

**UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO DE CIÊNCIAS AGRÁRIAS CURSO DE
ZOOTECNIA**

GABRIELA LEONI

**IMPUTAÇÃO UTILIZANDO PAINÉIS
PERSONALIZADOS EM BOVINOS DA RAÇA
CANCHIM**

**FLORIANÓPOLIS - SC
2022**

**UNIVERSIDADE FEDERAL DE SANTA CATARINA CENTRO
DE CIÊNCIAS AGRÁRIAS CURSO DE ZOOTECNIA**

GABRIELA LEONI

**IMPUTAÇÃO UTILIZANDO PAINÉIS PERSONALIZADOS
EM BOVINOS DA RAÇA CANCHIM**

Trabalho Trabalho de Conclusão de Curso apresentado como exigência para obtenção do Diploma de Graduação em Zootecnia da Universidade Federal de Santa Catarina.

Orientador (a): Prof.^a. Priscila Arigucci Bernardes.

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

LEONI, GABRIELA
IMPUTAÇÃO UTILIZANDO PAINÉIS PERSONALIZADOS EM BOVINOS
DA RAÇA CANCHIM / GABRIELA LEONI ; orientador, Priscila
Arrigucci Bernardes, 2022.
42 p.

Trabalho de Conclusão de Curso (graduação) -
Universidade Federal de Santa Catarina, Centro de Ciências
Agrárias, Graduação em Zootecnia, Florianópolis, 2022.

Inclui referências.

1. Zootecnia. 2. Bovinos de Corte. 3. Painéis
Personalizados. 4. SNP. I. Arrigucci Bernardes, Priscila.
II. Universidade Federal de Santa Catarina. Graduação em
Zootecnia. III. Título.

Gabriela Leoni

IMPUTAÇÃO UTILIZANDO PAINÉIS PERSONALIZADOS EM BOVINOS DA RAÇA CANCHIM

Esta Monografia de Trabalho de Conclusão de Curso foi julgada aprovada e adequada para obtenção do grau de Zootecnista.

Florianópolis, 21 de novembro de 2022.

Banca Examinadora:

Prof.^a Priscila Arrigucci Bernardes, Dr.^a
Orientadora
Universidade Federal de Santa Catarina

Prof. André Luís Ferreira Lima, Dr.
Universidade Federal de Santa Catarina

Prof.^a Márcio Cinachi Pereira, Dr.^a
Universidade Federal de Santa Catarina

DEDICATÓRIA

Dedico esse trabalho aos meus pais, Raquel de F. S. Leoni e Laércio J. Leoni, com todo o amor e gratidão pelo apoio nessa jornada.

AGRADECIMENTOS

Gostaria de agradecer a Deus, por sempre me guiar e me dar forças em toda essa caminhada,

Agradecer a minha família, em especial os meus pais, Raquiel e Laércio, por terem me proporcionado tudo de mais precioso que eu poderia ter, por não terem medido esforços em promover a educação que recebi e darem total apoio em toda a minha jornada, serei eternamente grata pelo modo como trabalharam para garantir que eu tivesse inúmeras oportunidades.

Ao meu namorado Matheus, que sempre me deu total apoio e suporte, entendeu todas as minhas ausências e me faz muito feliz.

A minha professora orientadora Priscila, que me proporcionou a oportunidade de conduzir o presente estudo, por me receber como aluna e orientada, tendo muita paciência e dedicação, por todos os ensinamentos e conselhos.

Por fim, sou grata a todos que de alguma forma, direta ou indiretamente, participaram da realização desse trabalho.

Muito obrigada!

A persistência é o menor
caminho do êxito.
(Charles Chaplin)

RESUMO

Para adequada resposta à seleção genômica, torna-se necessário obter grande quantidade de animais genotipados com muitos SNPs e a utilização de todos os animais da população genotipados com painéis de alta densidade ainda exige custo elevado. Dessa maneira, genotipar animais candidatos jovens com painéis de menor custo, como os de baixa densidade e então, realizar a imputação dos marcadores não conhecidos utilizando a informação de uma população referência genotipada com painéis de alta densidade, tornou-se uma alternativa para redução de custos para aplicação desta tecnologia. Para raças em que a população não é grande o suficiente, há indisponibilidade de grande quantidade de animais genotipados com alta densidade. Além disso, os painéis de baixa densidade comerciais, podem não favorecer a imputação para painéis de alta densidade comerciais. Dessa maneira, o objetivo deste estudo foi comparar a acurácia de imputação utilizando painéis comerciais de baixa densidade e diferentes painéis personalizados de baixa densidade para um painel de alta densidade comercial, além de verificar a acurácia ao incorporar genótipos de uma das raças fundadoras na população referência para imputação em animais da raça Canchim. Foram utilizados 801 animais Nelore e 396 animais Canchim e MA (resultante do cruzamento entre o grupo genético A e animais Charolês, sendo o grupo genético A resultante do cruzamento entre animais Canchim e Nelore) em 16 diferentes cenários, em que a população de animais referência é a população de animais que foram mantidos todos os SNPs do painel de alta densidade. Os demais animais da população tiveram seus SNPs “escondidos” com a finalidade de mimetizar que estes animais foram genotipados com painéis comerciais de baixa densidade contendo 20 mil SNPs (19.263k) ou 50 mil SNPs (26.342k), e considerou-se também 3 painéis personalizados, em que a partir do painel de 700k dos animais da raça Canchim e grupo MA, foram escolhidos diferentes SNPs formando painéis contendo a mesma quantidade de SNPs dos painéis comerciais de baixa densidade, ou seja, 19.263 (20k) e 26.342 SNPs (50k). A acurácia de imputação teve 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 os painéis personalizado 1 e o personalizado 2 e 3, indicando que a maneira como o painel é construído interfere nas acurácias de imputação. Apesar de ter sido observada acurácias moderadas, utilizar raças fundadoras na população referência pode ser uma alternativa para raças em que a população não é grande o suficiente.

Palavras chave: Bovinos de corte. Illumina. Painéis personalizados. SNPs.

LISTA DE FIGURAS

Figura 1. Esquema representando como foi realizada a terceira maneira de personalização.....	12
Figura 2. Taxa de concordância por cromossomo considerando os cenários Alto Parentesco com 20k.....	1

LISTA DE TABELAS

Tabela 1. Cenários utilizados para o estudo de imputação para os painéis BovineHD BeadChip (700k).....	11
Tabela 2. Número de SNPs que foram imputados (Nº de SNPs), média da taxa de concordância (TC) e média da correlação (Cor) para todos os animais e, considerando os dois grupos raciais (Canchim e MA), para os diferentes cenários.....	15
Tabela 3. Número de SNPs que foram imputados (Nº de SNPs), média da taxa de concordância (TC) e média da correlação (Cor) para todos os animais e, considerando os dois grupos raciais (Canchim e MA), para os diferentes cenários.....	16

SUMÁRIO

1. INTRODUÇÃO.....	1
2. OBJETIVO.....	2
3. REVISÃO BIBLIOGRÁFICA.....	3
3.1. Raça Canchim.....	3
3.2. Imputação.....	5
3.2.1. <i>Descrição da imputação</i>	5
3.2.2. <i>Desequilíbrio de Ligação</i>	6
3.2.2. <i>Frequência do Alelo de Menor Frequência</i>	7
3.3. Acurácia de imputação.....	7
4. METODOLOGIA.....	9
4.1. Animais e genotipagem.....	9
4.2. Imputação.....	10
4.2.1. <i>Cenários utilizados</i>	10
4.2.2. <i>Algoritmos para Imputação</i>	13
4.2.3. <i>Acurácia de imputação</i>	13
5. RESULTADOS E DISCUSSÃO.....	14
6. CONCLUSÕES.....	21
7. REFERÊNCIAS BIBLIOGRÁFICAS.....	21

1. INTRODUÇÃO

As tecnologias fundamentadas em informações genômicas favorecem o desenvolvimento científico na produção animal, principalmente no âmbito da genética e melhoramento. Estas tecnologias têm auxiliado programas de melhoramento a aprimorar inúmeras características, como taxa de crescimento, composição de produtos de origem animal, fertilidade, taxa de sobrevivência e resistência à doenças, entre outras. 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 explorados pela seleção genômica, que é uma técnica de obtenção dos valores genéticos utilizando informação de fenótipos, pedigree e marcadores genéticos, necessitando de um grande número de indivíduos genotipados com maior quantidade de marcadores para que os valores genéticos estimados tenham elevada acurácia (CHUD et al., 2015). Portanto, a seleção genômica viabiliza aumentar a resposta à seleção, reduzir o intervalo de geração, examinar de forma mais acurada características de baixa herdabilidade e ainda, elevar a acurácia da avaliação genética de animais candidatos à seleção (BOISON et al., 2015).

Para adequada resposta à seleção genômica, torna-se necessário obter grande quantidade de animais genotipados com muitos marcadores (SNPs) e embora tenha sido constatado uma redução no custo da genotipagem nos últimos anos, a utilização de todos animais da população genotipados com muitos marcadores (painéis de alta densidade) ainda exige custo elevado, tornando por muitas vezes inviável para os produtores e pesquisadores. Dessa maneira, uma alternativa para baratear o custo da aplicação da seleção genômica seria realizar a imputação dos genótipos. A imputação é uma ferramenta que envolve a genotipagem em maior densidade de uma parcela de indivíduos na população, chamada de referência, enquanto outra parcela dos indivíduos é genotipada com um painel SNPs de menor densidade, chamada de população de imputação. As informações da população referência são utilizadas para imputar os genótipos de marcadores ausentes de indivíduos genotipados com painéis de menor densidade.

Contudo, a imputação é influenciada pela estrutura da população, o número de SNPs no painel de imputação, o nível de parentesco entre a população de referência e de imputação, o tamanho efetivo da população e ainda, a precisão inerente do método usado para imputação e o grau aos quais os marcadores são ordenados corretamente pelo mapa do genoma (HICKEY et al. 2012; HOZÉ et al.2013; UEMOTO et al. 2015). Todos esses fatores irão influenciar na acurácia da imputação, ou seja, a proximidade do resultado da imputação, com o seu valor real, assim, quanto maior for a acurácia, maior será a precisão da imputação de genótipos.

Assim, devido às diferenças de estruturas populacionais, observa-se diferenças na acurácia de imputação de acordo com a raça utilizada. Em países tropicais, nota-se a criação de algumas raças de produção adaptadas ao local e de menor expressão mundial. O bovino de corte Canchim é um exemplo, sendo uma raça desenvolvida no Brasil com o propósito de reunir 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, próprios de animais taurinos (*Bos tauros tauros*) (ALENCAR, 1988).

Para raças em que a população não é grande o suficiente, há indisponibilidade de grande quantidade de animais genotipados com alta densidade. Além disso, os painéis de baixa densidade comerciais, podem não favorecer a imputação para painéis de alta densidade comerciais. A principal diferença entre painéis personalizados e painéis comerciais está relacionada com a construção do painel, painéis comerciais são construídos por empresas como Illumina e Affymetrix, onde determinam os critérios e normalmente são utilizadas populações taurinas para sua construção. Já painéis personalizados, são construídos e elaborados por pesquisadores e conhecedores da técnica para pesquisas e aplicações específicas, onde, estabelecem critérios próprios de construção e também podem utilizar populações zebuínas ou cruzadas.

2. OBJETIVO

O objetivo deste estudo foi comparar a acurácia de imputação

utilizando painéis comerciais de baixa densidade e diferentes painéis personalizados de baixa densidade para um painel de alta densidade comercial, além de verificar a acurácia da utilização de populações fundadoras na população referência e populações de animais cruzados na população de imputação.

3. REVISÃO BIBLIOGRÁFICA

3.1. Raça Canchim

O gado Zebu (*Bos indicus*) por ter grande aptidão em habitar os trópicos e ser um animal bastante rústico se desenvolveu muito bem no Brasil, porém tem inferioridade em rendimento de carne e precocidade quando comparado a raças europeias (MACHADO, 2015). Como características das raças europeias está a sua baixa resistência a climas tropicais, a elevada exigência quanto a qualidade de forragem e susceptibilidade aos parasitas, e com isso, esses animais não se adaptam muito bem às condições do Brasil. Assim, desenvolveu-se o interesse em obter uma raça que junte a rusticidade, fertilidade, pelo curto e liso, boa adaptação ao clima, às pastagens e aos parasitas do gado Zebu com o rendimento de carne e precocidade, ganho de peso, peso de carcaça, maciez da carne, libido e docilidade do gado europeu (MAIO, 2018).

Nesse contexto, em 1940, no estado de São Paulo, na fazenda Canchim, a qual hoje é a Unidade de Execução de Pesquisa de Âmbito Estadual da EMBRAPA, houve o desenvolvimento de acasalamentos para obtenção do bimestiço $5/8$ Charolês + $3/8$ Nelore, denominado posteriormente de Canchim. Porém, diferentes proporções de cruzamentos entre a raça Charolês e raças zebuínas podem estar presentes em animais Canchim devido aos vários esquemas de acasalamento que têm sido utilizados para expandir a base genética para esta raça (CHUD et al. 2015).

Dentre as diferentes proporções de acasalamento, existe o esquema “UEPAE”, que utiliza reprodutores Canchim cruzados com vacas zebuínas,

originando o grupo genético “A”, este grupo genético é cruzado com reprodutores Charolês, gerando o grupo genético “MA”. Por último, os reprodutores do grupo genético “MA” ou animais Canchim podem ser cruzados com fêmeas “MA”, gerando progênies Canchim (ABCCAN, 2020). Todos os diferentes sistemas de cruzamentos irão resultar em animais Canchim com proporções Charolês-Zebu semelhantes.

A raça Canchim tem se fortalecido no país, estendendo-se em maior volume para o Nordeste e Norte do Brasil, e também para outros países da América Latina (MAIO, 2018). Nota-se a utilização desses animais em cruzamentos industriais como touro de repasse em propriedades que efetuam a inseminação artificial e, também é utilizado em regiões onde não seria possível a criação de animais taurinos puros (ALENCAR, 2018).

No Brasil, muitos animais resultam de cruzamentos entre *Bos taurus taurus* e *Bos taurus indicus*. Esse fato proporciona grandes desafios para a imputação dos genótipos desses animais, sendo necessário o desenvolvimento de metodologias e estratégias de imputação que maximizem a precisão na população de interesse (CHUD et al. 2015).

O tamanho efetivo da população dentro do processo evolutivo e na genética quantitativa é considerado um parâmetro-chave, pois mediante este parâmetro se obtém a medida da taxa de oscilação genética e da endogamia da população. Se o tamanho efetivo da população em gerações anteriores é baixo, maior será o número de ancestrais comuns e, conseqüentemente, maior será o nível de endogamia (CARVAJAL, 2017). A população de Canchim é considerada pequena quando comparada a outras raças no Brasil. O rebanho é estimado em aproximadamente 30.000 animais (MOKRY et al. 2014). Porém, devido a formação recente da raça Canchim não se tem informações precisas sobre o tamanho efetivo da população, como também evidenciado por Buzanskas et al. (2017).

A frequência com que acasalamentos endogâmicos ocorrem em populações naturais depende do tamanho da população, sendo que em populações pequenas, após poucas gerações de acasalamento, qualquer par de indivíduos será relacionado através de pelo menos um ancestral comum (REGITANO, 1996). Alencar et al. (1981) analisaram o coeficiente de endogamia em animais da raça Canchim, possuindo uma média de 1,85%

para os machos e 1,87% para as fêmeas, e apesar do reduzido valor, o efeito da endogamia foi significativo para peso aos 205 dias, sendo que os machos foram mais afetados do que as fêmeas. Na população estudada neste trabalho, foi verificado uma média do coeficiente de endogamia de 0,1937%, tendo um coeficiente de endogamia alto justamente pela população analisada ser pequena.

3.2. Imputação

3.2.1. Descrição da imputação

A imputação trata-se de um método que envolve a genotipagem em maior densidade de uma parcela de indivíduos na população, chamada de referência, enquanto outra parcela dos indivíduos é genotipada com um painel SNPs de menor densidade, chamada de população de imputação. Os dados de genótipos da referência são utilizados para imputar os genótipos de marcadores ausentes nos indivíduos genotipados com painéis de menor densidade (TSAI et al, 2017). A imputação já foi aplicada com grande êxito em diversos programas de melhoramento para várias raças de bovinos (HAYES et al., 2012; HICKEY et al., 2012; PAUSCH et al., 2013; DAETWYLER et al., 2013; MOGHADDAR et al., 2015).

Ao realizar um estudo para quantificar a precisão da imputação de genótipos de um painel marcador de baixa densidade disponível comercialmente para um painel de marcadores de maior densidade disponível comercialmente em bovinos da raça Holandesa, os autores Berry e Kearney (2015) reportaram adequada precisão, afirmando que esta prática possui menor custo, facilitando assim uma redução no custo da seleção.

Contudo, a habilidade de imputação é influenciada pela estrutura da população, o número de SNPs no painel de imputação, o nível de parentesco entre a população de referência e de imputação, o tamanho efetivo da população e ainda, a precisão inerente do método usado para imputação e o grau aos quais os marcadores são ordenados corretamente pelo mapa do genoma (HICKEY et al., 2012; HOZÉ et al., 2013; UEMOTO et

al., 2015).

Diferentes estudos de imputação foram realizados em diferentes raças e em geral os valores de precisão são demonstrados em porcentagens, sendo que em Nelore, já foi relatado uma precisão de imputação superior a 97%, conforme estudo feito por Carvalheiro et al (2014). Os mesmo autores que reportaram essa habilidade de imputação, também compararam painéis personalizados com os comerciais, sendo que os personalizados obtiveram maiores precisões, ultrapassando os comerciais que tinham densidades semelhantes. Já em raças compostas torna-se mais difícil obter elevados valores de precisão de imputação, devido a composição genética. Oliveira Júnior et al (2017) verificaram que quando a população de referência formada por três raças (Gir, Holandês e Girolando) é utilizada para imputar genótipos na população alvo formada somente por animais Girolando, resulta em acurácias de 0,96 a 0,97.

Chud et al. (2015), verificando a acurácia de imputação em bovinos Canchim empregando diferentes painéis comerciais de baixas e médias densidades para a alta densidade, obtiveram maiores acurácias em painéis com mais de 50 mil marcadores. Estes autores também identificaram que as maiores precisões eram obtidas quando animais mais velhos eram considerados na população referência, pois animais mais velhos possuem maior distribuição do seu material genético quando comparado com animais mais novos.

3.2.2. *Desequilíbrio de Ligação*

O desequilíbrio de ligação (DL) é a associação não aleatória de alelos de diferentes loci. Este fenômeno está relacionado à ocorrência de recombinações ocorridas antes do início da genotipagem, sendo propriedade de uma população e da geração em que é medida (PEIXOTO et al, 2013). A exploração deste conhecimento é amplamente utilizada nas análises de seleção genômica . Estas análises buscam e estudam o desequilíbrio de ligação entre os marcadores e genes de interesse econômico para reconhecer regiões genômicas com potencial, ou sob influência, de seleção, e também a elaboração e formação de mapas de

desequilíbrio os quais são importantes para o estudo da diversidade genética dentro e entre raças (McKay et al., 2007). Os SNPs têm sido amplamente utilizados em estudos de DL (ZHANG et al., 2012).

Os mapas de desequilíbrio de ligação (LD) são ferramentas fundamentais para analisar e investigar a base genética de características economicamente importantes em bovinos. Os SNPs têm sido amplamente utilizados em estudos de DL (ZHANG et al., 2012), sendo que quantificar a extensão de LD no genoma bovino é um primeiro passo para estabelecer o número de marcadores que serão necessários para o mapeamento de QTL (locos de traços quantitativos) por desequilíbrio de ligação.

3.2.3. Frequência do Alelo de Menor Frequência

A frequência do alelo menos comum (MAF) é dada pela frequência em que este alelo ocorre em uma dada população, sendo que a MAF é amplamente utilizada em estudos de genética populacional porque fornece informações para diferenciar entre variantes comuns e raras na população. Assim, a MAF pode afetar a precisão da imputação, como visto no estudo de Wang et al. (2016), onde para todos os métodos, a precisão de imputação para genótipos portadores do alelo menor aumenta à medida que o MAF aumenta.

Abdol lahi-Arpanahi et al. (2014) revelou que os SNPs de baixa frequência com grandes efeitos podem fornecer melhor a capacidade preditiva para características quantitativas, incluindo peso corporal, medição ultra-sônica do músculo do peito, entre outras. Embora possam ser importantes, SNPs com baixa MAF demonstraram influenciar negativamente a precisão da imputação, pois os alelos que ocorrem em baixas frequências não são necessariamente apresentados nos haplótipos identificados (SCHROOTEN et al., 2014).

3.3. Acurácia de imputação

A acurácia é a precisão e exatidão de dados e informações, este termo é usado para definir a proximidade de um resultado experimental, com o seu valor real. Quanto maior for a acurácia, maior será a precisão da

imputação de genótipos. É importante avaliar a precisão da imputação para definir melhor qual método de imputação e painel são mais apropriados para aplicações genômicas (CARVALHEIRO et al., 2014).

Existem alguns fatores que podem interferir na acurácia da imputação dos genótipos, porém, segundo Hickey et al. (2012), o fator mais importante é o número de ancestrais imediatos genotipados na população referência. O número total de animais na população referência torna-se importante quando não há, ou há poucos, ancestrais imediatos com genótipos de menor densidade, pois ter poucos animais com genótipos no SNP a ser imputado, produz resultados de baixa acurácia, tornando a população limitada (WANG et al., 2012; DUARTE et al., 2013). Por outro lado, o impacto de ter apenas um pequeno número de animais na população referência torna a imputação limitada.

Outro fator que afeta a acurácia de imputação é o desequilíbrio de ligação (DL), para densidade muito baixa, o impacto do DL entre o SNP do painel de maior densidade e o SNP imputado no painel de baixa densidade também podem ser reduzido consideravelmente se os animais genotipados na densidade maior são parentes próximos dos indivíduos a serem imputados (HICKEY e KRANIS, 2013; WANG et al., 2013; WELLANN et al., 2013).

A imputação usando software que não utiliza diretamente as informações de pedigree, apresenta influência de outros fatores sobre a precisão de imputação (ZHANG e DRUET, 2010), como a relação entre indivíduos imputados e indivíduos genotipados em alta densidade (HICKEY et al., 2012).

A maioria dos fatores que interferem na acurácia de imputação interagem entre si, como o fato de a precisão de imputação aumentar quando a densidade do painel aumenta (BERRY et al., 2014). Observa-se também essa interação quando a informação de pedigree não é diretamente utilizada, em que nestes casos, o aumento do número de haplótipos no painel de referência tende a aumentar a precisão de imputação dos SNPs que possuem baixa frequência do alelo de menor frequência (MAF) (BADKE et al., 2013).

A precisão da imputação, sendo medida como a correlação entre o

genótipo verdadeiro e o imputado, é preferível quando comparada a medidas que contam o número de alelos ou genótipos corretos ou erroneamente imputados. Isso porque, considerando a imputação baseada apenas na frequência alélica apresentará acurácia igual a zero por essa medida, enquanto a acurácia medida pela taxa de erro terá um valor dependente da MAF, além disso, a correlação é mais sensível a erros em SNPs com MAF baixa, enquanto a taxa de erro de imputação mostra uma relação oposta (CALUS et al., 2014). Medir o desempenho de imputação corretamente será cada vez mais importante no futuro, quando muitas das tarefas de imputação envolverão dados de sequência, os quais têm uma proporção maior de loci com baixa MAF do que painéis de SNPs e, portanto, as métricas que dependem da MAF resultarão em valores imprecisos (CALUS et al, 2014).

4. METODOLOGIA

4.1. Animais e genotipagem

Inicialmente foram recebidos dados de 814 animais, sendo 780 novilhos machos, produzidos em três estações de monta, nascidos em 2007, 2008 e 2009, descendentes de 34 touros registrados da raça Nelore, constituindo famílias de meio-irmãos que foram gerados por inseminação artificial. Os animais Nelore foram criados nas fazendas da Embrapa Pecuária Sudeste, localizada na cidade de São Carlos, da Embrapa Gado de Corte, situada no município de Campo Grande e de propriedades particulares dos estados de Mato Grosso e Mato Grosso do Sul. Também foram recebidos dados de 285 animais da raça Canchim (CA) e 114 animais do grupo genético MA (MA), que são animais originários de um esquema de acasalamento entre a raça Canchim e Charolês. Na população estudada neste trabalho, foi verificado uma média do coeficiente de endogamia de 19,37%, tendo um coeficiente de endogamia alto justamente pela população analisada ser pequena.

Os 814 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). Foi então realizado o controle de qualidade dos genótipos, utilizando o programa PLINK, removendo SNPs localizados em regiões não autossômicas, os que possuíam posição desconhecidas, os que apresentaram MAF menor que 0,001 e os SNPs com desvios significativos ($p < 0,00001$) do equilíbrio de Hardy-Weinberg. Posteriormente ao controle de qualidade restaram 802 animais da raça Nelore e 396 animais Canchim e MA com 516.044 SNPs. As análises foram realizadas considerando apenas animais e SNPs que passaram no controle de qualidade, em que utilizou-se 400 animais da raça Nelore na população referência e os 396 animais da raça Canchim e MA na população de imputação.

4.2. Imputação

4.2.1. Cenários utilizados

Foram utilizados 16 diferentes cenários (Tabela 1), em que a população de animais referência é a população de animais Nelore, sendo escolhidos os 400 mais aparentados (alto grau de parentesco, com 0,005) e os 400 menos aparentados (baixo grau de parentesco, com -0,003), que foram mantidos todos os SNPs do painel de alta densidade (516.044 SNPs). A população de imputação foi formada pelos animais Canchim e MA, totalizando 396 animais, não havendo distinção quanto ao parentesco. Os animais da população de imputação que apresentavam 516.044 SNPs tiveram parte de seus SNPs “escondidos” com a finalidade de mimetizar que estes animais foram genotipados com painéis de baixa densidade 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.

Tabela 1. Cenários utilizados para o estudo de imputação para os painéis BovineHD BeadChip (700k).

População de Referência (Nelore)	Grau de Parentesco	População de Imputação (Canchim + MA)	Painel de menor densidade	Painel de maior densidade
400	Baixo	399	20k	700k
400	Baixo	399	20kP1	700k
400	Baixo	399	20kP2	700k
400	Baixo	399	20kP3	700k
400	Alto	399	20k	700k
400	Alto	399	20kP1	700k
400	Alto	399	20kP2	700k
400	Alto	399	20kP3	700k
400	Baixo	399	50k	700k
400	Baixo	399	50kP1	700k
400	Baixo	399	50kP2	700k
400	Baixo	399	50kP3	700k
400	Alto	399	50k	700k
400	Alto	399	50kP1	700k
400	Alto	399	50kP2	700k
400	Alto	399	50kP3	700k

20k- painel comercial contendo 19.263 mil SNPs; 50k- painel comercial contendo 26.342 mil SNPs; P1 - painel personalizado 1; P2 - painel personalizado 2; P3 - painel personalizado 3.

Considerou-se também 3 painéis personalizados, em que a partir do painel de 700k dos animais da raça Canchim e grupo MA, foram escolhidos diferentes SNPs formando painéis contendo a mesma quantidade de SNPs dos painéis comerciais de baixa densidade, ou seja, 19.263 (20k) e 26.342 SNPs (50k). Para a escolha dos SNPs para formar o primeiro painel

personalizado, 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 foi pela soma de DL entre cada SNP e os demais SNPs presentes nesta janela, o SNP que obteve maior soma de DL foi selecionado. O segundo painel personalizado considerou também o DL, porém também o SNP deveria atingir um critério mínimo de MAF (0,2 para 20k e 0,3 para 50K). Consideramos limites de MAF diferentes entre os painéis 20k e 50k de forma que fosse possível determinar ao menos um SNP dentro da janela que atendesse aos critérios.

Já no terceiro painel, 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. Esta terceira maneira de personalização está representada na Figura 1. Além dessas três formas de personalização foram também utilizados os SNPs presentes nos painéis comerciais de 20k e 50k, como mencionado.

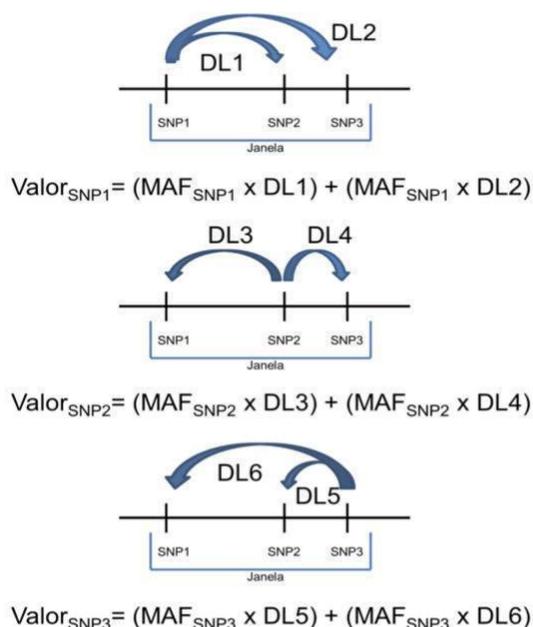


Figura 1. Esquema representando a terceira maneira de personalização. Dentro da janela considerada, foi calculado o desequilíbrio de ligação (DL1) entre o SNP1 e o SNP2 e este valor de DL1 foi multiplicado pela MAF do SNP1. Também foi calculado o desequilíbrio de ligação (DL2) entre o SNP1 e o SNP3, este valor de DL2 foi multiplicado pela MAF do SNP1. O resultado dos dois cálculos foi então somado. O mesmo cálculo foi repetido para cada SNP dentro da janela. O SNP que apresentou maior valor foi selecionado (BERNARDES, 2018).

Assim, a imputação do painel de baixa densidade para o painel de alta densidade foi realizada utilizando diferentes animais na população referência. Então foi verificada a acurácia de imputação desses diferentes cenários.

4.2.2. Algoritmos para Imputação

Um algoritmo que leva em consideração informações populacionais e de parentesco foi utilizado, como o implementado pelo programa FImpute (SARGOLZAEI; CHESNAIS; SCHENKEL, 2014). O algoritmo de imputação do FImpute utiliza como procedimentos a identificação de regiões que possam ser inferidas pelas informações dos pais ou da progênie na reconstrução dos haplótipos com alto grau de confiabilidade, (SARGOLZAEI et al., 2008) de forma que os haplótipos são reconstruídos de forma interativa. Esta interação é repetida até que a soma dos quadrados das probabilidades dos haplótipos seja suficientemente pequena. Os haplótipos das progênies são então combinados com os haplótipos dos pais ou descendentes e assim os genótipos não observados são inferidos, e a consistência da fase dos haplótipos é inferida por meio de sobreposição dos segmentos de cromossômicos.

4.2.3. Acurácia de imputação

A taxa de Concordância (TC) por animal foi estimada nos animais da população de imputação que foram genotipados com os painéis de alta densidade e tiveram seus SNPs “escondidos” para serem inferidos pelo programa. Esta Taxa de concordância foi calculada comparando os alelos dos marcadores imputados com os alelos observados e calculando a proporção dos genótipos imputados corretamente. Também foi utilizado a correlação (Cor), em que é a correlação entre a quantidade de alelos (alelo de efeito menor) dos marcadores imputados e a quantidade de alelos do genótipo verdadeiro (BROWNING; BROWNING, 2009). Sendo definido como acurácia de imputação os valores de TC e Cor.

5. RESULTADOS E DISCUSSÃO

Com base nos cenários apresentados foi possível comparar os diferentes valores de acurácia obtidos, sendo os cenários compostos por diferenças de parentesco entre os animais da população de referência, quantidade de SNPs que foram imputados e também, os diferentes painéis personalizados e comerciais utilizados.

As acurácias dos cenários em que a população referência apresentava alto parentesco tiveram maiores resultados quando comparados com os de baixo parentesco entre os animais Nelore, tal condição revela que os cenários tiveram maior influência nos resultados pelo grau de parentesco entre os animais Nelore do que quando avaliamos pelo número de SNPs, isso pode ter ocorrido em razão da MAF ter interferido positivamente na acurácia.

Esperava-se que os cenários com população referência com baixo parentesco (C1, C2, C3, C4, C9, C10, C11 e C12) apresentassem melhores resultados, por ter maior variabilidade de haplótipos, o que permite aumentar o número de variantes presentes na população de referência, representando melhor a diversidade genética e conseqüentemente auxiliando a correta imputação das variantes raras, conforme visto no estudo de Pausch et al. (2017). Estes autores, ao utilizarem uma população de referência multirracial, observaram aumento de 5% na acurácia de imputação dessas variantes menos frequentes. Porém, no presente estudo, a média de relacionamento entre os 400 animais com alto parentesco (0,005) não foi muito diferente da média dos 400 animais com baixo parentesco (-0,003), ou seja, a população considerando alto parentesco também apresentou variação de haplótipos. Além disso, possivelmente esses haplótipos podem ser maiores quando comparados aos obtidos com a população de baixo parentesco, pois a distribuição do DL pode se apresentar de maneira mais uniforme nos cenários de alto parentesco, mantendo-se mais constante por maior distância, aumentando o tamanho dos haplótipos e conseqüentemente, favorecendo os cenários de alto parentesco para que tivessem melhores valores de acurácia.

Tabela 2. Número de SNPs que foram imputados (N° de SNPs), média da taxa de concordância (TC) e média da correlação (Cor) para todos os animais e, considerando os dois grupos raciais (Canchim e MA), para os diferentes cenários.

Cenários	Grau de Parentesco	Painel de Menor Densidade	N° de SNPs	TC	Cor	Canchim		MA	
						TC	Cor	TC	Cor
C1-400baixo_20kcom_700k	baixo	20k	496781	56,62	0,4838	56,55	0,485	56,78	0,481
C2-400baixo_20kc1_700k	baixo	20k	496781	63,52	0,5702	63,45	0,57	63,7	0,57
C3-400baixo_20kc2_700k	baixo	20k	496781	63,72	0,5747	63,6	0,57	63,9	0,576
C4-400baixo_20kc3_700k	baixo	20k	496781	63,3	0,573	63,19	0,572	63,56	0,575
C5-400alto_20kcom_700k	alto	20k	496781	58,73	0,5198	58,67	0,52	58,89	0,518
C6-400alto_20kc1_700k	alto	20k	496781	65,57	0,6053	65,49	0,605	65,76	0,606
C7-400alto_20kc2_700k	alto	20k	496781	65,82	0,6108	65,75	0,61	66	0,612
C8-400alto_20kc3_700k	alto	20k	496781	65,46	0,6095	65,36	0,608	65,71	0,612

Tabela 3. Número de SNPs que foram imputados (N° de SNPs), média da taxa de concordância (TC) e média da correlação (Cor) para todos os animais e, considerando os dois grupos raciais (Canchim e MA), para os diferentes cenários.

Cenários	Grau de Parentesco	Painel de Menor Densidade	N° de SNPs	TC	Cor	Canchim		MA	
						TC	Cor	TC	Cor
C9-400baixo_50kcom_700k	baixo	50k	489702	58,68	0,5027	58,6	0,503	58,87	0,501
C10-400baixo_50kc1_700k	baixo	50k	489702	66,36	0,6074	66,29	0,607	66,53	0,608
C11-400baixo_50kc2_700k	baixo	50k	489702	65,97	0,6069	65,88	0,606	66,18	0,608
C12-400baixo_50kc3_700k	baixo	50k	489702	66,36	0,6127	66,27	0,612	66,6	0,614
C13-400alto_50kcom_700k	alto	50k	489702	60,8	0,5389	60,74	0,539	60,97	0,538
C14-400alto_50kc1_700k	alto	50k	489702	68,5	0,6426	68,44	0,642	68,64	0,643
C15-400alto_50kc2_700k	alto	50k	489702	68,3	0,6451	68,23	0,644	68,46	0,646
C16-400alto_50kc3_700k	alto	50k	489702	68,66	0,6501	68,57	0,649	68,86	0,652

Pei et al. (2008) avaliaram o efeito do DL na acurácia de imputação utilizando vários softwares (MACH, IMPUTE, fastPHASE, PLINK e Beagle) e em todos os softwares de imputação a acurácia de imputação aumentou quando o nível de DL foi maior. Dado o tamanho da população de referência de 90 indivíduos, a acurácia foi 62,8%, 75,9% e 95,1% para os níveis baixo, médio e forte de DL para o MACH. Porém, mais estudos precisam identificar o DL nas populações de alto e baixo parentesco consideradas no presente estudo.

Em geral os painéis personalizados obtiveram valores superiores de acurácia (63,52, 63,72, 63,3, 65,57, 65,82, 65,46, 66,36, 65,97, 66,36, 68,5, 68,3, 68,66) quando comparados com os comerciais (56,62, 58,73, 58,68, 60,8). Semelhantemente, Romero (2017) observou que painéis personalizados apresentaram vantagem em relação aos painéis comerciais em todos os casos estudados, tendo para o painel comercial de 3k uma acurácia de 0,91 e seu personalizado com acurácia de 0,95, e para o painel comercial de 7k uma acurácia de 0,95 e seu personalizado um valor igual a 0,98. Estes resultados indicam a potencialidade do uso desses painéis para estudos genômicos, reduzindo custos e proporcionando dados com qualidade, aptos para aplicação em estudos posteriores. No presente estudo, os painéis personalizados pela segunda e terceira maneira tiveram maiores acurácias (63,72; 63,3; 65,82; 65,46; 65,97; 66,36; 68,3; 68,66). Isso pode ser explicado pela participação do valor de MAF que foi considerado nesses painéis. De fato, resultados obtidos por Bernardes et al. (2019), quando comparados os painéis comerciais com os painéis personalizados, a terceira forma de personalização (Cust3) foi a que aumentou a acurácia de imputação para quase todas as situações estudadas, sendo que esta forma de personalização considera dois fatores que afetam a acurácia de imputação, como o DL e MAF.

A MAF demonstrou ter influência nas acurácias ao ser considerada na construção dos painéis personalizados 2 e 3, pois visto que maioria dos indivíduos são homozigotos para o alelo mais comum da população, torna a imputação desse alelo mais fácil do que a imputação do alelo raro (MA et al., 2013). No entanto, a imputação acurada de alelos raros é fundamental, pois estes alelos podem explicar variações genéticas que não são explicadas

pelos alelos comuns (CIRULLI ; GOLDSTEIN, 2010).

Ao comparar os painéis que possuem densidade de 20k (56,62; 63,52; 63,72; 63,3; 58,73; 65,57; 65,82; 65,46) com os de 50k (58,68; 66,36; 65,97; 66,36; 60,8; 68,5; 68,3; 68,66), nota-se que os painéis de 50k apresentaram maiores valores de acurácia, sendo justificado pelo fato de acurácia aumentar conforme a densidade do painel aumenta (BERRY et al., 2014). Assim, para aqueles cenários em que o painel de menor densidade for o de 50 mil SNPs esperava-se que a acurácia fosse maior, como observado no presente estudo. O fato de haver menor quantidade de SNPs a serem imputados ao utilizar a densidade de 50k quando comparado com 20k, favorece a construção de haplótipos pelo programa, o que pode reduzir os erros de imputação.

Boichard et al. (2012) perceberam que o ganho na precisão da imputação é atribuído principalmente ao aumento da densidade. Da mesma maneira, Khatkar et al. (2012), estudando imputação em bovinos da raça Holstein-Friesian australianos e com painéis de 3k e 35k, notaram que a precisão da imputação também aumenta com o número de SNPs no conjunto de teste para todos os cenários e métodos examinados, sendo que a taxa média de erro alélico diminuiu de 2,80%, utilizando o painel de 3K, para 0,76%, utilizando o painel 35K. Assim com também, Vasconcelos (2014) notou que a eficiência de imputação aumenta com o aumento do número de SNPs no painel de menor densidade, sendo que estes autores utilizaram painéis de menor densidade como sendo de 7k, 20k, 50k e 75k para painel de maior densidade de 780k e as porcentagem de genótipos imputados corretamente foi de 92,29%, 98,23%, 98,33% e 99,40, respectivamente.

A densidade de marcadores é importante fator que afeta a acurácia de imputação, pois painéis com muitos SNP apresentam maior desequilíbrio de ligação entre os marcadores, reduzindo assim os erros de imputação (HICKEY et al., 2012). Segundo Wang et al. (2012), quanto maior a taxa de genótipos não observados menor é a acurácia de imputação. Para populações que apresentam desequilíbrio de ligação relativamente grande e tamanho efetivo pequeno, permite-se prever genótipos provenientes de painéis com poucos marcadores para painéis de alta densidade com baixas taxas de erros de imputação (CHUD et al., 2015).

Todos os cenários apresentaram o mesmo perfil para a média de taxa de concordância de acordo com o cromossomo (Figura 2). A correlação também seguiu o mesmo perfil da taxa de concordância, porém com valores um pouco inferiores de acurácia. Nota-se que os cromossomos 8 e 14 apresentaram maiores valores de acurácia. Em seu estudo Chud et al. (2014) também verificaram resultados diferentes nos cromossomos, em que para o painel de 80K, os cromossomos 27 e 28 apresentaram maior erro de imputação alélico, 3,9% e 3,8%, respectivamente, enquanto que no painel de 90K, o cromossomo 13 apresentou maior erro de imputação, 4,27%. Segundo Sun et al. (2012), essa diferença de imputação observada por cromossomo é devido ao tamanho do cromossomo, pois imputar alelos corretamente é mais difícil na região inicial e final do cromossomo, com isso os cromossomos mais curtos são mais afetados do que cromossomos longos, promovendo menor taxa de alelos imputados incorretamente. Além disso, os cromossomos mais longos possuem maior número de marcadores do que cromossomos mais curtos.

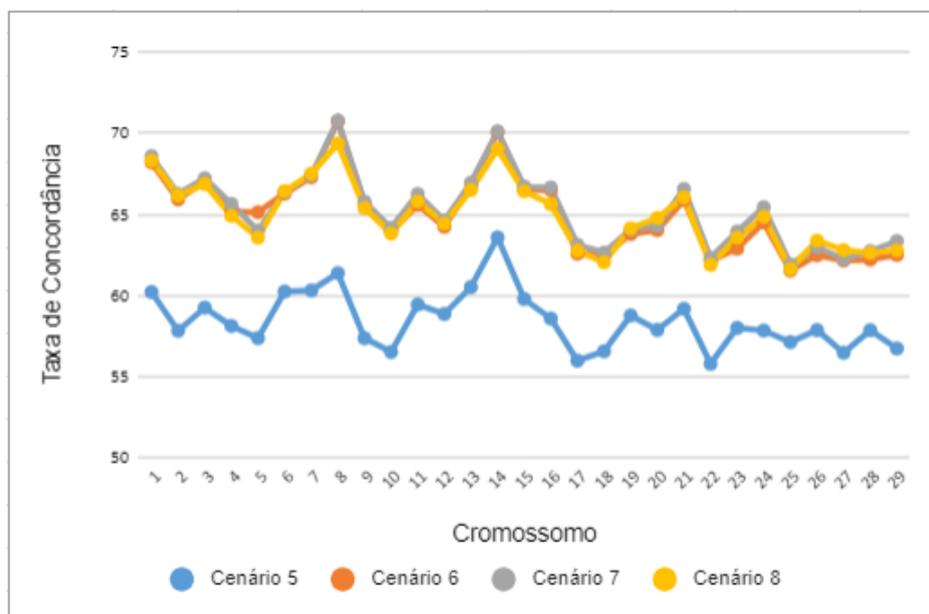


Figura 2. Taxa de concordância por cromossomo considerando os cenários Alto Parentesco com 20k.

Os resultados de acurácia dos animais Canchim e MA não apresentaram diferenças, indicando que a imputação foi similar para os

grupos raciais estudados. De acordo com Ventura et al. (2014), as acurácias de imputação em populações multirraciais de bovinos podem ser influenciadas pela composição racial existente entre as populações de referência e de imputação. Marciano et al. (2018), estudando um cenário que considerava a população referência sendo animais Canchim e Nelore, e população de imputação os animais MA, e outro cenário que considerava a população referência sendo animais Nelore e MA e população imputação os animais Canchim reportaram evidente aumento na correlação (0,91 e 0,86) e TC (89,69% e 84,44%), respectivamente, quando comparado com a população de referência sendo constituída apenas por animais Nelore (0,63 e 65,87). Chud et al. (2014) notaram que quando a população referência foi composta apenas por animais Canchim ou MA, os resultados obtidos foram insatisfatórios para aplicação da imputação, assim sugeriram que os dois grupos (canchim e MA) sejam considerados juntos, pois isto possibilita o aumento do parentesco médio e da população referência, além de auxiliar na construção dos haplótipos necessários para a imputação.

Duarte et al. (2018) compararam as acurácias de diferentes cenários, sendo o primeiro cenário formado pela população referência Charolês e população imputação Canchim e MA, o segundo composto por população referência Charolês e população imputação Canchim, o terceiro cenário sendo com população referência Charolês e população imputação MA, quarto cenário com população referência Charolês e Canchim e população imputação MA e quinto cenário composto por população referência Charolês e MA e população imputação Canchim. Os autores observaram aumento no quarto e quinto cenário para ambas as medidas de acurácia testadas. Este resultado pode estar associado ao maior parentesco existente entre as populações de animais Canchim e do grupo genético MA. Tanto para a formação da raça Canchim quanto para o grupo MA, são consideradas maiores contribuições da raça Charolês. Isso também pode ser uma explicação para as acurácias desse estudo não terem sido tão altas, pois foi utilizada uma população referência com animais Nelore, estes sendo mais distantes geneticamente dos animais Canchim e MA, quando comparado com os animais da raça Charolês.

6. CONCLUSÕES

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 diferença entre os painéis personalizados 1 e os personalizados 2 e 3, indicando que a maneira como o painel é construído interfere nas acurácias de imputação. Utilizar raças fundadoras na população referência pode ser uma alternativa para raças em que a população não é grande o suficiente.

7. REFERÊNCIAS BIBLIOGRÁFICAS

ABCCAN. Associação Brasileira de Criadores de Canchim. 2020. Disponível em: <http://www.abccan.com.br/canchim/index.php/a-raca.html>. Acesso em: 25/06/2022.

ALENCAR, M. M., Bovino - Raça Canchim: Origem e Desenvolvimento. Brasília: Embrapa-DMU, 1988. Disponível em: https://www.infoteca.cnptia.embrapa.br/infoteca/bitstream/doc/38887/1/PROCID_0c4MMA1985636.210824.pdf. Acesso em: 04/05/2022.

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. Disponível em: <https://ainfo.cnptia.embrapa.br/digital/bitstream/item/188409/1/Touro-Canchim-no-cruzamento-industrial.pdf>. Acesso em: 06/05/2022.

ALENCAR M.M.; SILVA, A.H.G.; BARBOSA, P.F. Efeitos da consanguinidade sobre o peso à desmama de bezerros da raça Canchim. Revista da Sociedade Brasileira de Zootecnia, 10: 156 - 172, 1981

ARPANAHI. R. A.; MOROTA G.; PEÑAGARICANO F. Predicting bull fertility using genomic data and biological information. *Journal of Dairy Science*. Vol. 100. 2017.

BADKE, Y. M.; BATES, R. O.; ERNST, C. W.; SCHWAB, C.; FIX, J.; VAN TASSELL, C. P.; STEIBEL, J. P., Methods of tagSNP selection and other variables affecting imputation accuracy in swine. *BMC Genetics* v. 14 n. 8, 2013. Disponível em: <https://bmcgenomdata.biomedcentral.com/articles/10.1186/1471-2156-14-8>. Acesso em: 17/05/2022.

BERRY, D. P. & KEARNEY, J. F., Imputation of genotypes from low- to high-density genotyping platforms and implications for genomic selection. *Animal*. v. 5 p. 1162–1169, 2011. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/22440168/>. Acesso em: 15/04/2022.

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* n. 131 p. 165–172, 2014. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/24906026/>. Acesso em: 27/05/2022.

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*, v. 9, p. 17920, 2019.

BOICHARD D.; CHUNG H.; DASSONNEVILLE R.; DAVID X.; EGGEN A.; FRITZ S.; GIETZEN K.J.; HAYES B.J.; LAWLEY C.T.; SONSTEGARD T.S.; VAN TASSELL C.P.; Design of a bovine low-density SNP array optimized for imputation. *PLoS one*. 2012.

BOISON, S. A.; SANTOS, D. J. A.; UTSONOMIYA, A. H. T.; GARCIA, J. F.; SOLKNER, J.; SILVA, M. V. G. B., 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*, v. 98, n. 7, p. 4969–4989, 2015. Disponível em: [https://www.journalofdairyscience.org/article/S0022-0302\(15\)00302-1/fulltext#](https://www.journalofdairyscience.org/article/S0022-0302(15)00302-1/fulltext#). Acesso em: 30/04/2022.

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*, Houston, v. 84, p. 210-223, 2009.

BUZANSKAS M. E.; VENTURA R. V.; CHUD T. C. S.; BERNARDES P. A.; SANTOS D. J. A.; REGITANO L. C. A.; ALENCAR M. M.; MADADU M. A.; ZANELA R.; SILVA M. V. G. B.; LI C.; SCHENKEL F. S.; MANURI D. P. Study on the introgression of beef breeds in Canchim cattle using single nucleotide polymorphism markers. *Plos One*. 2017. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0171660>. Acesso em: 06/12/2022.

CALUS, M. P. L.; BOUWMAN, A. C.; HICKEY, J. M.; VEERKAMP, R. F.; MULDER, H. A., Evaluation of measures of correctness of genotype imputation in the context of genomic prediction: a review of livestock applications. *Animal*. v. 8 n. 11 p. 1743–1753, 2014. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/25045914/>. Acesso em: 14/05/2022.

CARVAJAL A. B. Estrutura Populacional e Estimação de Parâmetros Genéticos para Características Associadas ao Desempenho Reprodutivo em Bovinos da Raça Caracu. Dissertação de mestrado apresentada à Faculdade de Ciências Agrárias e Veterinárias (UNESP). 2017. Disponível em:

https://repositorio.unesp.br/bitstream/handle/11449/151868/barreracarvajal_a_me_jabo.pdf?sequence=4. Acesso em: 05/12/2022.

CARVALHEIRO, R.; BOISIN, S. A.; NEVES, H. H. R.; SARGOLZAE, M.; SCHENKEL, F. S.; UTSUNOMIYA, Y. T.; O'BRIEN, A. M. P.; SOLKNER, J.; MCEWAN, J. C.; VAN TASSELL, C. P.; SONSTEGARD, T. S.; GARCIA, F. J., Accuracy of genotype imputation in Nelore cattle. *Genetics Selection Evolution*. p. 46:69, 2014. Disponível em: <https://repositorio.unesp.br/handle/11449/117329>. Acesso em: 01/05/2022.

CHUD, T. C. S.; VENTURA, R. V.; SCHENKEL, F. S.; CARVALHEIRO, R.; BUZANSKAS, M. E.; ROSA, J. O.; MUDADU, M. A.; SILVA, M. V. G. B.; MOKRY, F. B.; MARCONDES, C. R.; REGITANO, L. C. A.; MUNARI, D. P., Strategies for genotype imputation in composite beef cattle. *BMC Genetics* v. 16 n. 99, 2015. Disponível em: <https://bmcgenomdata.biomedcentral.com/articles/10.1186/s12863-015-0251-7>. Acesso em: 01/05/2022.

CIRULLI, E. T.; GOLDSTEIN, D. B. Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nature Reviews Genetics*, v. 11, p.415- 425, 2010.

DAETWYLER, H D.; CALUS, M. P. L.; WONG, P. R.; DE LOS CAMPOS, G.; HICKEY, J., M.Genomic prediction in animals and plants: simulation of data, validation, reporting, and benchmarking. *Genetics* v. 193 p. 347–365, 2013. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/23222650/>. Acesso em: 06/05/22.

DUARTE, J. L. G.; BATES, R. O.; ERNST, C. W.; RANEY, N. E.; CANTET, R. J.; STEIBEL, J. P., Genotype imputation accuracy in a F2 pig population using high density and low density SNP panels. *BMC Genetics*. n. 14 p. 1–13, 2013. Disponível em: <https://bmcgenomdata.biomedcentral.com/articles/10.1186/1471-2156-14-38>.

Acesso em: 28/05/2022.

DUARTE I. N. H; MAIA R. O .G.; SANTOS M. S.; MARCIANO L. E. A; BERNARDES P. A.; CHUD T. C. S.; REGITANO L. C. A; BUZANSKAS M. E., Utilização de Genótipos da Raça Charolês para Auxílio na Imputação de Animais da Raça Canchim. 28° Congresso Brasileiro de Zootecnia. 2018. Disponível em: <http://www.adaltech.com.br/anais/zootecnia2018/resumos/trab-0113.pdf>.

Acesso em: 04/11/2022.

HERMISDORFF, I. C.; COSTA, R. B.; ALBUQUERQUE, L. G.; PAUSCH, H.; KADRI, N. K., Investigating the accuracy of imputing autosomal variants in Nellore cattle using the ARS-UCD1.2 assembly of the bovine genome. BMC Genomics v. 21 n. 772, 2020. Disponível em: <https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-020-07184-8>.

Acesso em: 27/04/2022.

HICKEY, J. M.; CROSSA, J.; BABU, R.; DE LOS CAMPOS, G., Factors affecting the accuracy of genotype imputation in populations from several maize breeding programs. Crop Sci. v. 52 p. 654–663, 2012. Disponível em: <https://acess.onlinelibrary.wiley.com/doi/abs/10.2135/cropsci2011.07.0358>.

Acesso em: 05/05/2022.

HICKEY, J. M. & KRANIS, A., Extending long-range phasing and haplotype library imputation methods to impute genotypes on sex chromosomes. Genetics Selection Evolution. n. 45 p. 10, 2013. Disponível em: <https://gsejournal.biomedcentral.com/articles/10.1186/1297-9686-45-10>.

Acesso em: 28/05/2022.

HOZÉ, C.; FOUILLOUX, M. N.; VENOT, E.; GUILLAUME, F.; DASSONNEVILLE, R., Precisão de imputação de marcadores de alta densidade em dezesseis raças de bovinos franceses. Genet. Sel. v. 45 p.

1–11, 2013. Disponível em:
<https://gsejournal.biomedcentral.com/articles/10.1186/1297-9686-45-33>.
Acesso em: 17/05/2022.

KHATKAR M. S.; MOSER G.; HAYES B. J.; RAADSMA H. W., Strategies and utility of imputed SNP genotypes for genomic analysis in dairy cattle. *BMC Genomics*. 2012. Disponível em:
<https://link.springer.com/article/10.1186/1471-2164-13-538#citeas>. Acesso em: 08/11/2022.

MA, P.; BRØDUM, R. F.; ZHANG, Q.; LUND, M. S.; SU, G. Comparison of different methods for imputing genome-wide marker genotypes in Swedish and Finnish Red Cattle, *Journal of Dairy Science*, v. 96, p.4666-4677, 2013.

MACHADO, C. H. C., Análise Histórica das Raças Zebuínas no Brasil e Estudo no Consumo Alimentar Residual (CAR) Como Critério de Seleção. *Repositorio.uniube*. 2015. Disponível em:
<https://repositorio.uniube.br/bitstream/123456789/616/1/Disserta%C3%A7%C3%A3o.pdf>. Acesso em: 10/06/2022.

MCKAY, S.D., SCHNABEL, R.D., MURDOCH, B.M., Whole genome linkage disequilibrium maps in cattle. *BMC Genet.*, v.8, p.74, 2007. Disponível em:
<https://bmcgenomdata.biomedcentral.com/articles/10.1186/1471-2156-8-74>.
Acesso em: 10/06/2022.

MAIO, A., Mapa mostra expansão da raça Canchim. *Embrapa Pecuária Sudeste*, 2018. Disponível em:
<https://www.embrapa.br/busca-de-noticias/-/noticia/36342551/publicacao-mapeia-357-trabalhos-sobre-canchim-nos-ultimos-dez-anos>. Acesso em: 01/05/2022.

MARCIANO L. E. A; MAIA R. O .G.; SANTOS M. S.; DUARTE I. N. H.; BERNARDES P. A.; CHUD T. C. S.; REGITANO L. C. A; BUZANSKAS M. E.,

Estratégias de Imputação em Gado Canchim Utilizando População de Referência da Raça Nelore. 28º Congresso Brasileiro de Zootecnia. 2018. Disponível em: <https://www.alice.cnptia.embrapa.br/alice/bitstream/doc/1110873/1/2018trab0112.pdf>. Acesso em: 04/11/2022.

MOGHADDAR, N.; KLINT, P. G.; HANS, D. D.; HAYES, B. J.; VAN DER WERF, J. H. J., Accuracy of genotype imputation based on random and selected reference sets in purebred and crossbred sheep populations and its effect on accuracy of genomic prediction. *Genet. Sel.* vl. 47 p. 1–12, 2015. Disponível em: <https://gsejournal.biomedcentral.com/articles/10.1186/s12711-015-0175-8>. Acesso em: 04/05/2022.

MOKRY F. B.; BUZANSKAS M. E.; ALVARENGA M. M. et al. Linkage disequilibrium and haplotype block structure in a composite beef cattle breed. *BMC Genomics.* (2014). Disponível em: <https://link.springer.com/article/10.1186/1471-2164-15-S7-S6#citeas>. Acesso em: 05/12/2022.

OLIVEIRA JÚNIOR, G. A.; CHUD, T. C. S.; VENTURA, R. V.; GARRICK, D. J.; COLE, J. B.; MUNARI, D. P.; FERRAZ, J. B.S.; MULLART, E.; DENISE, S.; SMITH, S., Genotype imputation in a tropical crossbred dairy cattle population. *Journal of Dairy Science*, v. 100, n. 12, p. 9623–9634, 2017. Disponível em: <https://repositorio.unesp.br/handle/11449/175291>. Acesso em: 04/05/2022.

PAUSCH, H.; AIGNER, B.; EMMERLING, R.; EDEL, C.; GOTZ, K. U.; FRIES, R., Imputation of high-density genotypes in the Fleckvieh cattle population. *Genet. Sel.* v. 45 p. 1–10, 2013. Disponível em: <https://gsejournal.biomedcentral.com/articles/10.1186/1297-9686-45-3>. Acesso em: 30/05/2022.

PAUSCH, H.; MACLEOD, I. M.; FRIES, R.; EMMERLING, R.; BOWMAN, P. J.; DAETWYLER, H. D.; GODDARD, M. E. Evaluation of the Accuracy of Imputed Sequence Variant Genotypes and Their Utility for Causal Variant Detection in Cattle. *Genetics Selection Evolution*. v. 49 p. 24, 2017.

PEI, Y.; LI, J.; ZHANG, L.; PAPASIAN, C.J.; DENG, H. Analyses and comparison of accuracy of different genotype imputation methods. *PLoS ONE*, San Francisco, v. 3, n. 10, p. 1-8, 2008.

PEIXOTO, M. G. C. D.; SANTOS, D. J. A.; CARVALHO, M. R. S.; MACHADO, M. A.; SILVA, M. V. G. B.; TONHA, H., Desequilíbrio de ligação com um painel de alta densidade em genomas de bovinos da raça Guzerá. X Simpósio Brasileiro de Melhoramento Animal Uberaba, MG. 2013. Disponível em: <http://sbmaonline.org.br/anais/x/trabalhos/pdf/6MWP.pdf>. Acesso em: 10/06/2022.

PIMENTEL, E.; WENSCH-DORENDORF, M.; KONIG, S.; SWALVE, H., Enlarging a training set for genomic selection by imputation of un-genotyped animals in populations of varying genetic architecture. *Genetics Selection Evolution*. v. 45, 2013. Disponível em: <https://gsejournal.biomedcentral.com/articles/10.1186/1297-9686-45-12>. Acesso em: 03/05/2022.

REGITANO L. C. A. Polimorfismo molecular em gerações de bovinos da raça Canchim. Universidade de São Paulo. 1996. Disponível em: <https://www.teses.usp.br/teses/disponiveis/11/11137/tde-20200111-130155/publico/RegitanoLucianaCorreiaAlmeida.pdf>. Acesso em: 05/12/2022.

ROMERO, A. R. S.; Estudos Genômicos Aplicados ao Melhoramento Genético de Bovinos. UFGD/Programa de Pós - Graduação em Biologia Geral. 2017. Disponível em: <https://www.alice.cnptia.embrapa.br/alice/bitstream/doc/1082424/1/Estudosgenomicosaplicadosaomelhoramentogeneticodebovinos.pdf>. Acesso em:

04/11/2022.

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. 478, 2014.

SARGOLZAEI, M.; SCHENKEL, F.S.; JANSEN, G.B.; SCHAEFFER, L.R. Extent of linkage disequilibrium in Holstein cattle in North America. *Journal of Dairy Science*, Champaign, v. 91, n. 5, p. 2106–2117, 2008.

SCHROOTEN C.; DASSONNEVILLE R.; DUCROCQ V.; BRONDUM R. F.; Lund M. S.; Chen J.; Liu Z.; González-Recio O.; Pena J.; Druet T. Error rate for imputation from the Illumina BovineSNP50 chip to the Illumina BovineHD chip. *Genet. Seleçione. Evolução*, 46, p. 10. 2014

TSAI, H. Y.; MATIKA, O.; EDWARDS, S. M.; ANTOLÍN-SÁNCHEZ, R.; HAMILTON, A.; GUY, D. R.; TINCH, A. E.; GHARBI, K.; STEAR, M. J.; TAGGART, J. B.; BRON, J. E.; HICKEY, J. M.; HOUSTON, R. D., Genotype imputation to improve the cost-efficiency of genomic selection in farmed Atlantic salmon. *G3: Genes, genomes, genetics*, v. 7, n. 4, p. 1377-1383, 2017. Disponível em: <https://academic.oup.com/g3journal/article/7/4/1377/6031815?login=true>. Acesso em: 25/05/2022.

UEMOTO, Y.; SASAKI, S.; SUGIMOTO, Y.; WATANABE, T., Precisão da imputação genotípica de alta densidade em bovinos negros japoneses. *Anim. Genet.* v. 46 p. 388–394, 2015. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1111/age.12314>. Acesso em: 30/05/2022.

VENTURA, R. V., D. Lu, F. S. SCHENKELI, Z. WANG, C. Li, and S. P. MILLER. Impact of reference population on accuracy of imputation from 6K to 50K single nucleotide polymorphism chips in purebred and crossbreed beef cattle. *J Anim Sci* 92:1433–1444. 2014.

VASCONCELOS, F. O., Predição genômica utilizando painéis de marcadores moleculares com diferentes densidades, em bovinos da raça Nelore. Pós-graduação Genética e Melhoramento Animal - FCAV. 2014.

WANG, C.; HABIER, D.; PEIRIS, B. L.; WOLC, A.; KRANIS, A.; WATSON, K. A.; AVENDANO, S.; GARRICK, D. J.; FERNANDO, R. L.; LAMONT, S. J.; DEKKERS, J. C. M., Accuracy of genomic prediction using an evenly spaced, low- density single nucleotide polymorphism panel in broiler chickens. Poultry Science. n. 92 p. 1712–1723, 2013. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/23776257/>. Acesso em: 03/05/2022.

WANG, H.; WOODWARD, B.; BAUCK, S.; REKAYA, R., Imputation of missing SNP genotypes using low density panels. Livestock Science. n. 146 p. 80–83, 2012. Disponível em: <https://agris.fao.org/agris-search/search.do?recordID=US201400056373>. Acesso em: 06/05/2022.

WANG Y.; LIN G.; LI C.; STOTHARD P. Genotype Imputation Methods and Their Effects on Genomic Predictions in Cattle. Springer Science Reviews. Vol. 4, pag. 79–98. 2016.

WANG, Y.; ZHIPENG, C.; STOTHARD, P.; MOORE, S.; GOEBEL, R.; WANG, L.; LIN, G. Fast accurate missing SNP genotype local imputation. BMC Research Notes, v.5, p.404, 2012.

WELLMANN, R.; PREU, S.; THOLEN, E.; HEINKEL, J.; WIMMERS, K.; BENNEWITZ, J., Genomic selection using low density marker panels with application to a sire line in pigs. Genetics Selection Evolution n. 45 p. 28, 2013. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/23895218/>. Acesso em: 27/04/2022.

ZHANG, Z. & DRUET, T., Marker imputation with low-density marker panels in Dutch Holstein cattle. Journal of Dairy Science n. 93 p. 5487–5494, 2010.

Disponível em: [https://www.journalofdairyscience.org/article/S0022-0302\(10\)00590-4/fulltext](https://www.journalofdairyscience.org/article/S0022-0302(10)00590-4/fulltext). Acesso em: 30/05/2022.