

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

YAGOR BARBOSA DOS SANTOS RAMOS

**IMPUTAÇÃO COM DOIS PAINÉIS DE ALTA
DENSIDADE UTILIZANDO AS MONTAGENS DE GENOMA
BOVINO UMD3.1 E ARS-UCD1.2**

FLORIANÓPOLIS - SC

2022

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

YAGOR BARBOSA DOS SANTOS RAMOS

**IMPUTAÇÃO COM DOIS PAINÉIS DE ALTA
DENSIDADE UTILIZANDO AS MONTAGENS DE GENOMA
BOVINO UMD3.1 E ARS-UCD1.2**

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. Priscila Arrigucci Bernardes

FLORIANÓPOLIS - SC

2022

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

RAMOS, YAGOR BARBOSA DOS
IMPUTAÇÃO COM DOIS PAINÉIS DE ALTA DENSIDADE UTILIZANDO
AS MONTAGENS DE GENOMA BOVINO UMD3.1 E ARS-UCD1.2 / YAGOR
BARBOSA DOS RAMOS ; orientadora, Priscila Arrigucci
Bernardes, 2022.
35 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. Imputação. 3. SNP. 4. Genoma Bovino. I.
Arrigucci Bernardes, Priscila . II. Universidade Federal
de Santa Catarina. Graduação em Zootecnia. III. Título.

Yagor Barbosa dos Santos Ramos

IMPUTAÇÃO COM DOIS PAINÉIS DE ALTA DENSIDADE UTILIZANDO AS MONTAGENS DE GENOMA BOVINO UMD3.1 E ARS-UCD1.2

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

Florianópolis, 11 de julho 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 ao meu pai Vilmar Ramos (*in memoriam*), com todo o amor e gratidão pelo apoio nessa jornada.

AGRADECIMENTOS

Gostaria de agradecer a Deus, por ser um guia e estar sempre ao meu lado durante toda essa jornada, me guiando pela fé e não pela visão.

Aos meus pais Mari e Lucas por terem me proporcionado a educação que tenho hoje, sem medir esforços para me ajudar a alcançar todos os degraus que a vida me apresentou, terem trabalhado para me dar o apoio e estrutura para seguir em frente, por me ensinarem que o estudo é uma das melhores ferramentas para uma vida mais digna, e junto do trabalho honesto, o caminho para realizar os sonhos e objetivos de vida.

A minha professora orientadora Priscila, que me proporcionou a oportunidade de conduzir o presente estudo na área de conhecimento que mais tenho afinidade, ao mesmo tempo que compartilhou uma parcela do seu conhecimento tanto em sua área de estudo, como seu conhecimento de vida.

A todos os meus amigos e colegas, que compartilharam momentos de alegria, tristeza e ansiedade durante a confecção desse trabalho, vocês me ajudaram a superar momentos de dificuldade e compartilharam os pequenos e grandes avanços que hoje, constituem esse trabalho.

Muito obrigado!

Pelos caminhos difíceis, aos astros.

(Provérbio latim)

RESUMO

Para obter a melhor resposta possível para a seleção genômica, é fundamental uma ampla quantidade de animais genotipados com muitos SNPs. Todavia, painéis com alta densidade de SNPs ainda apresentam custo elevado. Uma alternativa para reduzir os custos, é a utilização da técnica da imputação, que permite obter painéis de alta densidade a partir de painéis de baixa densidade. A imputação ocorre a partir da genotipagem de animais jovens em painéis de menor custo, sendo posteriormente inferidos SNPs faltantes usando uma população referência genotipada com painéis de alta densidade. Para a utilização de animais genotipados a partir da técnica de imputação em programas de seleção, é necessário que a acurácia da imputação seja elevada e esta depende de uma série de fatores. Assim, o objetivo desse estudo foi avaliar a acurácia de imputação em 32 cenários distintos entre si quanto a densidade, parentesco e quantidade de animais na população referência, utilizando a montagem nova do genoma (ARS-UCD1.2) e a montagem antiga (UMD3.1). Os dados utilizados foram de 801 animais da raça Nelore, os quais foram genotipados com o BovineHD BeadChip (Illumina), sendo que destes, 94 animais também foram genotipados com o Axion Genome-Wide BOS 1 Array Plate (Affymetrix). Foram utilizados 32 diferentes cenários, em que a população de animais referência foi composta por animais que foram mantidos todos os SNPs do painel de alta densidade. Os demais animais da população tiveram seus SNPs “removidos” com a finalidade de mimetizar que estes animais foram genotipados com painéis comerciais de baixa densidade contendo 20 mil SNPs (20k) ou 50 mil SNPs (50k). Assim, foi realizada a imputação do painel de baixa densidade para o painel de alta densidade utilizando diferentes animais na população referência. A acurácia de imputação destes diferentes cenários foi verificada utilizando a taxa de concordância (TC), que compara os alelos dos marcadores imputados com os alelos observados, calculando a proporção dos genótipos imputados corretamente. De maneira geral, aumentar a densidade dos painéis a serem imputados representou aumento na acurácia de todos os cenários. Para o painel Axion Genome-Wide BOS 1 Array Plate (Affymetrix), aumentar a população referência elevou a acurácia de imputação para os cenários em que havia alto parentesco entre a população referência e população imputação. Nos cenários de maior população referência, aumentar o parentesco dos animais elevou a acurácia de imputação. No painel BovineHD BeadChip (Illumina) foi possível observar o aumento na acurácia de imputação com aumento da população referência na montagem ARS-UCD1.2. Enquanto que aumentar o parentesco dos animais elevou a acurácia de imputação apenas na montagem UMD3.1. Em geral, as diferenças na acurácia de imputação observadas entre as montagens ARS-UCD1.2 e UMD3.1 podem ter ocorrido devido ao efeito das mudanças de posição dos SNPs nos diferentes painéis, em que a acurácia média de alguns cromossomos pode ter sido reduzida na nova montagem. Assim, sugere-se estudos posteriores para maior compreensão das regiões cromossômicas e posição dos SNPs.

Palavras-chave: Acurácia de imputação. Montagem do genoma. SNP.

LISTA DE FIGURAS

Figura 1. Taxa de concordância por cromossomo na montagem UMD3.1 e ARS-UCD1.2 no painel BovineHd Beadchip (Illumina).....	15
---	----

LISTA DE TABELAS

Tabela 1. Cenários utilizados para o estudo de imputação para os painéis BovineHD BeadChip (700k) e Axion Genome-Wide BOS 1 Array Plate (600k).....	10
Tabela 2. Cenários, painéis, número de animais da população referência (N), parentesco, quantidade de SNPs imputados (SNPs) e taxa de concordância de imputações utilizando o painel comercial Affymetrix (Axion Genome-Wide BOS 1 Array Plate).	12
Tabela 3. Cenários, painéis, número de animais da população referência (N), parentesco, quantidade de SNPs imputados (SNPs) e taxa de concordância de imputações utilizando o painel comercial Illumina (BovineHD Beadchip).....	13

LISTA DE ABREVIATURAS E SIGLAS

SNP - **S**ingle **N**ucleotide **P**olymorphism - Nucleotídeo de polimorfismo único

MAF - **M**inor **A**llele **F**requency - Frequência do alelo de menor frequência

AA23B20 - Cenário de imputação para o painel comercial Axion Genome-Wide BOS 1 Array Plate (**A**ffymetrix) utilizando a montagem **A**ntiga do genoma (UMD3.1) com **23** animais de **B**aixo parentesco na população referência a partir de painéis contendo **20** mil SNPs.

AA23A20 - Cenário de imputação para o painel comercial Axion Genome-Wide BOS 1 Array Plate (**A**ffymetrix) utilizando a montagem **A**ntiga do genoma (UMD3.1) com **23** animais de **A**lto parentesco na população referência a partir de painéis contendo **20** mil SNPs.

AA46B20 - Cenário de imputação para o painel comercial Axion Genome-Wide BOS 1 Array Plate (**A**ffymetrix) utilizando a montagem **A**ntiga do genoma (UMD3.1) com **46** animais de **B**aixo parentesco na população referência a partir de painéis contendo **20** mil SNPs.

AA46A20 - Cenário de imputação para o painel comercial Axion Genome-Wide BOS 1 Array Plate (**A**ffymetrix) utilizando a montagem **A**ntiga do genoma (UMD3.1) com **46** animais de **A**lto parentesco na população referência a partir de painéis contendo **20** mil SNPs.

AA23B50 - Cenário de imputação para o painel comercial Axion Genome-Wide BOS 1 Array Plate (**A**ffymetrix) utilizando a montagem **A**ntiga do genoma (UMD3.1) com **23** animais de **B**aixo parentesco na população referência a partir de painéis contendo **20** mil SNPs.

AA23A50 - Cenário de imputação para o painel comercial Axion Genome-Wide BOS 1 Array Plate (**A**ffymetrix) utilizando a montagem **A**ntiga do genoma (UMD3.1) com **23** animais de **A**lto parentesco na população referência a partir de painéis contendo **50** mil SNPs.

AA46B50 - Cenário de imputação para o painel comercial Axion Genome-Wide BOS 1 Array Plate (**A**ffymetrix) utilizando a montagem **A**ntiga do genoma (UMD3.1) com **46** animais de **B**aixo parentesco na população referência a partir de painéis contendo **50** mil SNPs.

AA46A50 - Cenário de imputação para o painel comercial Axion Genome-Wide BOS 1 Array Plate (Affymetrix) utilizando a montagem Antiga do genoma (UMD3.1) com **46** animais de **Alto** parentesco na população referência a partir de painéis contendo **50** mil SNPs.

AN23B20 - Cenário de imputação para o painel comercial Axion Genome-Wide BOS 1 Array Plate (Affymetrix) utilizando a montagem Nova do genoma (ARS-UCD1.2) com **23** animais de **Baixo** parentesco na população referência a partir de painéis contendo **20** mil SNPs

AN23A20 - Cenário de imputação para o painel comercial Axion Genome-Wide BOS 1 Array Plate (Affymetrix) utilizando a montagem Nova do genoma (ARS-UCD1.2) com **23** animais de **Alto** parentesco na população referência a partir de painéis contendo **20** mil SNPs.

AN46B20 - Cenário de imputação para o painel comercial Axion Genome-Wide BOS 1 Array Plate (Affymetrix) utilizando a montagem Nova do genoma (ARS-UCD1.2) com **46** animais de **Baixo** parentesco na população referência a partir de painéis contendo **20** mil SNPs.

AN46A20 - Cenário de imputação para o painel comercial Axion Genome-Wide BOS 1 Array Plate (Affymetrix) utilizando a montagem Nova do genoma (ARS-UCD1.2) com **46** animais de **Alto** parentesco na população referência a partir de painéis contendo **20** mil SNPs.

AN23B50 - Cenário de imputação para o painel comercial Axion Genome-Wide BOS 1 Array Plate (Affymetrix) utilizando a montagem Nova do genoma (ARS-UCD1.2) com **23** animais de **Baixo** parentesco na população referência a partir de painéis contendo **50** mil SNPs.

AN23A50 - Cenário de imputação para o painel comercial Axion Genome-Wide BOS 1 Array Plate (Affymetrix) utilizando a montagem Nova do genoma (ARS-UCD1.2) com **23** animais de **Alto** parentesco na população referência a partir de painéis contendo **50** mil SNPs.

AN46B50 - Cenário de imputação para o painel comercial Axion Genome-Wide BOS 1 Array Plate (Affymetrix) utilizando a montagem Nova do genoma (ARS-UCD1.2) com **46** animais de **Baixo** parentesco na população referência a partir de painéis contendo **50** mil SNPs.

AN46A50 - Cenário de imputação para o painel comercial Axion Genome-Wide BOS 1 Array Plate (Affymetrix) utilizando a montagem Nova do genoma (ARS-UCD1.2) com

46 animais de **Alto** parentesco na população referência a partir de painéis contendo **50** mil SNPs.

IA23B20 - Cenário de imputação para o painel comercial BovineHD BeadChip (Illumina) utilizando a montagem **Antiga** do genoma (UMD3.1) com **23** animais de **Baixo** parentesco na população referência a partir de painéis contendo **20** mil SNPs.

IA23A20 - Cenário de imputação para o painel comercial BovineHD BeadChip (Illumina) utilizando a montagem **Antiga** do genoma (UMD3.1) com **23** animais de **Alto** parentesco na população referência a partir de painéis contendo **20** mil SNPs.

IA46B20 - Cenário de imputação para o painel comercial BovineHD BeadChip (Illumina) utilizando a montagem **Antiga** do genoma (UMD3.1) com **46** animais de **Baixo** parentesco na população referência a partir de painéis contendo **20** mil SNPs.

IA46A20 - Cenário de imputação para o painel comercial BovineHD BeadChip (Illumina) utilizando a montagem **Antiga** do genoma (UMD3.1) com **46** animais de **Alto** parentesco na população referência a partir de painéis contendo **20** mil SNPs.

IA23B50 -Cenário de imputação para o painel comercial BovineHD BeadChip (Illumina) utilizando a montagem **Antiga** do genoma (UMD3.1) com **23** animais de **Baixo** parentesco na população referência a partir de painéis contendo **50** mil SNPs.

IA23A50 - Cenário de imputação para o painel comercial BovineHD BeadChip (Illumina) utilizando a montagem **Antiga** do genoma (UMD3.1) com **23** animais de **Alto** parentesco na população referência a partir de painéis contendo **50** mil SNPs.

IA46B50 - Cenário de imputação para o painel comercial BovineHD BeadChip (Illumina) utilizando a montagem **Antiga** do genoma (UMD3.1) com **46** animais de **Baixo** parentesco na população referência a partir de painéis contendo **50** mil SNPs.

IA46A50 - Cenário de imputação para o painel comercial BovineHD BeadChip (Illumina) utilizando a montagem **Antiga** do genoma (UMD3.1) com **46** animais de **Alto** parentesco na população referência a partir de painéis contendo **50** mil SNPs.

IN23B20 - Cenário de imputação para o painel comercial BovineHD BeadChip (Illumina) utilizando a montagem **Nova** do genoma (UMD3.1) com **23** animais de **Baixo** parentesco na população referência a partir de painéis contendo **20** mil SNPs.

IN23A20 - Cenário de imputação para o painel comercial BovineHD BeadChip (Illumina) utilizando a montagem **Nova** do genoma (UMD3.1) com **23** animais de **Alto** parentesco na população referência a partir de painéis contendo **20** mil SNPs.

IN46B20 - Cenário de imputação para o painel comercial BovineHD BeadChip (Illumina) utilizando a montagem **Nova** do genoma (UMD3.1) com **46** animais de **Baixo** parentesco na população referência a partir de painéis contendo **20** mil SNPs.

IN46A20 - Cenário de imputação para o painel comercial BovineHD BeadChip (Illumina) utilizando a montagem **Nova** do genoma (UMD3.1) com **46** animais de **Alto** parentesco na população referência a partir de painéis contendo **20** mil SNPs.

IN23B50 - Cenário de imputação para o painel comercial BovineHD BeadChip (Illumina) utilizando a montagem **Nova** do genoma (UMD3.1) com **23** animais de **Baixo** parentesco na população referência a partir de painéis contendo **20** mil SNPs.

IN23A50 - Cenário de imputação para o painel comercial BovineHD BeadChip (Illumina) utilizando a montagem **Nova** do genoma (UMD3.1) com **23** animais de **Alto** parentesco na população referência a partir de painéis contendo **50** mil SNPs.

IN46B50 - Cenário de imputação para o painel comercial BovineHD BeadChip (Illumina) utilizando a montagem **Nova** do genoma (UMD3.1) com **46** animais de **Baixo** parentesco na população referência a partir de painéis contendo **50** mil SNPs.

IN46A50 - Cenário de imputação para o painel comercial BovineHD BeadChip (Illumina) utilizando a montagem **Nova** do genoma (UMD3.1) com **46** animais de **Alto** parentesco na população referência a partir de painéis contendo **50** mil SNPs.

SUMÁRIO

1. INTRODUÇÃO	1
2. OBJETIVO	2
3. REVISÃO BIBLIOGRÁFICA	2
3.1. Imputação	2
3.2. Fatores que influenciam na acurácia de imputação	4
3.2.1. Metodologia computacional	4
3.2.2. Frequência do alelo de menor frequência (MAF)	5
3.2.3. Tamanho da população referência	6
3.2.4. Parentesco entre os indivíduos	6
3.2.5. Desequilíbrio de ligação	7
3.2.6. Densidade dos painéis utilizados	7
4. METODOLOGIA	8
4.1. Animais	8
4.2. Genotipagem	9
4.3. Cenários utilizados para imputação	9
4.4. Algoritmos para imputação	10
4.5. Eficiência da imputação	11
5. RESULTADOS E DISCUSSÃO	11
6. CONCLUSÕES	16
7. REFERÊNCIAS BIBLIOGRÁFICAS	16

1. INTRODUÇÃO

As tecnologias fundamentadas em informações genômicas tornaram-se importantes ferramentas para o progresso de estudos genéticos em animais. Dentre estas tecnologias, a utilização de marcadores do tipo polimorfismo de nucleotídeo único (SNP) promoveram grandes avanços na área de melhoramento genético animal, uma vez que permitem a aplicação da seleção genômica para características que apresentam baixa resposta ao processo de seleção tradicional, características que são medidas em momentos tardios na vida do animal e características de difícil mensuração ou alto custo para obtenção de fenótipos.

O polimorfismo de nucleotídeo único (SNP) consiste na mutação de nucleotídeos únicos no DNA. Estas mutações podem ocorrer em qualquer região do genoma, inclusive em regiões que possuem efeitos em características de interesse econômico, funcionando como sinalizadores de alterações responsáveis por promover variabilidade nas características importantes para a seleção.

A seleção genômica disponibiliza uma forma diferente de realizar a obtenção de valores genéticos animais, a qual utiliza além das informações fenotípicas e de genealogia dos animais, os efeitos dos marcadores moleculares (SNPs), que estão espalhados por todo o genoma, para a predição dos valores genéticos dos indivíduos (MEUWISSEN et al, 2001). Esta técnica preconiza a utilização do máximo de marcadores disponíveis e pode ser aplicada para múltiplas características, pois tem como base a regressão do fenótipo sobre genótipo dos animais na população referência, permitindo a avaliação dos efeitos dos marcadores moleculares sobre a característica avaliada e assim permite a obtenção do valor genético dos animais da população alvo que não possuem a observação do fenótipo.

Para obter resultado adequado da aplicação da seleção genômica, torna-se necessário a disponibilidade de animais genotipados com grande quantidade de marcadores moleculares, organizados em painéis com alta densidade. No entanto mesmo com a redução do custo de genotipagem nos últimos anos devido aos avanços das tecnologias, a obtenção da informação de uma população grande genotipada com muitos marcadores demanda elevados recursos financeiros. Dessa maneira, a técnica de imputação é apresentada como uma alternativa para redução de custo para aquisição de diversos animais genotipados com alta densidade (HAYES et al., 2011).

A técnica de imputação consiste na genotipagem de animais com menor quantidade de marcadores (painéis de menor densidade) e então, inferir por meio de modelos matemáticos os marcadores presentes em um painel com maior quantidade de marcadores (painel de maior densidade) usando como referência uma população genotipadas com esses painéis de maior densidade.

A aplicação da seleção genômica utilizando genótipos imputados é dependente da precisão que os marcadores moleculares foram inferidos e a métrica para avaliar a precisão é denominada de acurácia de imputação. Esta é dependente de diversos fatores como o tamanho da população referência, a densidade dos painéis envolvidos, a frequência dos alelos, o desequilíbrio de ligação entre os marcadores do painel de menor densidade com os do painel de maior densidade, entre outros. Neste contexto, a posição física a qual os marcadores se encontram no genoma referência tornam-se essenciais para a obtenção de adequada acurácia.

Embora atualmente a sequência bovina seja precisa, a identificação da posição de alguns marcadores presentes em regiões repetitivas pode ser desafiadora (HERMISDORFF et al., 2020). No entanto, devido ao aumento das informações e ao avanço das diferentes tecnologias foi possível a elucidação de algumas dessas regiões complexas, gerando a montagem nova do genoma denominada, ARS-UCD1.2, a qual possui diferenças nas posições de alguns marcadores quando comparado à montagem antiga utilizada anteriormente, denominada UMD3.1.

2. OBJETIVO

O objetivo do presente estudo foi avaliar a acurácia de imputação em 32 cenários distintos quanto a densidade, parentesco e quantidade de animais na população referência, utilizando a montagem nova do genoma (ARS-UCD1.2) e a montagem antiga (UMD3.1).

3. REVISÃO BIBLIOGRÁFICA

3.1. Imputação

Um dos principais objetivos no melhoramento animal é maximizar a produção nas condições ambientais existentes e, para atingir este objetivo, uma das ferramentas mais utilizadas é a seleção genética. Esta permite melhorias consideráveis nos aspectos reprodutivo, imunológico, digestivo e produtivos, como por

exemplo na produção de leite (LI et al., 2009). Para selecionar os animais com maior potencial genético é necessário a estimação dos valores genéticos dos animais participantes dos programas de melhoramento.

Por muitos anos estes valores genéticos eram estimados utilizando informações de fenótipos e genealogia. No entanto, um método diferente de estimação foi proposto por Meuwissen, Goddard e Hayes (2001), sendo denominado de seleção genômica. Este tem como base à utilização de informações fenotípicas, de genealogia e os efeitos dos marcadores moleculares espalhados por todo o genoma para a predição do valor genético de cada indivíduo, permitindo a seleção de características de difícil mensuração ou alto custo para obtenção.

Para a prática da seleção genômica, o principal marcador molecular utilizado é o polimorfismo de nucleotídeos único (“Single Nucleotide Polymorphism” - SNP), o qual possui grande distribuição ao longo do genoma (Caetano, 2009) e tecnologia disponível para sua detecção de forma eficiente. A disponibilidade dessas tecnologias desencadeou o avanço da oferta comercial e tornou disponíveis painéis que fornecem informações dos genótipos dos animais para diferentes quantidades de SNPs encontrados através do genoma. Dessa forma, os painéis comerciais possuem variadas densidades, como os painéis com densidade de 6 mil SNPs (6K), 50 mil SNPs (50K) e painéis customizados, os quais se tornaram opções com maior viabilidade para condução da seleção e outros estudos genômicos.

Painéis de alta densidade com informação acima de 50 mil SNPs (50K) permitiram o avanço da seleção genômica em relação a características de interesse econômico e também auxiliam na condução de diferentes estudos (BOVINE HAPMAP CONSORTIUM, 2009). No entanto, painéis com maior densidade têm elevado custo quando comparados aos painéis de menor densidade e, uma vez que o aumento da eficiência da seleção genômica está relacionado à quantidade de marcadores moleculares dos animais, tornando-se necessário a presença de muitos animais genotipados com painéis de alta densidade, a viabilidade para a aplicação da seleção genômica pode ser prejudicada pelos elevados custos. Dessa maneira, é possível utilizar uma técnica chamada de imputação, que consiste na genotipagem de animais de interesse com painéis de menor densidade, e então, utilizar cálculos matemáticos para inferir os SNPs presentes em painéis de maior densidade, com o auxílio da informação de uma população referência de animais genotipados com esses painéis de maior densidade (SARGOLZAEI; CHESNAIS; SCHENKEL, 2011).

A imputação é uma alternativa para a redução de custos, com possível aumento da acurácia dos valores genéticos genômicos, permitindo a presença de mais animais com genótipos em painéis de maior densidade e com menor custo para a aplicação da seleção genômica (KHATKAR et al., 2012).

A fim de obter elevada acurácia na seleção genômica, é necessário que na imputação, os animais tenham os genótipos inferidos para painéis de maior densidade com alta precisão, esta que por sua vez é chamada de acurácia de imputação. Diferentemente do que ocorre na prática, quando há o estudo sobre imputação, realiza-se a redução de um painel de maior densidade para um painel com menor densidade, para simular a ausência de informação de alguns SNPs, e infere-se esses marcadores moleculares da população imputada. Assim é possível comparar os genótipos inferidos com as informações verdadeiras, que foram ocultadas e, quanto maior a taxa de acerto na inferência dos marcadores SNPs, maior a acurácia de imputação.

Diferentes fatores podem afetar a acurácia de imputação como o tamanho da população referência, o parentesco entre os indivíduos da população referência e da população de imputação, o desequilíbrio de ligação entre os SNPs dos painéis utilizados, a frequência do alelo de menor frequência (MAF) e pela metodologia do programa computacional escolhido para realizar a imputação.

3.2. Fatores que influenciam na acurácia de imputação

3.2.1. Metodologia computacional

Existem diferentes programas computacionais que são utilizados para realizar a imputação, sendo que alguns utilizam apenas informações populacionais e outros informações populacionais adicionada às informações de genealogia. Alguns dos programas mais utilizados na literatura são Findhap (VIEIRA et al, 2014), BEAGLE (BUCKLING et al, 2009), IMPUTE2 (MARCHINI et al, 2007), FImpute (SARGOLZAEI et al, 2014), e Minimac3 (DAS et al, 2016).

O programa BEAGLE utiliza o método de imputação baseada na população referência (BROWNING BL, 2009), não dependendo de qualquer informação de genealogia para fazer a inferência estatística dos genótipos. Diferentemente, o programa FImpute utiliza o método de imputação com informações populacionais, adicionadas das informações de pedigree, se estas estiverem disponíveis. A acurácia de imputação não depende necessariamente apenas do método de imputação

escolhido, pois existem muitos fatores relacionadas a população em si, no entanto é possível observar diferença na eficiência computacional e no tempo para conclusão de tarefas quando comparado os diferentes programas (CARVALHEIRO et al., 2014).

Ventura et al. (2011), estudando diversos cenários com animais de raça pura e cruzados para a imputação de painéis de 6mil SNPs para painéis de 50 mil SNPs, observaram valores de acurácia variando entre 94,20% e 97,93% quando utilizado o FIMPUTE, variação de 95,35% e 98,31% com o IMPUTE2 e variação de 90,02% e 96,38% com o BEAGLE. Estes mesmos autores observaram que nos cenários utilizando animais cruzados, os resultados obtidos para a acurácia de imputação foram de 54,15% a 97,53%, de 57,04% e 97,46% e de 54,35% a 95,64%, para os programas FIMPUTE, IMPUTE2 e BEAGLE, respectivamente.

O programa BEAGLE também apresentou menor desempenho em estudo, de imputação na raça bovina Simental, em que Pausch et al. (2013) avaliaram cenários com 50, 100, 200 e 400 animais na população referência, e utilizaram os programas BEAGLE, Findhap.f90, MaCH e Minimac para imputar diferentes painéis de menor densidade para o painel de 50k. Estes autores verificaram que os programas MaCH e Minimac atingiram uma performance de acurácia de imputação maior que os demais, juntamente com maior precisão da inferência de alelos raros, à medida que o tamanho da população referência foi aumentada.

Semelhantemente, Das et al. (2016) comparando a acurácia de imputação por meio de quatro cenários distintos utilizando o programa BEAGLE, IMPUTE2 e Minimac3, observaram que o programa Minimac3 obteve performance levemente superior, principalmente para os alelos de menor frequência e, assim como observado por Pausch et al. (2013), a acurácia de imputação aumentou quando ocorreu o aumento na quantidade de SNPs dos painéis.

3.2.2. Frequência do alelo de menor frequência (MAF)

Além da metodologia escolhida, a frequência do alelo de menor frequência (MAF) também pode apresentar efeitos na acurácia de imputação. Esta pode demonstrar informações sobre a quantidade de alelos comuns ($MAF > \sim 5\%$) e alelos raros ($MAF \leq \sim 5\%$). A correta imputação para alelos com baixa MAF é importante pois muitas vezes os alelos raros atuam em características genéticas que possuem baixa expressão fenotípica (CIRULLI.; GOLDSTEIN, 2010). A imputação correta de alelos raros é mais difícil que a imputação de alelos comuns (MA et al., 2013), o que reduz

a acurácia de imputação total quando medida pela correlação, uma vez que regiões qualitativas do locus podem não ser inferidas corretamente.

Bernardes et al. (2019) conduzindo estudos de imputação para a combinação de dois painéis comerciais de alta densidade a partir de diferentes painéis de baixa densidades, demonstrou a relação entre a acurácia de imputação medida pela correlação em função da MAF e, observaram que a acurácia aumentou à medida que a MAF se elevou. Da mesma maneira, Hermisdorff et al. (2020), investigando a acurácia de imputação de painéis 50K para 777K de uma população referência de 500 e 2900 bovinos Nelore, concluíram que a MAF está positivamente relacionada com a acurácia de imputação.

3.2.3. Tamanho da população referência

O tamanho da população referência possui relação positiva com a acurácia de imputação, em que Hermisdorff et al. (2020) verificaram em bovinos da raça Nelore que a acurácia de painéis 50K para 777K de SNPs, medida pela correlação, aumentou de 0,845 para 0,917 quando a quantidade de animais na população referência modificou de 250 para 2000 animais. Neste mesmo estudo, o aumento na acurácia de imputação foi observado também quando números de animais ainda maiores foram utilizados (500 animais referência e 2900 animais imputados), com a acurácia de imputação com acréscimos de 0,903 para 0,913.

O tamanho da população referência aumenta a informação dos haplótipos presentes na população, auxiliando nas inferências dos genótipos a serem imputados, ao mesmo tempo o tamanho populacional torna-se mais importante quando os animais a serem imputados tem baixo parentesco com os animais da população referência (CALUS et al., 2014).

3.2.4. Parentesco entre os indivíduos

Desta forma, o parentesco entre os indivíduos também influencia a acurácia da imputação, de maneira que os indivíduos na população de imputação que possuem parentesco com animais na população referência, possuem maior acurácia de imputação. Esta acurácia pode ser ainda maior quando parentes próximos têm SNPs presentes na população referência (ZHANG et al., 2010). Neste contexto, Pausch et al. (2013), identificou que a acurácia de imputação é positivamente relacionada à quantidade de animais e do parentesco na população referência, em que os cenários

com 50 a 100 animais na população referência com alto grau de parentesco apresentaram maior acurácia de imputação quando comparados com cenários de animais com baixo parentesco.

Hozé et al. (2013), ao realizar a imputação do painel comercial BovineSNP50 BeadChip com 50 mil SNP (50k) para o painel de alta densidade BovineHD BeadChip (700k) utilizando o software BEAGLE em uma população com 16 raças bovinas e 5153 animais, constatou que a acurácia de imputação foi de 99%, variando alguns pontos percentuais entre as raças, por exemplo 97,06% e 99,96% para as raças Simmental e Normande, respectivamente.

3.2.5. Desequilíbrio de ligação

O desequilíbrio de ligação, que é interpretado como a associação não-aleatória de dois alelos em um ou mais genes em uma população (PEIXOTO et al., 2013). Painéis que apresentam SNPs com grande desequilíbrio de ligação entre eles, apresentam maior acurácia de imputação devido à baixa variabilidade de haplótipos presentes na população imputada (MEUWISSEN et al., 2009). A baixa variabilidade produzida pelo alto desequilíbrio de ligação acaba por reduzir as possibilidades de combinações de haplótipos para serem imputados, além de aumentar a probabilidade de localização de regiões que abrigam genes de características quantitativas de interesse (HABIER et al., 2009). Existe uma relação entre o desequilíbrio de ligação e o parentesco entre os indivíduos da população referência com a população de imputação, além da relação deste desequilíbrio com a densidade dos painéis.

3.2.6. Densidade dos painéis utilizados

A densidade também apresenta efeitos na acurácia de imputação, em que, em geral, quanto maior é a quantidade de SNPs que deve ser imputado, menor é a acurácia de imputação. No entanto, painéis com baixa densidade, mas que possuem alto desequilíbrio de ligação entre os SNPs, podem permitir uma acurácia de imputação maior quando imputados para painéis de alta densidade.

Boichard et al. (2012), ao realizar a imputação de dois painéis de baixa densidade (BovineLD - 7K e Bovine3K - 3K) para um painel de maior densidade (BovineSNP50 - 50K), constataram que a imputação do painel BovineLD apresentava maior acurácia de imputação em todos os cenários estudados e, este resultado foi justificado devido a maior quantidade de SNPs do painel de baixa densidade.

Semelhantemente, Khatkar et al. (2012), na avaliação da acurácia de imputação de painéis de menor densidade (3K até 50K) imputados para painéis de alta densidade (700K) em animais da raça Holandesa concluíram que painéis de alta densidade imputados a partir de painéis de 50K possuem alta taxa de genótipos imputados corretamente, com resultado igual a 99,21% e, quando imputado a partir de painéis de baixa densidade 3K e 7K, obtiveram taxas iguais a 95,22% e 98%, respectivamente.

Zhang et al. (2010) realizaram a imputação de painéis de baixa densidade para um painel customizado de 60K em bovinos da raça Holandesa. Estes autores avaliaram cinco painéis de menor densidade, contendo 384 SNPs, 768 SNPs, 1536 SNPs, 3000 SNPs (3K) e 6000 SNPs (6K) e utilizando uma população referência de 1000 indivíduos, resultando em acurácia de imputação de 88,3% e 98% ao utilizar o painel 384 SNP e o painel 6K respectivamente. Neste estudo foi possível verificar que à medida que o número de SNPs presentes no painel de menor densidade cresce, a taxa de erro na acurácia de imputação é reduzida. Esta relação entre o tamanho do painel de menor densidade e a acurácia também foi descrita por Druet et al. (2010), concluindo que a acurácia pode ser aumentada com a adição de SNPs no painel a ser imputado.

4. METODOLOGIA

4.1. Animais

Foram utilizadas as informações cedidas pelo Laboratório de Melhoramento Animal – UFSC de 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 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.

A seleção dos progenitores foi realizada a partir de consultas aos catálogos das centrais de inseminação artificial disponíveis no país. A partir de um número total inicial de 616 touros Nelore (mochos ou aspados), foram selecionados 34 touros ativos na população para a composição desta amostra, ou seja, touros que possuíam sêmen

disponível no mercado, cujo valor não excedesse R\$ 50,00 (tousos acessíveis ao produtor típico da raça). Outro critério de seleção dos progenitores foi que os mesmos representassem as principais linhagens e genealogias que compõem a raça Nelore, ao mesmo tempo em que se buscou minimizar o grau de parentesco entre os progenitores.

4.2. Genotipagem

Os 814 animais da raça Nelore envolvidos nesse estudo foram genotipados com o BovineHD BeadChip – Illumina, e uma parcela de 94 animais (23 touros e 71 progênes) que fazem parte do grupo total dos 814 animais, também foram genotipados com o Axion Genome-Wide BOS 1 Array Plate – Affymetrix no Laboratório Multiusuários Centralizado de Genômica Funcional Aplicada à Agropecuária e Agroenergia, Piracicaba, São Paulo. O controle de qualidade dos genótipos foi feito para cada painel separadamente e teve como finalidade de remover erros de genotipagem excluindo SNPs localizados em regiões não autossômicas, com posições desconhecidas, com desvios significativos ($p < 0,00001$) do equilíbrio de Hardy-Weinberg, com frequência do alelo de menor frequência (MAF) menor que 0,001, com taxa de leitura (“call rate”) menor que 0,98 e amostras que apresentaram “call rate” menor que 0,90. Após o controle de qualidade realizado no programa PLINK, restaram 801 animais com 537.741 SNPs para o painel Illumina e, 94 animais com 428.060 SNPs para o painel Affymetrix. Assim as análises foram realizadas utilizando apenas estes animais que passaram do controle de qualidade, sendo 801 animais para o painel Illumina e 94 para o painel Affymetrix. Os genótipos destes animais foram faseados utilizando o programa EAGLE e posteriormente imputados utilizando o programa computacional Minimac3.

4.3. Cenários utilizados para imputação

Foram utilizados 16 diferentes cenários considerando a montagem ARS-UCD1.2 do genoma bovino e os mesmos 16 cenários considerando a montagem UMD3.1 (Tabela 1), em que a população de animais referência é a população que foi mantida todos os SNPs do painel de alta densidade. Os demais animais da população tiveram seus SNPs “removidos” com a finalidade de mimetizar que estes foram genotipados com painéis comerciais de baixa densidade contendo 20 mil SNPs (20k) ou 50 mil SNPs (50k). Para formar alguns cenários foram considerados 23 animais na

população referência, sendo selecionados os 23 animais com maior parentesco entre si e, em outros cenários utilizaram-se 23 animais com menor parentesco. Outros cenários foram formados por 46 animais na população referência, sendo alguns destes utilizando 46 animais com maior parentesco entre si, e outros utilizando 46 animais com menor parentesco. Essas diferentes constituições de população referência foram consideradas para ambos painéis estudados (Illumina e Affymetrix). Assim, foi realizada a imputação do painel de baixa densidade para o painel de alta densidade utilizando estes diferentes animais na população referência. Então foi verificada a acurácia de imputação destes diferentes cenários entra a montagem nova (ARS-UCD1.2) e montagem antiga (UMD3.1) do genoma bovino.

Tabela 1. Cenários utilizados para o estudo de imputação para os painéis BovineHD BeadChip (700k) e Axion Genome-Wide BOS 1 Array Plate (600k).

Animais na população referência	Parentesco	Painel imputação	Painel referência
23	Baixo	20K	700k
			600k
		50K	700K
			600K
46	Baixo	20K	700k
			600k
		50K	700K
			600K
23	Alto	20K	700k
			600k
		50K	700K
			600K
46	Alto	20K	700k
			600k
		50K	700K
			600K

4.4. Algoritmos para imputação

Foi utilizado o algoritmo que leva em consideração informações do desequilíbrio de ligação entre os SNPS, pré-faseada pelo programa EAGLE e posteriormente imputadas pelo programa Minimac3. O algoritmo de imputação do Minimac3 utiliza como procedimentos a identificação de regiões que possam ser inferidas pelas informações da população na reconstrução dos haplótipos com alto grau de confiabilidade.

4.5. Eficiência da imputação

A eficiência de imputação por animal foi estimada nos animais da população de imputação que foram genotipados com os painéis de alta densidade e tiveram seus SNPs “removidos” para serem inferidos pelo programa. Assim, esta acurácia de imputação realizada comparando os alelos dos marcadores imputados com os alelos observados e calculando a razão da proporção de genótipos imputados corretamente sobre todos os genótipos imputados, chamada de taxa de concordância (TC), esta que por sua vez, será considerada como a acurácia de imputação dos cenários descritos.

5. RESULTADOS E DISCUSSÃO

A partir dos cenários propostos é possível comparar as acurácias de imputação, sendo que as principais diferenças entre os cenários (Tabelas 2 e 3) foram a quantidade de SNPS imputados (densidade), a quantidade de animais considerados na população referência, o parentesco entre os animais da população referência com os animais da população imputação e o mapa utilizado (UMD3.1 e ARS-UCD1.2).

As diferenças de acurácia entre os cenários considerando 23 touros na população referência e considerando 46 touros, genotipados com painel Axion Genome-Wide BOS 1 Array Plate (Affymetrix), podem ser observadas ao comparar os cenários AA23A20 com AA46A20 em um aumento de 68,02% para 75,13%, AN23A20 com AN46A20 com um aumento de 68% para 75,16%, AA23A50 com AA46A50 com um aumento de 77,56% para 86,96% e AN23A50 com AN46A50 com um aumento de 78,49% e 86,88%. (Tabela 2). É possível verificar aumento na acurácia de imputação quando considerado maior número de indivíduos na população referência para os cenários em que havia alto parentesco entre a população referência e população imputação.

Pausch et al. (2013), ao realizar a imputação para um painel de alta densidade em 4 cenários distintos, com 50, 100, 200 e 400 animais na população referência, os quais tiveram respectivamente 78,03%, 89,21%, 97,47% e >99% de representatividade genética da população total, obtiveram resultados semelhantes, em que a acurácia de imputação aumentou conforme o número de animais com alto parentesco aumentava na população referência. Estes autores observaram elevada

acurácia de imputação mesmo em cenários com poucos animais na população referência e utilizando algoritmos computacionais baseados apenas em informações populacionais.

Tabela 2. Cenários, painéis, número de animais da população referência (N), parentesco, quantidade de SNPs imputados (SNPs) e taxa de concordância de imputações utilizando o painel comercial Affymetrix (Axion Genome-Wide BOS 1 Array Plate).

Cenários	Painel comercial	N	Parentesco	Painel imputação	Painel referência	SNPs	TC
AA23B20	Affymetrix	23	Baixo	20K	600k	418164	69,59
AA23B50	Affymetrix	23	Baixo	50K	600k	401708	80,55
AA46B20	Affymetrix	46	Baixo	20K	600k	400732	68,01
AA46B50	Affymetrix	46	Baixo	50K	600k	385478	78,49
AA23A20	Affymetrix	23	Alto	20K	600k	409667	68,02
AA23A50	Affymetrix	23	Alto	50K	600k	393778	77,56
AA46A20	Affymetrix	46	Alto	20K	600k	412832	75,13
AA46A50	Affymetrix	46	Alto	50K	600k	396791	86,96
AN23B20	Affymetrix	23	Baixo	20K	600k	418157	69,58
AN23B50	Affymetrix	23	Baixo	50K	600k	401699	80,5
AN46B20	Affymetrix	46	Baixo	20K	600k	409660	68
AN46B50	Affymetrix	46	Baixo	50K	600k	393770	77,59
AN23A20	Affymetrix	23	Alto	20K	600k	400725	68
AN23A50	Affymetrix	23	Alto	50K	600k	385469	78,49
AN46A20	Affymetrix	46	Alto	20K	600k	412825	75,16
AN46A50	Affymetrix	46	Alto	50K	600k	396782	86,88

Em cenários que a imputação foi realizada para o painel BovineHD BeadChip (Illumina) observaram-se diferenças na acurácia de imputação entre IN23B20 e IN46B20 com um aumento de 74,02% para 85,46%, IN23A20 e IN46A20 com um aumento de 72,36% para 81,82%, IN23B50 e IN46B50 com aumento de 85,89% para 91,68% e, por fim, entre IN23A50 e IN46A50 com aumento de 83,5% para 90,23% na taxa de concordância (Tabela 3). O aumento do número de animais na população referência apresentou melhora na acurácia apenas quando considerado a montagem ARS-UCD1.2 do genoma bovino, entretanto, esse resultado não se repete considerado a montagem UMD3.1. Provavelmente isso pode ter ocorrido devido a modificação da posição de alguns SNPs, reduzindo a homogeneidade da distribuição de SNPs separados com distância média de 3,4Kb que foi considerado ao formar o painel BovineHD BeadChip. Assim, o maior número de animais na população referência possivelmente permitiu melhor representatividade de haplótipos, elevando a acurácia de imputação.

Tabela 3. Cenários, painéis, número de animais da população referência (N), parentesco, quantidade de SNPs imputados (SNPs) e taxa de concordância de imputações utilizando o painel comercial Illumina (BovineHD Beadchip).

Cenários	Painel comercial	N	Parentesco	Painel imputação	Painel referência	SNPs	TC
IA23B20	Illumina	23	Baixo	20K	700k	500867	73,92
IA23B50	Illumina	23	Baixo	50K	700k	405433	85,92
IA46B20	Illumina	46	Baixo	20K	700k	495378	72,29
IA46B50	Illumina	46	Baixo	50K	700k	479140	83,39
IA23A20	Illumina	23	Alto	20K	700k	522599	84,57
IA23A50	Illumina	23	Alto	50K	700k	504953	91,61
IA46A20	Illumina	46	Alto	20K	700k	511816	81,58
IA46A50	Illumina	46	Alto	50K	700k	494791	90,14
IN23B20	Illumina	23	Baixo	20K	700k	500867	74,02
IN23B50	Illumina	23	Baixo	50K	700k	484330	85,89
IN46B20	Illumina	46	Baixo	20K	700k	522599	85,46
IN46B50	Illumina	46	Baixo	50K	700k	504953	91,68
IN23A20	Illumina	23	Alto	20K	700k	495378	72,36
IN23A50	Illumina	23	Alto	50K	700k	479140	83,5
IN46A20	Illumina	46	Alto	20K	700k	511816	81,82
IN46A50	Illumina	46	Alto	50K	700k	494791	90,23

A densidade do painel utilizado para realizar a imputação interfere diretamente na acurácia de imputação, uma vez que considerando os cenários que a imputação foi realizada para o painel Axion Genome-Wide BOS 1 Array Plate (Affymetrix), quando houve menor quantidade de SNPs a serem inferidos, ocorreu aumento na acurácia de imputação. O mesmo foi observado quando o painel de maior densidade era o BovineHD BeadChip (Illumina), em que quando o painel de menor densidade possuía maior quantidade de SNPs, resultou em alta acurácia.

Estes resultados são semelhantes aos obtidos por Boichard et al. (2012), Khatkar et al. (2012) e Zhang et al. (2010), que realizaram imputações em cenários distintos partindo de painéis imputação de menor densidade e maior densidade, obtendo acurácia de imputação superiores em todos os cenários que foram utilizados painéis imputação de maior densidade quando comparado com outros cenários de seus respectivos estudos.

Nos cenários em que o parentesco dos animais era maior, espera-se maior acurácia de imputação, uma vez que a proximidade na genealogia possui elevada influência sob a acurácia, sendo mais notável quando parentes de animais da população de imputação estão na população referência (SARGOLZAEI et al., 2014). Assim, ao comparar os cenários AA46B20 com AA46A20, AA46B50 com AA46A50, AN46B20 com AN46A20, e AN46B50 com AN46A50 (Tabela 2), foi possível identificar

aumento na acurácia de imputação quando o painel de maior densidade era o Axion Genome-Wide BOS 1 Array Plate (Affymetrix) e considerados 46 animais na população referência com alto parentesco. De maneira semelhante, ao investigar 4 cenários imputação com diferentes populações referências, García-Ruiz et al. (2015), observaram que a acurácia de imputação aumentou conforme o número de ancestrais utilizados na população referência aumentou.

Os cenários de imputação para o painel BovineHD BeadChip (Illumina) elevaram a acurácia de imputação ao considerar maior parentesco na população referência apenas quando considerado a montagem UMD3.1 do genoma bovino, como observado na comparação entre os cenários IA23B20 com IA23A20, IA23B50 com IA23A50, IA46B20 com IA46A20 e IA46B50 com IA46A50, com aumentos na taxa de concordância de 73,92% para 84,57%, 85,92% para 91,61%, 72,29% para 81,58% e 83,39% para 90,14%, respectivamente (Tabela 3). Esses resultados observados, quando considerado aumento no número de animais e aumento de parentesco da população referência, sugeriram que a alteração no tamanho da população referência apresentou mais efeito quando considerada a montagem ARS-UCD1.2 do genoma bovino e a alteração no parentesco teve maior influência quando considerada a montagem UMD3.1.

Quando considerado o mesmo tamanho de população referência e o mesmo parentesco, os cenários em que o painel de maior densidade era BovineHD BeadChip (Illumina) apresentaram diferenças na acurácia de imputação ao comparar as diferentes montagens do genoma bovino (UMD3.1 com ARS-UCD1.2). Os cenários IA46B20 em comparação ao IN46B20 e IA46B50 comparado ao IN46B50 (Tabela 3) apresentaram maior acurácia de imputação na nova montagem do genoma (ARS-UCD1.2), sendo que estes eram cenários com maior população referência, porém com menor parentesco entre os animais. Já os cenários IA23A20 comparado ao IN23A20 e IA23A50 comparado ao IN23A50 (Tabela 3), os quais apresentaram menor população referência com maior parentesco, resultaram em melhor acurácia de imputação na montagem antiga do genoma bovino (UMD3.1).

A diferença na acurácia de imputação dos painéis imputados para o painel BovineHd Beadchip (Illumina) nas montagens UMD3.1 e ARS-UCD 1.2 foi observada possivelmente pela alteração na posição de alguns SNPs afetando a acurácia de algumas regiões. Em contraste a estes resultados, Milanesi et al. (2014) ao realizarem a comparação da acurácia de imputação em 4 diferentes cenários de imputação de

baixa para média densidade, utilizando as montagens dos painéis genômicos BTAU4.2, BTAU 4.6, e UMD3.1 e, 4 diferentes métodos de imputação computacional, não encontraram diferenças significativas entre os cenários imputados.

Hermisdorff et al. (2020), ao compararem as montagens UMD3.1 e ARS-UCD1.2, observaram pequena diferença na média da acurácia de imputação entre as duas montagens. Entretanto, relataram regiões com baixa acurácia de imputação, potencialmente associadas a SNPs incorretamente posicionados na montagem do painel. Estes autores relataram regiões da montagem UMD3.1 que obtiveram maior acurácia de imputação quando comparada as mesmas regiões da montagem ARS-UCD1.2. Também observaram regiões da montagem ARS-UCD1.2 com maior acurácia de imputação quando comparada as mesmas regiões da montagem UMD3.1. Os resultados do presente estudo podem ter sido influenciados por regiões como estas, em que foi possível observar menor taxa de concordância em cromossomos como o 9 e o 21 na montagem ARS-UCD1.2 (Figura 1). No entanto sugere-se mais estudos sobre essas regiões e maior compreensão sobre a interação destas com a acurácia de imputação, recomendando-se explorar também outros cenários de imputação.

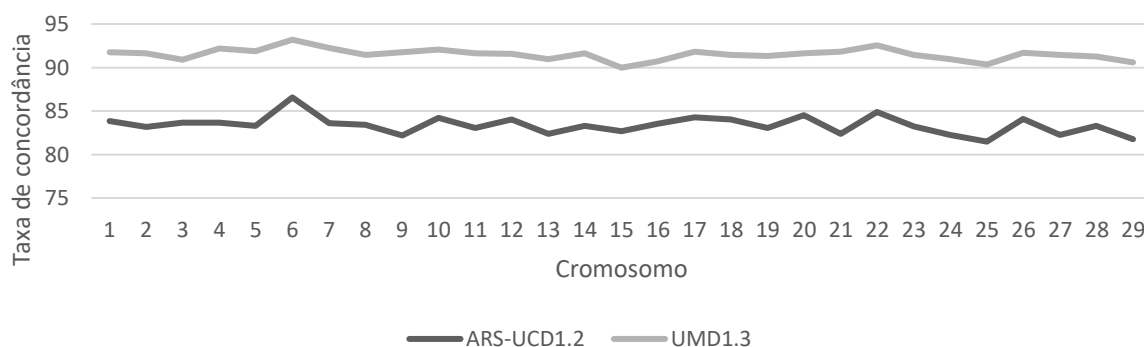


Figura 2. Taxa de concordância por cromossomo na montagem UMD3.1 e ARS-UCD1.2 no painel BovineHd Beadchip (Illumina).

Diferentemente dos painéis imputados para o Illumina, os cenários que foram imputados para o Axion Genome-Wide BOS 1 Array Plate (Affymetrix) não apresentaram diferenças na acurácia quando comparados as diferentes montagens de genoma (Tabela 2). Isso pode ter ocorrido devido as diferenças nas escolhas dos SNPs para compor o painel, as quais consideraram maximizar as informações de grupos SNPs menores com maior desequilíbrio de ligação, o que pode ter resultado

em menor influência na acurácia ao modificar as posições dos SNPs quando ocorreu a nova montagem do genoma.

De maneira geral, aumentar a densidade dos painéis a serem imputados representou aumento na acurácia de todos os cenários. Para o painel Axion Genome-Wide BOS 1 Array Plate (Affymetrix), aumentar a população referência elevou a acurácia de imputação para os cenários em que havia alto parentesco entre a população referência e população imputação. Nos cenários de maior população referência, aumentar o parentesco dos animais elevou a acurácia de imputação. No painel BovineHD BeadChip (Illumina) foi possível observar o aumento na acurácia de imputação com aumento da população referência na montagem ARS-UCD1.2. Enquanto que aumentar o parentesco dos animais elevou a acurácia de imputação apenas na montagem UMD3.1. Em geral, as diferenças na acurácia de imputação observadas entre as montagens ARS-UCD1.2 e UMD3.1 podem ter ocorrido devido ao efeito das mudanças de posição dos SNPs nos diferentes painéis, em que a acurácia média de alguns cromossomos pode ter sido reduzida na nova montagem. Assim, sugere-se estudos posteriores para maior compreensão das regiões cromossômicas e posição dos SNPs.

6. CONCLUSÕES

A acurácia de imputação em muitos cenários aumentou com a maior densidade dos painéis a serem imputados, com o aumento do parentesco e da quantidade de animais na população referência. Houve modificações das acurácias de imputação de acordo com a montagem utilizada (UMD3.1 ou ARS-UCD1.2) para alguns cenários, podendo esta ter sido influenciada pela mudança de posição dos SNPs, o que facilitou a realização da imputação em alguns cenários e dificultou em outros.

7. REFERÊNCIAS BIBLIOGRÁFICAS

BERNARDES P.A.; NASCIMENTO G.B.; SAVEGNAGO R.P.; BUZANSKAS M.E.; WATANABE R.N.; DE ALMEIDA REGITANO L.C.; 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*. 2019 Nov 29;9(1):1-0.

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 Mar 28;7(3):e34130.

BOVINE GENOME SEQUENCING AND ANALYSIS CONSORTIUM.; ELSIK C.G.; TELLAM R.L.; WORLEY K.C.; GIBBS R.A.; MUZNY D.M.; WEINSTOCK G.M.; ADELSON D.L.; EICHLER E.E.; ELNITSKI L.; GUIGÓ R.; The genome sequence of taurine cattle: a window to ruminant biology and evolution. Science. 2009 Apr 24;324(5926):522-8.

BOVINE HAPMAP CONSORTIUM.; GIBBS R.A.; TAYLOR J.F.; VAN TASSELL C.P.; BARENDSE W.; EVERSOLE KA.; GILL C.A.; GREEN R.D.; HAMERNIK D.L.; KAPPES S.M.; LIEN S.; Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. Science. 2009 Apr 24;324(5926):528-32.

BUCKLING A, CRAIG MACLEAN R, BROCKHURST M.A.; COLEGRAVE N. The Beagle in a bottle. Nature. 2009 Feb;457(7231):824-9.

BROWNING, B.L.; BROWNING, S.R. Improving the accuracy and efficiency of identity-by-descent detection in population data. Genetics, New York, v. 194, n. 2, p. 459–471, 2013

CAETANO A.R.; Marcadores SNP: conceitos básicos, aplicações no manejo e no melhoramento animal e perspectivas para o futuro. Revista Brasileira de Zootecnia. 2009; 38:64-71.

CARVALHEIRO R.; BOISON SA.; NEVES H.H.; SARGOLZAEI M.; SCHENKEL F.S.; UTSUNOMIYA Y.T.; O'BRIEN A.M.; SÖLKNER J.; MCEWAN J.C.; VAN TASSELL C.P.; SONSTEGARD TS.; Accuracy of genotype imputation in Nelore cattle. Genetics Selection Evolution. 2014 Dec;46(1):1-1.

CALUS M.P.; 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. 2014 Nov;8(11):1743-53.

CIRULLI E.T.; GOLDSTEIN D.B.; Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nature Reviews Genetics*. 2010 Jun;11(6):415-25.

DAS S.; FORER L.; SCHÖNHERR S.; SIDORE C.; LOCKE AE.; KWONG A.; VRIEZE SI.; CHEW EY.; LEVY S.; MCGUE M.; SCHLESSINGER D.; Next-generation genotype imputation service and methods. *Nature genetics*. 2016 Oct;48(10):1284-7.

DRUET T.; SCHROOTEN C.; DE ROOS AP.; Imputation of genotypes from different single nucleotide polymorphism panels in dairy cattle. *Journal of dairy science*. 2010 Nov 1;93(11):5443-54.

HAYES, B. J.; BOWMAN, P. J.; DAETWYLER, H. D.; KIJAS, J. W.; VAN DER WERF, J. H. J. Accuracy of genotype imputation in sheep breeds. *Animal Genetics*, v 43, p. 72-80, 2011.

HERMISDORFF I.D.; COSTA R.B.; DE ALBUQUERQUE L.G.; PAUSCH H.; KADRI NK.; Investigating the accuracy of imputing autosomal variants in Nellore cattle using the ARS-UCD1. 2 assembly of the bovine genome. *BMC genomics*. 2020 Dec;21(1):1-2.

HOZÉ C.; FOUILLOUX MN.; VENOT E.; GUILLAUME F.; DASSONNEVILLE R.; FRITZ S.; DUCROCQ V.; PHOCAS F.; BOICHARD D.; CROISEAU P.; High-density marker imputation accuracy in sixteen French cattle breeds. *Genetics Selection Evolution*. 2013 Dec;45(1):1-1.

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 Dec;13(1):1-2.

MA P.; BRØNDUM 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*. 2013 Jul 1;96(7):4666-77.

MARCHINI J.; HOWIE B.; MYERS S.; MCVEAN G.; DONNELLY P.; A new multipoint method for genome-wide association studies by imputation of genotypes. *Nature genetics*. 2007 Jul;39(7):906-13.

MEUWISSEN, T. H. E.; GODDARD, M. E.; HAYES, B. J. Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, v. 157, p. 1819-1829, 2001.

MILANESI M, VICARIO D, STELLA A, VALENTINI A, AJMONE-MARSAN PA, BIFFANI S, BISCARINI F, JANSEN G, NICOLAZZI EL., Imputation accuracy is robust to cattle reference genome updates. *Animal Genetics*. 2015 Feb;46(1):69-72.

PAUSCH H.; AIGNER B.; EMMERLING R.; EDEL C.; GÖTZ K.U.; FRIES R.; Imputation of high-density genotypes in the Fleckvieh cattle population. *Genetics Selection Evolution*. 2013 Dec;45(1):1-0.

PEIXOTO M.G.; DE ABREU SANTOS D.J.; SANTOS M.R.; CARVALHO M.A.; DA SILVA M.V.; TONHATI H. Desequilíbrio de ligação com um painel de alta densidade em genomas de bovinos da raça Guzerá1.

SARGOLZAEI, M.; CHESNAIS, J. P.; SCHENKEL, F.S.; FImpute. An efficient imputation algorithm for dairy cattle populations. *Journal of Animal Science*, v. 89(E-Suppl. 1)/*Journal of Dairy Science*, v. 94(E-Suppl. 1), p. 421 (abstr. 333). 2011

SARGOLZAEI, M.; CHESNAIS, J. P.; SCHENKEL, F.S.; A new approach for efficient genotype imputation using information from relatives. *BMC genomics*. 2014 Dec;15(1):1-2.

VENTURA R.V.; LU D.; SCHENKEL F.S.; WANG Z.; LI C.; MILLER S.P.; 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*. 2014 Apr 1;92(4):1433-44.

VIEIRA G.; CARNICER M.; PORTAIS J.C.; HEUX S.; FindPath: a Matlab solution for in silico design of synthetic metabolic pathways. *Bioinformatics*. 2014 Oct 15;30(20):2986-8.

YANG W.; YANG Y.; ZHAO C.; YANG K.; WANG D.; YANG J.; NIU X.; GONG J.; Animal-ImputeDB: a comprehensive database with multiple animal reference panels for genotype imputation. *Nucleic Acids Research*. 2020 Jan 8;48(D1):D659-67

MA P, BRØNDUM RF, ZHANG Q, LUND MS, SU G. Comparison of different methods for imputing genome-wide marker genotypes in Swedish and Finnish Red Cattle. *Journal of dairy science*. 2013 Jul 1;96(7):4666-77.

ZHANG Z.; DRUET T.; Marker imputation with low-density marker panels in Dutch Holstein cattle. *Journal of Dairy Science*. 2010 Nov 1;93(11):5487-94.