



**UNIVERSIDADE FEDERAL DA PARAÍBA  
CAMPUS II – AREIA-PB  
CENTRO DE CIÊNCIAS AGRÁRIAS  
BACHARELADO EM ZOOTECNIA**

**MARIA VICTÓRIA HENRIQUE GENUÍNO**

**ESTUDO DA ACURÁCIA DE IMPUTAÇÃO EM BOVINOS DA RAÇA CANCHIM**

**AREIA  
2021**

**MARIA VICTÓRIA HENRIQUE GENUÍNO**

**ESTUDO DA ACURÁCIA DE IMPUTAÇÃO EM BOVINOS DA RAÇA CANCHIM**

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Zootecnia da Universidade Federal da Paraíba, como requisito para obtenção do título de Bacharel em Zootecnia.

**Orientador:** Prof. Dr. Marcos Eli Buzanskas

**AREIA  
2021**

**Catálogo na publicação**  
**Seção de Catalogação e Classificação**

G341e Genuíno, Maria Victoria Henrique.

Estudo da acurácia de imputação em bovinos da raça  
Canchim / Maria Victoria Henrique Genuíno. -

Areia:UFPB/CCA, 2021.

22 f.

Orientação: Marcos Eli Buzanskas.

TCC (Graduação) - UFPB/CCA.

1. Zootecnia. 2. Bovino de corte. 3. Genômica. 4.  
Imputação. 5. Melhoramento animal. I. Buzanskas, Marcos  
Eli. II. Título.

UFPB/CCA-AREIA

CDU 636(02)

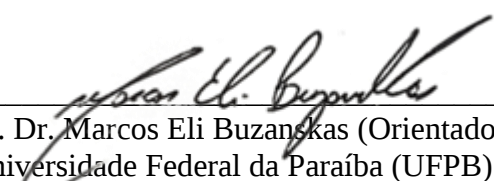
**MARIA VICTÓRIA HENRIQUE GENUÍNO**

**ESTUDO DA ACURÁCIA DE IMPUTAÇÃO EM BOVINOS DA RAÇA CANCHIM**

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Zootecnia da Universidade Federal da Paraíba, como requisito para obtenção do título de Bacharel em Zootecnia.

Aprovada em: 14/07/2021.

**BANCA EXAMINADORA**



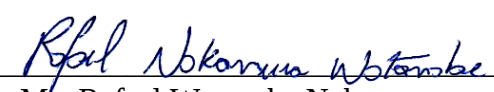
---

Prof. Dr. Marcos Eli Buzanskas (Orientador)  
Universidade Federal da Paraíba (UFPB)



---

Profa. Dra. Priscila Arrigucci Bernardes  
Universidade Federal de Santa Catarina (UFSC)



---

Me. Rafael Watanabe Nakamura  
Universidade Estadual Paulista (UNESP)

A minha família, por todo apoio e companheirismo.

## **AGRADECIMENTOS**

Agradeço a Deus e a Nossa Senhora da Penha, por toda força e apoio durante essa caminhada acadêmica.

Meus eternos agradecimentos ao meu orientador Marcos Eli Buzanskas, por ser tão solícito e me ajudar em todas as etapas da construção desse trabalho.

Aos professores do Curso da UFPB, em especial, Yirina, Alexandre, Bruna e Marcos, que foram meus orientadores de extensão, monitoria e Pibic, Lindomarcia, Adriana, Luciana, Edilson, Juliana, Ariosvaldo, Carla, Marcelo e Ludmila que contribuíram ao longo desses semestres, por meio das disciplinas e debates, para o meu desenvolvimento profissional.

Aos meus pais, Antônio Genuíno e Simone Henrique, por sempre acreditarem e apoiarem os meus sonhos. À minha irmã, Judithe Genuíno, por sempre acreditar no meu potencial. Ao meu namorado, Andrei Targino, por todo apoio durante a finalização desse processo acadêmico.

Aos colegas de classe, Glenda, Maria Isabelly, Antônio Cavalcanti, Camila, Laisy, Orlando e João, pelos momentos de amizade e apoio.

Aos meus amigos, Maria Luiza, Marina, Jéssica, Thayná, Carol, Yasmin, Julius e Emanuel, por sempre acreditarem em mim.

## RESUMO

Métodos de imputação visam prever informações de genótipos presentes em uma população alvo genotipada com painéis de menor densidade a partir de informações de uma população de referência genotipada com painéis de maior densidade, sendo utilizado para redução de custos de genotipagem e possível utilização na seleção genômica. Assim, o objetivo deste trabalho foi avaliar potenciais benefícios na incorporação de genótipos de raças fundadoras (Nelore e Charolês) para imputação em animais da raça composta Canchim. Foram utilizados genótipos de 285 animais da raça Canchim (CA) e de 114 animais do grupo genético MA (MA), sendo denominados de população alvo. As populações de referência foram compostas por 814 animais da raça Nelore (NE) e 897 animais da raça Charolês (CH) e em alguns casos, foram utilizados animais CA ou MA. A população de referência foi genotipada em painéis de alta densidade (BovineHD - BeadChip Illumina, com 777.962 SNPs) e a população alvo genotipada em painéis de baixa densidade (BovineSNP50 BeadChip Illumina, com 54.609 SNPs). Foram elaborados 17 cenários para analisar a acurácia de imputação por meio das medidas de taxa de concordância e  $r^2$  alélico. A variação observada entre os cenários de imputação avaliados por meio do  $r^2$  alélico foi de 0,59 (NE x MA) a 0,93 (CH+CA x MA). A incorporação de genótipos de animais da raça Charolês resultou em taxas de concordância para animais CA e MA que variaram de 70,93% a 73,61%, sendo superiores aos cenários nos quais animais Nelore foram utilizados, onde verificou-se taxas de concordância que variaram de 63,24% a 65,42%. Ao combinar as raças fundadoras na população de referência, foram observadas taxas de concordância que variaram de 83,47% a 85,99%, no entanto, as quantidades de marcadores na população de referência nestes cenários eram de cerca de metade em comparação aos cenários descritos anteriormente. A incorporação de genótipos das raças fundadoras promoveu aumento na acurácia de imputação para CA e MA. A inclusão da raça Charolês como população de referência em conjunto com animais Canchim, promoveu os maiores ganhos em acurácia observados. Pode-se inferir que a utilização de populações de referência multirraciais poderá auxiliar na redução de custos de genotipagem em raças compostas.

**Palavras-Chave:** bovino de corte; genômica; imputação; melhoramento animal.

## ABSTRACT

Imputation methods are applied to predict genotypes in a target population genotyped with lower density panels using genotypes from reference population genotyped with higher density panels, in order to reduce genotyping costs and usage in genomic selection. Thus, the aim of this study was to evaluate potential benefits of incorporating founder breed genotypes (Nelore and Charolais) for imputation in the Canchim composite breed. Genotypes from 285 Canchim (CA) and 114 MA genetic group (MA) individuals were used, being called as target population. The reference populations consisted of 814 Nelore (NE) and 897 Charolais (CH) animals, and in some cases, CA or MA animals were used. The panels considered were: BovineHD - BeadChip Illumina with 777,962 markers (reference population) and BovineSNP50 BeadChip Illumina with 54,609 markers (target population). A total of 17 scenarios were created to analyze the accuracy of imputation through concordance rate and  $r^2$  allelic. The variation observed between the imputation scenarios using the  $r^2$  allelic ranged from 0.59 (NE x MA) to 0.93 (CH+CA x MA). The integration of genotypes from Charolais resulted in concordance rates that ranged from 70.93% to 73.61%, being higher than the scenarios in which Nelore were used (varying from 63.24% to 65.42%). When combining the founders breed into the reference population, concordance rates ranging from 83.47% to 85.99% were observed; however, these scenarios presented half the quantity of markers in the reference populations when compared to the previous scenarios. The incorporation of founders breed genotypes increased the imputation accuracy for CA and MA. The inclusion of the Charolais as a reference population together with Canchim promoted the greatest gains in accuracy observed. It can be inferred that the use of multibreed reference populations may assist to reduce genotyping costs in composite breeds.

**Keywords:** animal breeding; beef cattle; genomics; imputation.

## SUMÁRIO

|                                                           |           |
|-----------------------------------------------------------|-----------|
| <b>1 INTRODUÇÃO .....</b>                                 | <b>8</b>  |
| <b>2 REVISÃO DE LITERATURA.....</b>                       | <b>10</b> |
| <b>2.1A raça Canchim .....</b>                            | <b>10</b> |
| <b>2.2 Seleção genômica e imputação de genótipos.....</b> | <b>11</b> |
| <b>3 METODOLOGIA.....</b>                                 | <b>14</b> |
| <b>4 RESULTADOS E DISCUSSÃO .....</b>                     | <b>16</b> |
| <b>5 CONCLUSÃO.....</b>                                   | <b>20</b> |
| <b>6 REFERÊNCIAS .....</b>                                | <b>21</b> |

## 1 INTRODUÇÃO

A raça Canchim, foi desenvolvida no Brasil com o intuito de combinar as características de rusticidade, adaptabilidade ao clima tropical e resistência à parasitas, presentes em raças zebuínas (*Bos taurus indicus*), com as características de rendimento de carcaça e produtividade, típicos de animais taurinos (*Bos taurus taurus*). Durante os estudos iniciais na década de 40, foi verificado que a composição racial de 3/8 zebuino e 5/8 taurino apresentou resultados satisfatórios em relação às demais composições avaliadas, para a formação inicial foram utilizadas as raças zebuínas Indubrasil, Nelore e Guzerá, enquanto que a raça taurina utilizada foi a Charolês (ALENCAR, 1988). Dentre os resultados observados neste animais, destaca-se a precocidade de desenvolvimento corporal e sexual, resistência ao calor, ótima conformação para corte e uniformidade de pelagem (VIANNA; PIMENTEL-GOMES; SANTIAGO, 1978). Estudos demonstraram os benefícios em utilizar o Canchim em cruzamentos com outras raças para obter progênies com melhor acabamento de gordura e velocidade de crescimento (KUSS *et al.*, 2008; PEROTTO; MOLETTA; LESSKIU, 2002; RUBIANO *et al.*, 2009).

O desenvolvimento científico e tecnológico na produção animal, em especial no campo da genética e melhoramento, tem auxiliado a identificar e intensificar o uso de animais com elevado potencial, com o objetivo de promover incrementos nos ganhos genéticos anuais. Neste sentido, a seleção genômica visa incorporar marcadores denominados de polimorfismos de nucleotídeos únicos (SNP), obtidos por meio de genotipagem, em conjunto com informações utilizadas na seleção tradicional para prever valores genéticos genômicos de indivíduos candidatos a seleção. Assim, a seleção genômica possibilita aumentar a resposta à seleção, diminuir o intervalo de geração, avaliar de forma mais acurada características de baixa herdabilidade e aumentar a acurácia da avaliação genética de animais candidatos à seleção (BOISON *et al.*, 2015). São utilizadas populações de referência que possuem informações de pedigree, de desempenho e dos SNPs para a predição dos efeitos dos marcadores e populações de candidatos que possuem apenas informações de pedigree e dos SNP. Além disto, os efeitos dos marcadores devem ser re-estimados com o passar das gerações, pois a população de referência será constantemente substituída (CALUS, 2010).

A genotipagem a larga escala apresenta custos elevados e, embora estes tenham decaído ao longo dos anos, ainda representam um dos fatores limitantes para a aplicação da seleção genômica. Opções com painéis de genotipagem de menores densidades e de custos mais acessíveis, possibilitaram o aumento do número de animais genotipados. A escolha da densidade destes painéis pode impactar na aplicabilidade da seleção genômica e de sua acurácia. Logo, uma alternativa seria utilizar estratégias de imputação, onde indivíduos da população de

referência são genotipados com o painel de alta densidade de marcadores, já a população de candidatos à seleção será genotipada com painéis de média ou baixa densidade. A imputação visa inferir marcadores não observados em animais genotipados em painéis de baixa densidade com base nos indivíduos genotipados em painéis de alta densidade. Utilizando os painéis de alta densidade é possível obter mais informações, por isso é vantajoso ter os animais em alta densidade. Como resultado desta metodologia, todos os indivíduos possuirão genótipos em alta densidade (BOISON *et al.*, 2015).

Diante do exposto, o objetivo deste estudo será avaliar a acurácia de imputação e os potenciais benefícios na incorporação de genótipos de raças fundadoras (Nelore e Charolês) para imputação em animais da raça Canchim.

## 2 REVISÃO DE LITERATURA

### 2.1A raça Canchim

Em 1940, estudos sobre cruzamentos foram realizadas nas dependências da Unidade Experimental de Pesquisa de Âmbito Estadual (UEPAE), localizada no município de São Carlos-SP, pertencente à Empresa Brasileira de Pesquisa Agropecuária (Embrapa), pelo pesquisador Dr. Antônio Teixeira Vianna. O intuito dos cruzamentos foi avaliar o desempenho de animais mestiços em comparação aos puros e obter indivíduos com composições genéticas adequadas ao sistema de produção brasileiro. Foram utilizados animais taurinos da raça Charolês, oriunda da França e que apresentavam grande musculosidade e precocidade de acabamento de carcaça; e zebuínos das raças Indubrasil, Nelore e Guzerá, oriundos da Índia e que apresentavam elevada rusticidade e adaptabilidade às condições de criação de clima tropical (ALENCAR, 1988).

No primeiro esquema avaliado, foram feitos cruzamentos entre reprodutores Charolês com fêmeas zebuínas, gerando o “meio-sangue”  $\frac{1}{2}$  Charolês  $\frac{1}{2}$  Zebu, em seguida indivíduos “meio-sangue” eram cruzados com zebuínos, gerando o  $\frac{3}{4}$  Zebu  $\frac{1}{4}$  Charolês e, por fim, os animais  $\frac{3}{4}$  Zebu  $\frac{1}{4}$  Charolês eram cruzados com animais Charoleses, gerando o  $\frac{5}{8}$  Charolês  $\frac{3}{8}$  Zebu. O cruzamento entre animais com composição  $\frac{5}{8}$  Charolês  $\frac{3}{8}$  Zebu resultou em progênes que possuíam maior produtividade, precocidade, rusticidade e adaptabilidade em relação às demais composições genéticas, sendo indicado para posteriores estudos (ALENCAR, 1988).

Os primeiros animais denominados de Canchim, nasceram em 1953 na Fazenda de Criação São Carlos. No ano de 1971, foi fundada a Associação Brasileira de Criadores de Canchim (ABCCAN) por criadores que tinham interesse de formar seus rebanhos com esta raça e em 1983, o Canchim foi reconhecido como raça pelo Ministério da Agricultura, Pecuária e Abastecimento. Foram utilizados diferentes esquemas de acasalamento para obtenção de novas linhagens Canchim, denominados de alternativo, absorvente, cruzado/absorvente e UEPAE. O objetivo destes esquemas foram o de ampliar a base genética da raça.

O esquema denominado “alternativo”, resulta em animais  $\frac{5}{8}$  Charolês  $\frac{3}{8}$  Zebu. O esquema “cruzado ou absorvente” utiliza touros Canchim em cruzamentos com vacas zebuínas, gerando o grupo genético “A”, que cruzadas com touros Canchim, geram o grupo genético “T1”, e que por sua vez, cruzadas novamente com touros Canchim, geram o grupo genético “V”. Por fim, o grupo genético “V” cruzado com touros Canchim resultará em progênes da raça Canchim.

O esquema “cruzado e absorvente” utiliza touros Charolês em cruzamentos com vacas zebuínas, gerando o grupo genético “T2”, que cruzadas com touros Canchim, geram o grupo genético “V”, e que por sua vez, cruzadas novamente com touros Canchim, geram progênie da raça Canchim. De forma semelhante, o esquema “UEPAE” utiliza reprodutores Canchim cruzados com vacas zebuínas, gerando o grupo genético “A”, este grupo genético é cruzado com reprodutores Charolês, gerando o grupo genético “MA”. Por fim, reprodutores do grupo genético “MA” ou animais Canchim podem ser cruzados com fêmeas “MA”, gerando progênie Canchim (ABCCAN, 2020). Todos os diferentes sistemas de cruzamentos irão resultar em animais Canchim com proporções Charolês-Zebu similares. Estudos de introgressão racial realizados por Buzanskas *et al.* (2017) confirmaram a existência de variabilidade presente em animais oriundos destes diferentes esquemas de cruzamentos.

A raça Canchim tem se consolidado no país, expandindo-se para as regiões Nordeste e Norte do Brasil, assim como para outros países da América Latina (MAIO, 2018). O Canchim tem sido utilizado em cruzamentos industriais como touro de repasse em propriedades que realizam inseminação artificial e em regiões onde o uso de animais taurinos puros não seria viável (ALENCAR, 2018). Além disto, a progênie oriunda destes cruzamentos apresenta elevado nível de heterose e, conseqüentemente, maiores ganhos em produtividade são esperados (ALENCAR; PACKER, 2005).

## 2.2 Seleção genômica e imputação de genótipos

Estudos realizados por Nejati-Javaremi, Smith e Gibson (1997) e Meuwissen, Hayes e Goddard (2001) indicaram que a utilização de uma densa quantidade de marcadores moleculares para a predição de valores genéticos genômicos poderia auxiliar no aumento da taxa de ganhos genéticos e redução do intervalo de geração em animais e plantas. Schaeffer (2006) indicou que a seleção genômica poderia resultar em maiores ganhos genéticos quando comparado à seleção tradicional, podendo aumentar a acurácia da avaliação genética de animais jovens e reduzir o intervalo de gerações, conseqüentemente, a intensidade de seleção seria aumentada.

Diante deste cenário e com o sequenciamento do genoma bovino (ELSIK *et al.*, 2009), observou-se a rápida evolução das tecnologias que resultaram na redução dos custos de genotipagem (MILLER, 2010) e no desenvolvimento de painéis de marcadores moleculares de diferentes densidades, aplicações e valores. O marcador molecular mais utilizado nestes painéis são os SNPs, sendo estes bi-alélicos, amplamente distribuídos no genoma das espécies e, muitas

vezes, funcionalmente relevantes para características produtivas dos animais (CAETANO, 2009).

Para a implementação da seleção genômica ou para outros estudos que utilizem informações de marcadores, os animais que são escolhidos para serem genotipados em alta densidade e até mesmo sequenciados são indivíduos de elevado valor genético para determinadas características de interesse econômico, também denominados de população referência. Já a população denominada como alvo, é composta por candidatos à seleção, ou seja, indivíduos jovens, com poucas progênies avaliadas e, até mesmo, sem idade para a avaliação de uma característica de interesse, sendo que estes animais poderiam ser genotipados com painéis de menores densidades para a redução de custos (VANRADEN, 2020).

Algoritmos de imputação de genótipos podem auxiliar na combinação de populações genotipadas com painéis de diferentes densidades. É possível inferir os marcadores não-observados em uma população genotipada com painéis de menores densidades utilizando informações de indivíduos genotipados com painéis de maiores densidades (SARGOLZAEI; CHESNAIS; SCHENKEL, 2014). A qualidade da imputação dependerá da densidade de marcadores utilizados, estrutura da população, quantidade de indivíduos nas populações de referência e alvo, a qualidade dos marcadores, do parentesco entre as populações e do método de imputação adotado (CHUD *et al.*, 2015; HUANG *et al.*, 2012; PAUSCH *et al.*, 2013).

De acordo com Carvalheiro *et al.* (2014), os métodos de imputação dependem parcialmente do desequilíbrio de ligação (associação não-aleatória dos *loci*) entre os marcadores para inferir genótipos não observados. Chud *et al* (2015) afirma que os métodos de imputação também dependem das informações de parentesco. Dentre os métodos disponíveis para imputação, o presente no programa FImpute (SARGOLZAEI; CHESNAIS; SCHENKEL, 2014) assume que todos os indivíduos considerados possuem algum grau de parentesco e, também, considera a montagem de janelas de segmentos cromossômicos (haplótipos) de diferentes tamanhos e que se sobrepõem, sendo utilizadas para predizer o genótipo faltante.

A imputação tem sido testada em diversas espécies e raças. Na raça Nelore, Carvalheiro *et al.* (2014) observaram acurácia de imputação, medida pela porcentagem de genótipos imputados corretamente, que variou 90,56 a 99,05%. Estes autores observaram que o cenário de melhor resultado era composto apenas por touros nas populações de referência, quando foram consideradas informações de parentesco e um painel customizado de 48 mil marcadores imputado para o painel de alta densidade. Em Gir leiteiro, Boison *et al.* (2015) estudaram diferentes cenários de imputação e observaram que a acurácia de imputação foi maior à medida que a densidade dos painéis LD eram maiores. Também observaram que melhores acurácias

eram observadas quando o número de animais na população de referência era maior, sendo a acurácia variou entre 0,957 a 0,988 quando a imputação foi realizada entre animais genotipados com o painel de 50 mil marcadores para a alta densidade.

Estudos com raças compostas ou multirraciais são mais complexos devido à composição genética dos indivíduos ser oriunda de duas ou mais raças. Oliveira Júnior *et al.* (2017) realizaram estudo com as raças Gir, Holandês e Girolando, sendo que a raça Girolando é composta e oriunda do cruzamento entre Gir e Holandês. Os autores verificaram que a população de referência formada pelas três raças e utilizada para imputar genótipos na população alvo formada apenas por animais Girolando resultou em acurácias que variaram 0,96 a 0,97. Na raça Canchim, Chud *et al.* (2015) avaliaram a acurácia de imputação utilizando diferentes painéis comerciais de baixas e médias densidades para a alta densidade e observaram maiores acurácias em painéis com mais de 50 mil marcadores. Além disto, dentre os cenários testados, maiores acurácias foram verificadas quando animais mais velhos eram considerados na população de referência, enquanto que os mais novos na população alvo, em que foi obtida acurácia de 97,74%.

### 3 METODOLOGIA

Foram utilizados genótipos de 285 animais da raça Canchim (CA) e de 114 animais do grupo genético MA (MA), sendo denominados como populações alvo para a imputação. A população alvo constituída por CA e MA em um mesmo conjunto de dados foi denominada de CAN. Já as populações de referência foram compostas por 814 animais da raça Nelore (NE) e 897 animais da raça Charolês (CH). Em algumas análises, animais CA e MA também foram considerados na população de referência. Os dados de CA, MA e NE foram fornecidos pela Embrapa Pecuária Sudeste (São Carlos/SP), enquanto que os dados de CH foram fornecidos pela TEAGASC - Athenry Animal & Grassland Research Centre (Athenry, Irlanda). Todos os indivíduos foram genotipados com o painel de alta densidade BovineHD - BeadChip Illumina (HD), que possui 777.962 SNPs.

Para cada cenário foi feito o controle de qualidade dos genótipos, realizado por meio do programa PLINK v.1.9. (CHANG *et al.*, 2015), considerou a exclusão de SNPs e de indivíduos com taxa de leitura inferior a 90%, frequência de alelos raros inferiores a 1% e SNPs com desvios significativos do equilíbrio de Hardy-Weinberg ( $p < 0.0001$ ). Apenas SNPs com posições conhecidas e localizados nos autossomos foram utilizados, de acordo com o mapa ARS-UCD1.2 (ROSEN *et al.*, 2020).

Para a elaboração dos cenários de imputação, os genótipos em HD foram reduzidos para a densidade do painel BovineSNP50 BeadChip Illumina (50k), com 54.609 SNPs, apenas para a população alvo. As populações de referência permaneceram com genótipos em alta densidade, alternando-se as raças (NE, CH, ou NE+CH), onde também foram incluídos os genótipos de CA ou MA para cenários específicos. A quantidade de SNPs e de indivíduos após o controle de qualidade, para cada cenário avaliado, é apresentada na Tabela 1.

Utilizou-se o programa FImpute v.3 (SARGOLZAEI; CHESNAIS; SCHENKEL, 2014) para realizar a imputação de genótipos faltantes, o qual considera informações de desequilíbrio de ligação entre os marcadores para identificação de haplótipos longos e curtos. Estes haplótipos estão, respectivamente, relacionados ao parentesco próximo ou distante entre os indivíduos. Não foram utilizados dados de pedigree para as populações estudadas.

A acurácia de imputação foi avaliada por meio das medidas de taxa de concordância (TC) e correlação de Pearson ( $r^2$  alélico) utilizando o programa computacional R. Para TC, foi verificada a porcentagem de genótipos imputados corretamente ao comparar genótipos presentes no painel HD original e imputado. O  $r^2$  alélico por marcador pode ser definido como:

$$r^2 = \frac{\sum_{k=1}^{N_j} (g_{jk} - \bar{g})(\hat{g}_{jk} - \bar{\hat{g}})}{\sqrt{\sum_{k=1}^{N_j} (g_{jk} - \bar{g})^2 \sum_{k=1}^{N_j} (\hat{g}_{jk} - \bar{\hat{g}})^2}}$$

onde  $N$  é o número total de marcadores sendo imputados;  $g_{jk}$  e  $\hat{g}_{jk}$  são os genótipos observados e imputados, respectivamente, para o SNP  $j$  do indivíduo  $k$ ; e  $\bar{g}$  e  $\bar{\hat{g}}$  são os valores médios para os genótipos observados e imputados, respectivamente.

#### 4 RESULTADOS E DISCUSSÃO

A variação observada entre os cenários de imputação avaliados por meio do  $r^2$  alélico foi de 0,59 (C5: NE x MA) a 0,93 (C11: CH+CA x MA) (Tabela 1). Os mesmos cenários apresentaram as menores (63,24%) e maiores (92,16%) taxas de concordância. Já o cenário C16 (NE+CH+CA x MA) apresentou a segunda maior acurácia. A inclusão de genótipos CH na população de referência em conjunto com CA (cenário C11) contribuiu com o aumento de 0,02 e 2,23% para o  $r^2$  alélico e TC, respectivamente, quando comparado ao cenário C1 (CA x MA). Já entre os cenários C2 (MA x CA) e C12 (CH+MA x CA), houve aumento de 0,04 e 4,49% para o  $r^2$  alélico e TC, respectivamente. O número de marcadores, a maior quantidade de animais presentes na população de referência e a incorporação de genótipos de animais CH poderá promover benefícios na acurácia de imputação para animais Canchim e MA. Devido à maior contribuição do Charolês na constituição genética do Canchim e MA, pode-se supor que existam haplótipos longos entre as raças, desta forma, CH contribuiu diretamente para este resultado.

A incorporação de apenas NE (C3, C4 e C5) e CH (C8, C9 e C10) como populações de referência para a imputação de CAN, CA e MA, não resultou em acurácias de imputação elevadas. No entanto, o  $r^2$  alélico e TC para C8, C9 e C10 (apenas CH na população de referência) foram cerca de 0,11 e 8% superiores quando comparados com C3, C4 e C5 (apenas NE na população de referência). Estes resultados também podem indicar que, devido à maior contribuição genética de CH com Canchim e MA, são esperadas maiores acurácias de imputação. A combinação das raças fundadoras na população de referência (cenários C13, C14 e C15) resultou no aumento das acurácias de imputação quando comparados com os cenários em que foram utilizadas as raças fundadoras separadamente. Logo, haplótipos curtos que estejam relacionados ao parentesco distante entre os indivíduos, ou seja, entre Canchim e MA com Nelore, podem ter contribuído para este resultado. Deve-se destacar que, nos cenários C13, C14, C15, C16 e C17, a combinação de NE e CH na população de referência resultou em aproximadamente 300 mil marcadores após o controle de qualidade de genótipos. Este fato pode ter contribuído para as acurácias observadas nestes cenários, uma vez que nos demais, o número de marcadores nas populações de referência eram de aproximadamente 600 mil.

Os resultados do presente estudo podem ser comparados ao estudo semelhante de Oliveira Júnior *et al.*, (2017), que utilizou as raças Gir, Holandês e Girolando (raça composta oriunda do cruzamento entre as raças puras mencionadas), em que verificou, ao utilizar animais Girolando, Gir e Holandês, separadamente, como população de referência (HD) para imputar genótipos em animais Girolando (50k), o  $r^2$  alélico foi igual a 96%, 59% e 74%,

respectivamente, enquanto que TC foi igual a 0,96, 0,51 e 0,75, respectivamente. Encontramos resultados semelhantes no presente estudo, onde os cenários C1 (CA x MA), C3 (NE x CAN) e C8 (CH x CAN) resultaram em  $r^2$  alélico de 89,93%, 65,13% e 72,97%, respectivamente, e TC igual a 0,91, 0,61 e 0,72, respectivamente.

Tabela 1. Cenários de imputação (C1 a C17),  $r^2$  alélico\*, taxa de concordância#, número de marcadores nas populações referência e alvo (em negrito e itálico, respectivamente) e número de indivíduos nas populações referência e alvo (entre parênteses, em negrito e itálico, respectivamente).

| População de referência | População alvo                                                       |                                                                      |                                                                      |
|-------------------------|----------------------------------------------------------------------|----------------------------------------------------------------------|----------------------------------------------------------------------|
|                         | CAN                                                                  | CA                                                                   | MA                                                                   |
| CA                      | -                                                                    | -                                                                    | C1: 0,91* - 89,93#<br><b>705.272 (283)</b><br><i>39.490 (112)</i>    |
| MA                      | -                                                                    | C2: 0,88* - 86,25#<br><b>658.480 (112)</b><br><i>39.490 (283)</i>    | -                                                                    |
| NE                      | C3: 0,61* - 65,13#<br><b>654.258 (809)</b><br><i>35.185 (395)</i>    | C4: 0,61* - 65,42#<br><b>654.258 (809)</b><br><i>35.277 (283)</i>    | C5: 0,59* - 63,24#<br><b>654.258 (809)</b><br><i>33.814 (112)</i>    |
| NE+CA                   | -                                                                    | -                                                                    | C6: 0,91* - 89,61#<br><b>654.258 (1.092)</b><br><i>35.185 (112)</i>  |
| NE+MA                   | -                                                                    | C7: 0,87* - 86,17#<br><b>607.043 (921)</b><br><i>35.233 (283)</i>    | -                                                                    |
| CH                      | C8: 0,72* - 72,97#<br><b>643.733 (897)</b><br><i>39.669 (395)</i>    | C9: 0,72* - 73,61#<br><b>643.733 (897)</b><br><i>39.782 (283)</i>    | C10: 0,73* - 70,93#<br><b>643.733 (897)</b><br><i>38.275 (112)</i>   |
| CH+CA                   | -                                                                    | -                                                                    | C11: 0,93* - 92,16#<br><b>635.757 (1.180)</b><br><i>37.969 (112)</i> |
| CH+MA                   | -                                                                    | C12: 0,92* - 90,74#<br><b>705.567 (1.009)</b><br><i>40.935 (283)</i> | -                                                                    |
| NE+CH                   | C13: 0,86* - 84,27#<br><b>301.572 (1.706)</b><br><i>19.965 (395)</i> | C14: 0,85* - 83,47#<br><b>301.572 (1.706)</b><br><i>19.857 (283)</i> | C15: 0,88* - 85,99#<br><b>301.572 (1.706)</b><br><i>19.369 (112)</i> |
| NE+CH+CA                | -                                                                    | -                                                                    | C16: 0,93* - 91,53#<br><b>306.859 (1.989)</b><br><i>19.748 (112)</i> |
| NE+CH+MA                | -                                                                    | C17: 0,89* - 87,72#<br><b>310.726 (1.818)</b><br><i>20.514 (283)</i> | -                                                                    |

CA = Canchim, MA = grupo genético MA, CAN = Canchim e MA, NE = Nelore, CH = Charolês, C1 a C17 = Cenários de imputação.

Cenários que combinaram as diferentes raças na formação das populações de referência foram estudados por Oliveira Júnior *et al.* (2017), que utilizaram animais Girolando e Gir,

Girolando e Holandês, Gir e Holandês, e as três raças juntas. Estes autores observaram nestes cenários,  $r^2$  alélico que variou de 0,94 a 0,97 e TC que variou de 93% a 96%. Os resultados mais próximos aos obtidos por estes autores foram observados nos cenários C11 (CH+MA x CA), iguais a 0,93 e 92,16%, e C12 (CH+CA x MA), iguais a 0,92 e 90,74%.

Piccoli *et al.* (2014) avaliaram a acurácia de imputação na raça Hereford e Braford (raça composta oriunda de cruzamentos entre Hereford e Nelore) e utilizaram populações de referência e alvo multirraciais, compostas pelas raças Braford, Hereford e Nelore. Os autores observaram  $r^2$  alélico e TC iguais a 0,88 e 93%, respectivamente, ao imputarem do painel 50k para HD. Além disto, não foi encontrada diferença na acurácia de imputação quando à população de referência incluía animais da raça Nelore. Ao testarem diferentes cenários de imputação na raça Canchim, Chud *et al.* (2015) verificaram maiores acurácias de imputação ao definir que a população de referência seria formada por animais nascidos antes de 2005 (296 animais), enquanto que animais nascidos a partir deste ano iriam compor a população alvo (99 animais), sendo observado  $r^2$  alélico e TC ao redor de 0,94 e 96%.

Em estudo com as raças Angus e Charolês e com uma população multirracial composta por animais oriundos de cruzamentos entre as raças Angus, Charolês, Galloway, Hereford, Simental, Holandês e Pardo-Suíço. Ventura *et al.* (2014) verificaram que a utilização de uma população de referência multirracial em conjunto com as raças puras resultou em maiores acurácias de imputação do painel 6k (6.909 SNPs) para 50k, onde os cenários contendo a população alvo Angus e a população alvo da estação experimental Elora (animais com composição genética de 46,86% Angus, 24,77% Simental, 6,01% Piemontês, 5,05% Charolês, 4,14% Gelbvieh, 1,93% Limousin e 11,24% de outras raças) apresentaram TC iguais a 97,93% e 97,53%, respectivamente.

Ao realizarem a imputação utilizando os painéis 50k e HD customizado (denominado de 850k), Rowan *et al.* (2019) observaram acurácia de imputação de acima de 0,97 ao utilizarem populações multirraciais. Os autores concluíram que o aumento da acurácia ocorreu devido número de indivíduos na população de referência, sugerindo que isto beneficiará populações alvo pouco representativas numericamente, ainda mais quando estas são raças compostas. Além disto, foi indicado que alelos raros foram acuradamente imputados nos cenários avaliados.

Li, Sargolzaei e Schenkel (2014) avaliaram a acurácia de imputação do painel HD para a sequência completa do genoma utilizando as raças Angus, Pardo-Suíço, Holandês, Jersey, Limousin e Simental. Ao compor a população de referência considerando múltiplas raças sequenciadas, os autores observaram TC média e  $r^2$  alélico médio para todas as raças imputadas (populações alvo) igual a 92,90% e 0,89, respectivamente. Como conclusões, indicaram que

existem benefícios ao utilizar a população multirracial como referência na imputação do painel HD para a sequência completa do genoma, principalmente em raças que possuem poucos indivíduos sequenciados.

## 5 CONCLUSÃO

A incorporação de genótipos das raças fundadoras promoveu aumento na acurácia de imputação para animais Canchim e MA. A inclusão da raça Charolês como população de referência em conjunto com animais Canchim, promoveu os maiores ganhos em acurácia observados. Pode-se inferir que a utilização de populações de referência multirraciais poderá auxiliar na redução de custos de genotipagem em raças compostas. O número de marcadores nas populações de referência e o número de animais utilizados, além da presença dos genótipos de Charolês, podem ter interferido nos resultados das acurácias de imputação.

## 6 REFERÊNCIAS

- ABCCAN. **Associação Brasileira de Criadores de Canchim**. Disponível em: <<http://www.abccan.com.br/canchim/index.php/a-raca.html>>. Acesso em: 5 dez. 2020.
- ALENCAR, M. M. **Bovino - Raça Canchim: Origem e Desenvolvimento**. Brasília: Embrapa-DMU, 1988.
- ALENCAR, M. M. Touro Canchim no cruzamento industrial – resultados de pesquisas com uso do Canchim. **Anais da V Convenção Nacional da Raça Canchim**, p. 19–26, 2018.
- ALENCAR, M. M.; PACKER, I. U. Competitividade depende do cruzamento entre raças. **Visão agrícola**, v. 2, p. 11–13, 2005.
- BOISON, S. A. et al. Strategies for single nucleotide polymorphism (SNP) genotyping to enhance genotype imputation in Gyr (*Bos indicus*) dairy cattle: Comparison of commercially available SNP chips. **Journal of Dairy Science**, New York, v. 98, n. 7, p. 4969–4989, 2015.
- BUZANSKAS, M. E. et al. Study on the introgression of beef breeds in Canchim cattle using single nucleotide polymorphism markers. **PLOS ONE**, San Francisco, v. 12, n. 2, p. e0171660, 2017.
- CAETANO, A. R. Marcadores SNP: conceitos básicos, aplicações no manejo e no melhoramento animal e perspectivas para o futuro. **Revista Brasileira de Zootecnia**, Viçosa, v. 38, p. 64–71, jul. 2009.
- CALUS, M. P. L. Genomic breeding value prediction: methods and procedures. **Animal**, Amsterdam, v. 4, n. 2, p. 157–164, 2010.
- CARVALHEIRO, R. et al. Accuracy of genotype imputation in Nelore cattle. **Genetics Selection Evolution**, London, v. 46, n. 1, p. 69, 2014.
- CHANG, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. **GigaScience**, Oxford, v. 4, n. 7, 2015.
- CHUD, T. C. S. et al. Strategies for genotype imputation in composite beef cattle. **BMC Genetics**, London, v. 16, n. 1, p. 99, 2015.
- ELSIK, C. G. et al. The Genome Sequence of Taurine Cattle: A window to ruminant biology and evolution. **Science**, Washington, v. 324, n. 5926, p. 522–528, 2009.
- HUANG, Y. et al. Assessment of alternative genotyping strategies to maximize imputation accuracy at minimal cost. **Genetics Selection Evolution**, London, v. 44, n. 1, p. 25, 2012.
- KUSS, F. et al. Carcaça e carne de novilhos cruzas Pardo Suíço x Canchim e Purunã x Canchim terminados em confinamento. **Ciência Rural**, Santa Maria, v. 38, n. 4, p. 1061–1066, ago. 2008.
- LI, H.; SARGOLZAEI, M.; SCHENKEL, F. S. **Accuracy of whole-genome sequence genotype imputation in cattle breeds**. 10th World Congress on Genetics Applied to Livestock Production. **Anais...**Vancouver: 2014
- MAIO, A. Mapa mostra expansão da raça Canchim. **Embrapa Pecuária Sudeste**, 23 jul. 2018.
- MEUWISSEN, T. H. E.; HAYES, B. J.; GODDARD, M. E. Prediction of total genetic value using genome-wide dense marker maps. **Genetics**, London, v. 157, n. 4, p. 1819–1829, 2001.
- MILLER, S. Genetic improvement of beef cattle through opportunities in genomics. **Revista**

**Brasileira de Zootecnia**, Viçosa, v. 39, p. 247–255, 2010.

NEJATI-JAVAREMI, A.; SMITH, C.; GIBSON, J. P. Effect of total Allelic Relationship on Accuracy of Evaluation and Response to Selection. **Journal of Animal Science**, Cary, v. 75, p. 1738–1745, 1997.

OLIVEIRA JÚNIOR, G. A. et al. Genotype imputation in a tropical crossbred dairy cattle population. **Journal of Dairy Science**, New York, v. 100, n. 12, p. 9623–9634, 2017.

PAUSCH, H. et al. Imputation of high-density genotypes in the Fleckvieh cattle population. **Genetics Selection Evolution**, London, v. 45, n. 1, p. 3, 2013.

PEROTTO, D.; MOLETTA, J. L.; LESSKIU, C. Desempenho em confinamento de machos bovinos inteiros Canchim, Aberdeen angus e cruzamentos recíprocos. **Ciência Rural**, Santa Maria, v. 32, n. 4, p. 669–673, 2002.

PICCOLI, M. L. et al. Accuracy of genome-wide imputation in Braford and Hereford beef cattle. **BMC Genetics**, London, v. 15, n. 1, p. 157, 2014.

ROSEN, B. D. et al. De novo assembly of the cattle reference genome with single-molecule sequencing. **GigaScience**, Oxford, v. 9, n. 3, 2020.

ROWAN, T. N. et al. A multi-breed reference panel and additional rare variants maximize imputation accuracy in cattle. **Genetics Selection Evolution**, London, v. 51, n. 1, p. 77, 2019.

RUBIANO, G. A. G. et al. Desempenho, características de carcaça e qualidade da carne de bovinos superprecoces das raças Canchim, Nelore e seus mestiços. **Revista Brasileira de Zootecnia**, Viçosa, v. 38, n. 12, p. 2490–2498, 2009.

SARGOLZAEI, M.; CHESNAIS, J. P.; SCHENKEL, F. S. A new approach for efficient genotype imputation using information from relatives. **BMC Genomics**, London, v. 15, n. 1, p. 478, 2014.

SCHAEFFER, L. R. Strategy for applying genome-wide selection in dairy cattle. **Journal of Animal Breeding and Genetics**, Hoboken, v. 123, n. 4, p. 218–223, 2006.

VANRADEN, P. M. Symposium review: How to implement genomic selection. **Journal of Dairy Science**, New York, v. 103, n. 6, p. 5291–5301, 2020.

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**, Cary, v. 92, n. 4, p. 1433–1444, 2014.

VIANNA, A. T.; PIMENTEL-GOMES, F.; SANTIAGO, M. **Formação do Gado Canchim pelo Cruzamento Charolês-Zebu**. São Paulo: Nobel, 1978.