação recente é a utilização de raças fundadoras ou de origem próxima na população referência, porém algumas vezes esta prática resulta em baixa acurácia. Assim, o objetivo deste estudo foi comparar a acurácia de imputação de painéis comerciais com painéis personalizados de baixa densidade imputados para um painel de alta densidade em bovinos da raça Canchim, utilizando população referência formada por animais da raça Nelore.

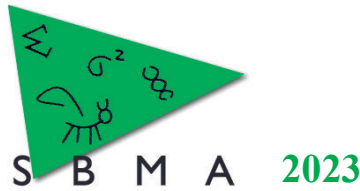
#### Material e Métodos

Os 400 animais da raça Nelore, 285 animais da raça Canchim (CA) e 114 animais do grupo genético MA envolvidos nesse estudo foram genotipados com o BovineHD BeadChip (700k). Os genótipos dos animais Nelores passaram por controle de qualidade utilizando o programa PLINK e foram removidos SNPs localizados em regiões não autossômicas, os que possuíam posição desconhecidas, os que apresentaram MAF menor que 0,001, os SNPs com desvios significativos ( $p < 0,00001$ ) do equilíbrio de Hardy-Weinberg, com taxa de leitura menor que 0,98 e amostras que apresentaram taxa de leitura menor que 0,90. O mesmo foi feito para os demais animais, sendo que os animais Canchim e MA foram considerados como uma única população. Assim, foram considerados para as análises 400 animais da raça Nelore e 396 animais Canchim e MA com 516.044 SNPs.

O total de 8 diferentes cenários foram avaliados. Em todos os cenários, a população referência considerou 400 animais Nelore em que foram mantidos todos os SNPs que passaram pelo controle de qualidade. A população de imputação foi formada por animais Canchim e MA, na qual o total de 516.044 SNPs foi “mascarado” para mimetizar densidades de painéis comerciais contendo 20 mil SNPs (20k) ou 50 mil SNPs (50k), restando um total de 19.263 SNPs no painel de 20k e 26.342 SNPs no painel de 50k, após o controle de qualidade.

Considerou-se também 3 diferentes painéis personalizados, em que a partir do painel de 700k dos animais da raça Canchim e MA, foram escolhidos diferentes SNPs formando painéis contendo a mesma quantidade de SNPs dos painéis comerciais de baixa densidade (50k e 20k). Para a escolha dos SNPs, considerou-se janelas de tamanho de 26 SNPs, para o painel de 20k, e 19 SNPs, para o painel de 50k, e em cada uma destas janelas foi selecionado um único SNP. O critério de escolha desse único SNP pela primeira maneira de personalização foi pela soma de desequilíbrio de ligação (DL - medido pelo  $r^2$ ) entre cada SNP e os demais SNPs presentes nesta janela, em que o SNP que obteve maior soma de DL foi selecionado. A segunda maneira de personalização considerou também o DL, porém o SNP com maior soma de DL deveria atingir um critério mínimo de frequência do alelo de menor frequência (MAF), sendo de 0,2 para 20k e 0,3 para 50k. Por fim, para a terceira maneira, a MAF de cada SNP foi multiplicada pelo DL calculado entre este SNP e os outros presentes na janela. Então, o SNP da janela que apresentou maior valor para soma dos resultados foi selecionado. A imputação dos oito cenários foi realizada utilizando o programa Fimpute (SARGOLZAEI; CHESNAIS; SCHENKEL, 2014). A acurácia de imputação foi verificada pela média da correlação (Cor) entre a quantidade de alelos dos marcadores imputados e a quantidade de alelos do genótipo verdadeiro (BROWNING; BROWNING, 2009).

#### Resultados e Discussão



Em geral, os painéis personalizados apresentaram valores superiores de acurácia quando comparados com os comerciais. O painel de 20k personalizado pela primeira forma obteve maior valor de acurácia (0,6053) quando comparado com o painel de 20k comercial (0,5198). Da mesma maneira, o painel de 50k personalizado pela primeira forma também apresentou maior valor de acurácia (0,6426) do que o painel de 50k comercial (0,5389). O mesmo foi observado para todas as comparações entre personalizados e comerciais, semelhante ao reportado por Romero (2017), em que os painéis personalizados apresentaram vantagem em relação aos painéis comerciais em todos os casos estudados pelos autores.

No presente estudo, o cenário que utilizou o painel de 20k personalizado pela primeira forma (0,6053) apresentou menor valor do que o painel de 20k personalizado pela segunda (0,6108) e terceira forma (0,6095). Essa mesma diferença foi observada nos cenários em que se utilizou a densidade de 50k, em que a segunda e terceira forma de customização resultaram em acurácias iguais a 0,6451 e 0,6501, respectivamente. Esse resultado pode ser explicado pela participação do valor de MAF que foi considerado nesses painéis, o que também foi reportado por Bernardes et al. (2019), que ao comparar os painéis comerciais com os painéis personalizados em imputação na raça Nelore, observaram que a terceira maneira de personalização apresentou maior acurácia de imputação para as situações estudadas, sendo que esta forma de personalização considera dois fatores que afetam a acurácia de imputação, como o DL e MAF.

Ao comparar os painéis pela densidade de SNPs, nota-se que o cenário que era constituído pelo painel de 20k comercial, apresentou menor valor de acurácia (0,5198) quando comparado com o cenário formado pelo painel de 50k comercial (0,5389). Entre os painéis personalizados a diferença foi maior quando a densidade aumentou, sendo que os painéis 20k personalizado pela primeira, segunda e terceira forma apresentaram valores menores de acurácia (0,6053; 0,6108; 0,6095, respectivamente) quando comparado aos cenários constituídos pelo painel de 50k personalizado pela primeira, segunda e terceira forma (0,6426; 0,6451; 0,6501, respectivamente). Estes resultados podem ser justificados pelo fato de acurácia aumentar conforme a densidade do painel aumenta (BERRY et al., 2014).

### Conclusão

A acurácia de imputação apresentou valores superiores nos painéis personalizados quando comparados com os comerciais e também, conforme a densidade dos painéis eram maiores as acurácias tiveram seu valor aumentado. Verificou-se diferenças entre os painéis personalizados, indicando que a maneira como o painel é construído poderá interferir nas acurácias de imputação.

### Literatura citada

#### Artigos:

Bernardes P. A., Nascimento G. B. D., Savegnago R. P., Buzanskas M. E., Watanabe R. N., De Almeida R., Correia L., Coutinho L. L., Gondro C., Munari D. P. Evaluation of imputation accuracy using the combination of two high-density panels in Nelore beef cattle. **Scientific Reports**, 9, 17920.

Berry D. P., McClure M. C., Mullen M. P., Within- and across-breed imputation of high-density genotypes in dairy and beef cattle from medium- and low-density genotypes. **Journal of Animal Breeding and Genetics**, 131, 165–172.

Browning B. L., Browning S. R. A Unified Approach to Genotype Imputation and Haplotype-Phase Inference for Large Data Sets of Trios and Unrelated Individuals. **The American Journal of Human Genetics**, 84, 210-223.

Romero A. R. S., Estudos Genômicos Aplicados ao Melhoramento Genético de Bovinos. **UFGD/Programa de Pós - Graduação em Biologia Geral**, 2017.

Sargolzaei M., Chesnais J. P., Schenkel F. S. A new approach for efficient genotype imputation using information from relatives. **BMC Genomics**, 15, 478.