

Universidade de São Paulo  
Faculdade de Medicina Veterinária e Zootecnia

ALANA SELLI

**Impactos da inclusão de corridas de homozigosidade em processos de acasalamento dirigido em uma população simulada de bovinos de corte**

Pirassununga  
2024

ALANA SELLI

**Impactos da inclusão de corridas de homozigosidade em processos de acasalamento dirigido em uma população simulada de bovinos de corte**

**VERSÃO CORRIGIDA**

Dissertação apresentada ao Programa de Pós-Graduação em Nutrição e Produção Animal da Faculdade de Medicina Veterinária e Zootecnia da Universidade de São Paulo para obtenção do título de Mestre em Ciências.

**Departamento:**

Nutrição e Produção Animal

**Área de concentração:**

Nutrição e Produção Animal

**Orientador:**

Prof. Dr. Ricardo Vieira Ventura (VNP-FMVZ/USP)

**Coorientadores:**

Prof. Dr. Luiz Fernando Brito (Purdue University)  
Prof. Dr. Marcos Eli Buzanskas (CCA-UFPB)

Pirassununga  
2024

Autorizo a reprodução parcial ou total desta obra, para fins acadêmicos, desde que citada a fonte.

#### Catálogo na Publicação

Biblioteca Virginie Buff D'Ápice da Faculdade de Medicina Veterinária e Zootecnia da  
Universidade de São Paulo  
Ficha catalográfica gerada automaticamente com os dados fornecidos pelo(a) autor(a).

Selli, Alana  
Impactos da inclusão de corridas de homozigossidade em processos de acasalamento dirigido em uma população simulada de bovinos de corte / Alana Selli ; orientador Prof. Dr. Ricardo Vieira Ventura ; coorientador Prof. Dr. Luiz Fernando Brito e Prof. Dr. Marcos Eli Buzanskas - versão corrigida.-- Pirassununga, 2024.  
110 f. : il.

Dissertação (Mestrado - Programa de Pós-Graduação em Nutrição e Produção Animal - Departamento de Nutrição e Produção Animal) - Faculdade de Medicina Veterinária e Zootecnia, Universidade de São Paulo, 2024.

1. Homozigossidade. 2. Seleção genômica. 3. Simulação de dados. I. Título.

Bibliotecária responsável pela estrutura de catalogação  
na publicação: Maria Aparecida Laet - CRB 5673-8.

## FOLHA DE AVALIAÇÃO

Autor: SELLI, Alana

Título: Impactos da inclusão de corridas de homozigosidade em processos de acasalamento dirigido em uma população simulada de bovinos de corte

Dissertação apresentada ao Programa de Pós-Graduação em Nutrição e Produção Animal da Faculdade de Medicina Veterinária e Zootecnia da Universidade de São Paulo para obtenção do título de Mestre em Ciências.

Data: 19 / 03 / 2024

### Banca Examinadora

Prof. Dr. Ricardo Vieira Ventura

Instituição: VNP/FMVZ/USP Julgamento: Aprovado

Prof. Dr. Hinayah Rojas de Oliveira

Instituição: Purdue University Julgamento: Aprovado

Prof. Dr. Fernando de Oliveira Bussiman

Instituição: University of Georgia Julgamento: Aprovado

## DEDICATÓRIA

*Àqueles que permitiram que eu me tornasse quem sou hoje:  
Meus pais Alcir e Sirlei.*

*E àqueles que acompanharam de perto esta jornada:  
Daiana e Felipe.*

*Dedico.*

## AGRADECIMENTOS

*Aos meus pais Alcir e Sirlei por terem me ensinado desde cedo a ter compaixão pelas pessoas e animais, e por não medirem esforços para proporcionar os meios para a minha felicidade e sucesso.*

*À minha irmã Daiana, minha melhor amiga e um grande exemplo de dedicação e esforço, e ao meu cunhado Bruno, por cuidar de mim desde pequena como uma filha. Sem a ajuda dos dois, é possível que este trabalho nunca tivesse sido idealizado.*

*Ao meu parceiro Felipe, por proporcionar suporte inestimável ao longo desta jornada, me aconselhando e me dando forças para continuar mesmo nos momentos mais desafiadores.*

*À minha irmã Sílvia, por sempre torcer por mim, e ao meu sobrinho Miguel, por me inspirar a ser uma pessoa digna de sua admiração.*

*Ao meu irmão Daniel, por ter me ensinado tantas coisas importantes sobre a vida.*

*À minha companheira de todas as horas Nala, que nunca deixa o meu lado e oferece um suporte emocional que só um bichinho de quatro patas é capaz de oferecer.*

*À Ingrid, minha irmã que nasceu em uma outra família, pelos longos “podcasts” trocados, sempre me dando forças e acreditando no meu potencial.*

*Ao meu amigo Diógenes pelas importantes conversas de bandeirão, sempre ouvindo ativamente e oferecendo valiosos conselhos.*

*Aos meus “irmãos mais velhos” do BioMA, Lucas e Wecksley, por proporcionarem um ambiente acolhedor e me guiarem nos meus primeiros passos no mundo do melhoramento animal e programação.*

*Ao meu “primo” Fernando Bussiman, por ter me auxiliado em momentos chave desde a minha graduação.*

*Aos meus amigos do BioMA-Tech: Alexandre, Aline, Lígia, Lívia e Paula, por contribuírem cada um à sua maneira para o sucesso do nosso grupo.*

*Ao meu orientador e amigo Prof. Dr. Ricardo Ventura, por ter acreditado no meu potencial desde as aulas de estatística em 2016, e mesmo nas fases de maior incerteza e inseguranças. Agradeço também por ter sido um exemplo de caráter e empatia durante toda essa jornada, e por manter um olhar humano e sensível que faz toda a diferença nas vidas daqueles que cruzam o seu caminho.*

*Aos meus coorientadores, Prof. Dr. Luiz Fernando Brito, por todo o apoio desde antes do início do mestrado, e pelas importantes sugestões em pontos cruciais deste trabalho. E Prof. Dr. Marcos Eli Buzanskas, pelas reuniões que foram fundamentais no delineamento deste projeto e pelo suporte ao longo do desenvolvimento deste trabalho.*

*Ao Prof. Dr. Júlio Balieiro, por todo o suporte oferecido desde as aulas de estatística até a estruturação e funcionamento da nossa família BioMA, além de sugestões que elevaram muito a qualidade final deste trabalho.*

*Ao Dr. Gordon Vander Voort, pelas diversas chamadas de vídeo que inspiraram e melhoraram toda a parte de visualização, simulação e análise de dados deste trabalho.*

*Ao Dr. Steve Miller, por confiar a mim a execução das perspicazes ideias desenvolvidas junto ao Prof. Ricardo, e também por compartilhar deste espírito tão humano e empático, capaz de proporcionar grandes mudanças nas vidas por onde passa.*

*A todos que de alguma forma contribuíram para a conclusão desta dissertação, mesmo que indiretamente, e permitiram que eu me tornasse a pessoa e profissional que sou hoje.*

*À Faculdade de Zootecnia e Medicina Veterinária (FZEA/USP), bem como docentes e funcionários, que foram a minha primeira casa em Pirassununga. Em especial à Tânia, que tornou a minha graduação possível através do seu apoio admirável.*

*À Faculdade de Medicina Veterinária e Zootecnia (FMVZ/USP), pela estrutura oferecida e por viabilizar a execução deste mestrado.*

*À Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) pela bolsa (Processo Nº 2021/11156-8) que viabilizou o desenvolvimento deste estudo.*

*À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela concessão da bolsa de mestrado pelo período que antecedeu a concessão da bolsa FAPESP.*

## RESUMO

SELLI, Alana. **Impactos da inclusão de corridas de homozigidade em processos de acasalamento dirigido em uma população simulada de bovinos de corte.**

Dissertação (Mestrado em Ciências) – Faculdade de Medicina Veterinária e Zootecnia, Universidade de São Paulo, Pirassununga, 2024.

Este trabalho teve como objetivo aplicar diferentes estratégias de seleção em relação aos perfis de corridas de homozigidade em uma população (via dados simulados), além de investigar o impacto de tais estratégias sob o aumento ou redução do valor genético estimado (EBV) de uma característica e outros parâmetros relacionados à variabilidade genética da população. Para isso, uma população de bovinos de corte foi simulada com o pacote R AlphaSimR, contendo 5 cromossomos com diferentes tamanhos e números de *loci* de interesse econômico (QTL, do inglês *Quantitative Trait Loci*), correspondentes a uma característica com herdabilidade de 0,3. A partir da simulação inicial, foram testadas três estratégias de seleção e acasalamentos dirigidos. (1) EBV + Fped: valores genéticos estimados através do pedigree e fenótipo, e o coeficiente de endogamia baseado em Pedigree. (2) GEBV + Fg: valores genéticos estimados através de informações de marcadores genéticos do tipo SNP (*Single Nucleotide Polymorphism*), e coeficiente de endogamia genômico. (3) GEBV + Froh: valores genéticos estimados através de informações de marcadores genéticos do tipo SNP, e coeficiente de endogamia baseado em corridas de homozigidade. A estratégia GEBV + Froh produziu consistentemente valores genéticos verdadeiros e fenótipos superiores. O Fg foi mais baixo quando utilizada a estratégia GEBV + Fg, no entanto a mesma estratégia promoveu a formação de corridas de homozigidade maiores e mais frequentes no genoma. A estratégia mais eficiente para minimizar a porcentagem do genoma coberta por ROHs, no entanto, foi a EBV + Fped. As avaliações genéticas de cada animal da população, assim como os respectivos coeficientes de endogamia, foram desenvolvidos via mescla de ativações em tempo real de softwares já existentes (BLUF90, PLINK), por meio de uma interface desenvolvida em linguagem R. Algumas limitações foram pontuadas, tais quais o tamanho populacional utilizado e a adequação de certos parâmetros da simulação, e possíveis soluções foram propostas.

Palavras-chave: Homozigidade. Seleção genômica. Simulação de dados.



## ABSTRACT

SELLI, Alana. **Impacts of the inclusion of runs of homozygosity in the mating strategy processes in a beef cattle simulated population.** Dissertation (Master of Science) – College of Veterinary Medicine and Animal Science, University of Sao Paulo, Pirassununga, 2024.

This work aimed to apply different selection strategies in relation to runs of homozygosity profiles in a population (via simulated data), in addition to investigating the impact of such strategies on the increase or decrease of the Estimated Breeding Value (EBV) of a trait and other parameters related to the genetic variability of the population. For this, a population of beef cattle was simulated with the R package AlphaSimR, containing 5 chromosomes with different sizes and numbers of Quantitative Trait *Loci* (QTL), corresponding to an additive trait with heritability of 0.3. From the initial simulation, three selection and mating strategies were tested. (1) EBV + Fped: genetic values estimated through pedigree and phenotype, and the inbreeding coefficient based on Pedigree. (2) GEBV + Fg: genetic values estimated through information from Single Nucleotide Polymorphism (SNP) markers, and genomic inbreeding coefficient. (3) GEBV + Froh: genetic values estimated through information from SNP genetic markers, and inbreeding coefficient based on homozygosity runs. The GEBV + Froh strategy consistently produced superior true breeding values and phenotypes. Fg was lower when using the GEBV + Fg strategy, however the same strategy promoted the formation of larger and more frequent runs of homozygosity in the genome. The most efficient strategy to minimize the percentage of the genome covered by ROHs, however, was EBV + Fped. The genetic evaluations of each animal in the population, as well as the respective inbreeding coefficients, were developed via a mix of real-time activations of existing software (BLUF90, PLINK), through an interface developed in R language. Some limitations were highlighted, such as the population size used and the adequacy of certain simulation parameters, and possible solutions were proposed.

Keywords: Homozygosity. Genomic selection. Data simulation.

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b> .....	<b>11</b>
1.1	HIPÓTESE .....	14
1.2	OBJETIVOS .....	14
1.3	REFERÊNCIAS.....	15
<b>2</b>	<b>REVISÃO DE LITERATURA</b> .....	<b>17</b>
2.1	DESEQUILÍBRIO DE LIGAÇÃO .....	17
2.2	CORRIDA DE HOMOZIGOSIDADE E SUA RELAÇÃO COM A ENDOGAMIA....	19
<b>2.2.1</b>	<b>Coeficiente de endogamia</b> .....	<b>19</b>
<b>2.2.2</b>	<b>Corridas de homozigosidade</b> .....	<b>20</b>
2.3	SELEÇÃO GENÉTICA.....	23
<b>2.3.1</b>	<b>Tradicional</b> .....	<b>23</b>
<b>2.3.2</b>	<b>Genômica</b> .....	<b>24</b>
2.4	ACASALAMENTOS DIRIGIDOS .....	26
2.5	FERRAMENTAS COMPUTACIONAIS .....	27
<b>2.5.1</b>	<b>Simulação de dados</b> .....	<b>27</b>
<b>2.5.2</b>	<b>Linguagens de programação</b> .....	<b>28</b>
<b>2.5.3</b>	<b>Visualização de dados</b> .....	<b>30</b>
2.6	REFERÊNCIAS.....	32
<b>3</b>	<b>INTEGRAÇÃO DE SIMULAÇÃO DE DADOS GENÔMICOS E FERRAMENTAS DE ANÁLISES GENÔMICAS EM BOVINOS DE CORTE: UM PIPELINE PARA ESTUDOS DE SELEÇÃO GENÉTICA E ACASALAMENTOS DIRIGIDOS COM FOCO NA MENSURAÇÃO DE CORRIDAS DE HOMOZIGOSIDADE.</b> .....	<b>42</b>
3.1	RESUMO.....	42
3.2	INTRODUÇÃO .....	43
3.3	MATERIAL E MÉTODOS.....	45
<b>3.3.1</b>	<b>Simulação inicial de dados</b> .....	<b>46</b>
<b>3.3.2</b>	<b>Progênes Hipotéticas com Ausência de Segregação Mendeliana (PHASM)..</b>	<b>49</b>
<b>3.3.3</b>	<b>Avaliações genômicas</b> .....	<b>50</b>
<b>3.3.4</b>	<b>Corridas de homozigosidade</b> .....	<b>51</b>
<b>3.3.5</b>	<b>Pipeline de análises e seleção</b> .....	<b>52</b>
<b>3.3.6</b>	<b>Validação da simulação inicial</b> .....	<b>55</b>

<b>3.3.7</b>	<b>Avaliação dos cenários .....</b>	<b>56</b>
3.4	RESULTADOS E DISCUSSÃO .....	57
<b>3.4.1</b>	<b>Simulação e validação de dados .....</b>	<b>57</b>
<b>3.4.2</b>	<b>Perfil de corridas de homozigosidade .....</b>	<b>61</b>
<b>3.4.3</b>	<b>Comparação dos cenários de seleção e direcionamento de acasalamentos .</b>	<b>68</b>
<b>3.4.4</b>	<b>Limitações e sugestões para estudos futuros .....</b>	<b>81</b>
3.5	CONCLUSÃO.....	83
3.6	REFERÊNCIAS.....	84
<b>4</b>	<b>O USO DE VISUALIZAÇÕES INTERATIVAS PARA RASTREAR</b>	
	<b>HERANÇA HAPLOTÍPICA NA PECUÁRIA .....</b>	<b>87</b>
4.1	RESUMO.....	87
4.2	INTRODUÇÃO .....	87
4.3	MATERIAL E MÉTODOS.....	91
<b>4.3.1</b>	<b>Simulação e Validação dos dados .....</b>	<b>91</b>
<b>4.3.2</b>	<b>Regressão de SNP Único e Preparação de Dados .....</b>	<b>93</b>
<b>4.3.3</b>	<b>Ferramenta de Visualização .....</b>	<b>94</b>
<b>4.3.4</b>	<b>Cenários para ilustrar a ferramenta de visualização.....</b>	<b>96</b>
4.3.4.1	Comparação de trios .....	96
4.3.4.2	Visualização de avós.....	96
4.4	RESULTADOS E DISCUSSÃO .....	96
<b>4.4.1</b>	<b>Simulação e Validação de Dados .....</b>	<b>96</b>
<b>4.4.2</b>	<b>Visualização de Trios .....</b>	<b>98</b>
<b>4.4.3</b>	<b>Visualização dos avós .....</b>	<b>101</b>
4.5	CONCLUSÃO.....	105
4.6	REFERÊNCIAS.....	106
<b>5</b>	<b>CONSIDERAÇÕES FINAIS .....</b>	<b>110</b>

## 1 1 INTRODUÇÃO

2 Segundo dados da Associação Brasileira das Indústrias Exportadoras de Carne  
3 (ABIEC, 2022), a cadeia produtiva da pecuária de corte no Brasil movimentou cerca de R\$  
4 913,14 bilhões no ano de 2021, o que representa quase 10% do produto interno bruto do  
5 país no mesmo ano (AGÊNCIA GOV, 2023). Em 2022, o rebanho bovino brasileiro alcançou  
6 a marca de 234,4 milhões de animais, incluindo rebanhos de corte e leiteiros (MAPA, 2023).  
7 Em relação ao volume de exportações de carne bovina, dados do departamento de  
8 agricultura dos Estados Unidos apontam que o Brasil liderou as exportações mundiais em  
9 2023 (USDA, 2024). Estes indicadores evidenciam o papel fundamental que a pecuária de  
10 corte desempenha na geração de empregos, garantia da segurança alimentar, e na balança  
11 comercial brasileira. No entanto, para manter-se competitivo o setor deve continuar  
12 investindo em produtividade, sustentabilidade ambiental e adaptação às mudanças  
13 climáticas.

14 Uma das principais ferramentas capazes de impulsionar o aumento da  
15 produtividade, a qualidade dos produtos e a adaptação ao meio ambiente dos rebanhos  
16 bovinos, diz respeito ao melhoramento genético. Inicialmente importador de animais zebu,  
17 após diversos avanços através da pesquisa e do desenvolvimento de programas nacionais  
18 de melhoramento genético, o Brasil é hoje exportador de animais rústicos, resistentes a  
19 doenças e parasitas, e que ao mesmo tempo exibem alto desempenho e qualidade de carne  
20 (EMBRAPA, 2024). Com o aumento do uso da genômica a partir de 2009 (GODDARD;  
21 HAYES, 2009; MISZTAL; LEGARRA; AGUILAR, 2009; VANRADEN et al., 2009), o ganho  
22 genético dos rebanhos passou a ser mais expressivo, através do aumento da acurácia de  
23 predição dos valores genéticos e redução do intervalo de gerações (PRYCE; HAYES;  
24 GODDARD, 2012).

25 No entanto, um dos desafios do melhoramento genético é a manutenção do nível  
26 de endogamia da população, que tende a aumentar com o acasalamento de animais  
27 aparentados, que geralmente são selecionados durante a escolha de indivíduos  
28 geneticamente superiores (CLARK et al., 2013; DAETWYLER et al., 2007). A endogamia,  
29 por sua vez, pode apresentar efeitos positivos, como a fixação de alelos favoráveis e o  
30 aumento da homogeneidade da progênie, bem como efeitos negativos, a exemplo do

31 aumento do aparecimento de doenças recessivas na população e da piora da performance  
32 ou vigor dos indivíduos (BOURDON, 2014; DICKERSON, 1963).

33 Tradicionalmente, os programas de melhoramento genético fazem uso do  
34 coeficiente de endogamia calculado através do pedigree para realizar a seleção de quais  
35 animais serão acasalados. Entretanto, uma grande desvantagem deste método é que os  
36 coeficientes calculados não capturam os efeitos da segregação mendeliana. Por exemplo,  
37 estima-se que irmãos completos compartilham em média 50% de conteúdo genético, porém  
38 na prática este valor pode apresentar grande variação. Este método também está sujeito a  
39 limitações como o número de gerações que se tem registro, além de possíveis erros no  
40 pedigree.

41 Uma outra forma de se quantificar a endogamia é através da proporção do genoma  
42 de um indivíduo que é coberta por corridas de homoziguidade (do inglês Runs of  
43 Homozygosity – ROH), que correspondem a cópias idênticas de haplótipos que são  
44 herdadas de um mesmo indivíduo (CEBALLOS et al., 2018). Tais trechos homozigotos  
45 podem ser identificados por meio de análises de marcadores moleculares do tipo SNP (do  
46 inglês, Single Nucleotide Polymorphism). Uma das vantagens deste método corresponde à  
47 possibilidade de identificar em quais regiões do genoma os indivíduos apresentam  
48 homoziguidade, e assim verificar a probabilidade de sua progênie também ser homozigota  
49 em tais regiões, a depender de ambos os progenitores, além de levar em conta quais  
50 características são afetadas por tais regiões. Assim, é possível delinear diferentes  
51 estratégias de melhoramento animal que utilizam tais informações.

52 Por isso, é de extrema importância entender como os mecanismos biológicos que  
53 permeiam a seleção genômica funcionam, além de antecipar os impactos de longo prazo  
54 das estratégias adotadas antes da sua implementação. Para este fim, a elaboração de  
55 simulações computacionais é fundamental, além da simplificação de sistemas altamente  
56 complexos. Programas como QMSim (SARGOLZAEI; SCHENKEL, 2009) e AlphaSimR  
57 (GAYNOR; GORJANC; HICKEY, 2021) permitem a criação de dados genômicos  
58 complexos e que refletem em grande parte a realidade, possibilitando um bom  
59 entendimento acerca do impacto de diversos processos evolutivos e demográficos sobre a

60 variação da sequência gênica, assim como a predição dos resultados de diferentes  
61 métodos de seleção (YUAN et al., 2012).

62 Uma das grandes vantagens do simulador AlphaSimR, diferente de outros  
63 simuladores, é que a sua implementação se dá através da linguagem de programação R.  
64 Isso permite a execução da simulação por meio de scripts ou de forma interativa, além de  
65 possibilitar o uso de recursos de programação como loops e o armazenamento de arquivos  
66 em objetos. Além disso, tal recurso facilita a manipulação dos dados, e o uso de  
67 ferramentas adicionais, também disponíveis via pacotes R. O uso de um simulador ativado  
68 por meio de scripts pertencentes a uma linguagem de programação, também permite a  
69 execução de comandos Linux, de modo que softwares externos sejam incorporados ao  
70 mesmo fluxo de processos. A integração destas ferramentas pode parecer trivial, no entanto  
71 ela permite que esquemas mais complexos e parecidos com o que é desenvolvido na  
72 indústria, por programas existentes de melhoramento genético, sejam executados e  
73 testados. Em contraste, outros simuladores não permitem que a simulação seja retomada  
74 após finalizado o processo, e uma vez que os arquivos da simulação são obtidos e  
75 analisados através de programas externos, não é possível criar novas gerações a partir da  
76 simulação desenvolvida.

77 Após a obtenção dos dados através da simulação e da execução de importantes  
78 análises genômicas ou estatísticas, é imprescindível a comunicação de tais resultados para  
79 o público, seja ele acadêmico, da indústria ou os próprios produtores (ROUGIER;  
80 DROETTBOOM; BOURNE, 2014). A forma como os resultados são apresentados têm um  
81 forte impacto na capacidade de entendimento acerca do trabalho desenvolvido, e de nada  
82 adianta o desenvolvimento de um estudo rigoroso e inovador se o mesmo não for  
83 compartilhado e compreendido. Uma das formas mais eficientes e convidativas de  
84 apresentar-se resultados se dá por meio do uso de gráficos e figuras, conforme muito bem  
85 descrito por Durant et al. (DURANT et al., 2022), ao lembrar como a frase "vamos achatar  
86 a curva!" tornou-se famosa durante a pandemia de COVID-19, em referência aos gráficos  
87 de transmissão do vírus.

88 Este trabalho buscou, via uso de simulação de dados e aplicação de ferramentas  
89 computacionais focadas em programação e visualização de dados, integrar diversos

90 conceitos de melhoramento genético e análises genômicas, desenvolvendo-se por fim, uma  
91 etapa final de apresentação dos resultados de forma dinâmica e de fácil acesso e  
92 entendimento do usuário. A proposta do presente estudo não foi a de desenvolver novas  
93 metodologias de seleção genômica, nem propor os melhores parâmetros para a simulação  
94 de dados ou análises genômicas, mas sim oferecer uma base bem estruturada onde se  
95 estabelece a integração entre diversas ferramentas já usadas por estudos do  
96 melhoramento animal, que por si só apresentam um desafio para novos estudantes da área.

## 97 1.1 HIPÓTESE

98 1. A inclusão da análise dos perfis de corridas de homozigidade na seleção  
99 genômica permite a análise pormenorizada dos efeitos da endogamia na população, assim  
100 como seus possíveis impactos no valor genético estimado de uma característica,  
101 comparado ao uso de valores únicos de coeficiente de endogamia obtidos por análise do  
102 pedigree.

103 2. Verificar a possibilidade da elaboração de uma interface (via Power BI, Tableau  
104 ou Dash) que possibilite ao usuário interagir com diferentes momentos da simulação e  
105 favoreça o entendimento sobre como os diferentes componentes do sistema (fenótipos,  
106 genótipos, desequilíbrio de ligação, entre outros) são afetados ao longo do processo de  
107 seleção.

## 108 1.2 OBJETIVOS

109 1. Aplicar diferentes estratégias de seleção em relação aos perfis de corridas de  
110 homozigidade em uma população (via dados simulados), além de investigar o impacto  
111 de tais estratégias sob o valor genético estimado de uma característica e outros parâmetros  
112 relacionados à variabilidade genética da população.

113 2. Buscar perfis de corridas de homozigidade que permitam a execução de  
114 acasalamentos dirigidos em um processo refinado de seleção.

115 3. Desenvolver uma interface que permita o acompanhamento de tal processo de  
116 seleção e a observação de seus impactos sobre cada componente do sistema.

### 117 1.3 REFERÊNCIAS

118 ABIEC. ASSOCIAÇÃO BRASILEIRA DAS INDÚSTRIAS EXPORTADORAS DE CARNE.  
119 **BEEF REPORT 2022**. DISPONÍVEL EM: <<https://www.abiec.com.br/en/publicacoes/beef-report-2022-2/>>. Acesso em: 07 mar. 2024.

121  
122 AGÊNCIA GOV. **IBGE revisa resultado do PIB de 2021 de 5,0% para 4,8%**. Publicado  
123 em: 08/11/2023. DISPONÍVEL EM: <<https://agenciagov.ebc.com.br/noticias/202311/pib-cresce-4-8-e-totaliza-r-9-trilhoes-em-2021#:~:text=O%20valor%20adicionado%20dos%20serviços,de%20Contas%20Nacionais%20do%20IBGE>>. Acesso em: 07 mar. 2024.

127  
128 BOURDON, R. M. **Understanding Animal Breeding**. Harlow: Pearson, 2014.

129  
130 BRASIL. Ministério da Agricultura e Pecuária. Produção da pecuária municipal: rebanho  
131 bovino brasileiro alcançou recorde de 234,4 milhões de animais em 2022. Publicado em:  
132 25/09/2023. Atualizado em: 25/09/2023. Disponível em: <https://www.gov.br/agricultura/pt-br/assuntos/noticias/rebanho-bovino-brasileiro-alcancou-recorde-de-234-4-milhoes-de-animais-em-2022#>. Acesso em: **07 mar.** 2024.

135  
136 CEBALLOS, F. C.; JOSHI, P. K.; CLARK, D. W.; RAMSAY, M.; WILSON, J. F. Runs of  
137 homozygosity: Windows into population history and trait architecture. **Nature Reviews**  
138 **Genetics**, v. 19, n. 4, p. 220–234, 2018.

139  
140 CLARK, S. A.; KINGHORN, B. P.; HICKEY, J. M.; VAN DER WERF, J. H. J. The effect of  
141 genomic information on optimal contribution selection in livestock breeding programs.  
142 **Genetics Selection Evolution**, v. 45, n. 1, p. 1–8, 2013.

143  
144 DAETWYLER, H. D.; VILLANUEVA, B.; BIJMA, P.; WOOLLIAMS, J. A. Inbreeding in  
145 genome-wide selection. **Journal of Animal Breeding and Genetics**, v. 124, n. 6, p. 369–  
146 376, 2007.

147  
148 DICKERSON, G. E. Experimental evaluation of selection theory in poultry. In:  
149 INTERNATIONAL CONGRESS OF GENETICS, 11., 1963, The Hague. **Proceedings...**  
150 The Hague: Pergmon, 1963. p. 747–761. (Genetics Today, v. 1)

151  
152 DURANT, E.; ROUARD, M.; GANKO, E. W.; MULLER, C.; CLEARY, A. M.; FARMER, A.  
153 D.; CONTE, M.; SABOT, F. Ten Simple Rules for Developing Visualization Tools in  
154 Genomics. **PLOS Computational Biology**, v. 18, n. 11, p. e1010622, 10 nov. 2022.

155  
156 EMBRAPA. Empresa Brasileira de Pesquisa Agropecuária. **Melhoramento genético**.  
157 Disponível em: <https://www.embrapa.br/qualidade-da-carne/carne-bovina/producao-de-carne-bovina/melhoramento-genetico>. Acesso em: 08 mar. 2024.



- 159  
160 GAYNOR, R. C.; GORJANC, G.; HICKEY, J. M. AlphaSimR: an R package for breeding  
161 program simulations. **G3 Genes|Genomes|Genetics**, v. 11, n. 2, 12 abr. 2021. Disponível  
162 em: <<https://academic.oup.com/g3journal/article/doi/10.1093/g3journal/jkaa017/6025179>>.  
163 Acesso em: 31 mar. 2022.
- 164  
165 GODDARD, M. E.; HAYES, B. J. Mapping genes for complex traits in domestic animals  
166 and their use in breeding programmes. **Nature Reviews Genetics**, v. 10, n. 6, p. 381–  
167 391, jun. 2009.
- 168  
169 MAPA. Ministério da Agricultura, Pecuária e Abastecimento. **Rebanho bovino brasileiro**  
170 **alcançou recorde de 234,4 milhões de animais em 2022**. Disponível em:  
171 [https://www.gov.br/agricultura/pt-br/assuntos/noticias/rebanho-bovino-brasileiro-alcancou-](https://www.gov.br/agricultura/pt-br/assuntos/noticias/rebanho-bovino-brasileiro-alcancou-recorde-de-234-4-milhoes-de-animais-em-2022#)  
172 [recorde-de-234-4-milhoes-de-animais-em-2022#](https://www.gov.br/agricultura/pt-br/assuntos/noticias/rebanho-bovino-brasileiro-alcancou-recorde-de-234-4-milhoes-de-animais-em-2022#). Acesso em: 08 mar. 2024.
- 173  
174 MISZTAL, I.; LEGARRA, A.; AGUILAR, I. Computing procedures for genetic evaluation  
175 including phenotypic, full pedigree, and genomic information. **Journal of Dairy Science**, v.  
176 92, n. 9, p. 4648–4655, 2009.
- 177  
178 PRYCE, J. E.; HAYES, B. J.; GODDARD, M. E. Novel strategies to minimize progeny  
179 inbreeding while maximizing genetic gain using genomic information. **Journal of Dairy**  
180 **Science**, v. 95, n. 1, p. 377–388, 2012.
- 181  
182 ROUGIER, N. P.; DROETTBOOM, M.; BOURNE, P. E. Ten Simple Rules for Better  
183 Figures. **PLoS Computational Biology**, v. 10, n. 9, p. e1003833, 11 set. 2014.
- 184  
185 SARGOLZAEI, M.; SCHENKEL, F. S. QMSim: A large-scale genome simulator for  
186 livestock. **Bioinformatics**, v. 25, n. 5, p. 680–681, 2009.
- 187  
188 USDA. Departamento de Agricultura dos Estados Unidos. **Livestock & Poultry: World**  
189 **Markets and Trade**. Disponível em: [https://downloads.usda.library.cornell.edu/usda-](https://downloads.usda.library.cornell.edu/usda-esmis/files/73666448x/cn69nh059/zs25zp436/livestock_poultry.pdf)  
190 [esmis/files/73666448x/cn69nh059/zs25zp436/livestock\\_poultry.pdf](https://downloads.usda.library.cornell.edu/usda-esmis/files/73666448x/cn69nh059/zs25zp436/livestock_poultry.pdf). Acesso em: 08 mar.  
191 2024.
- 192  
193 VANRADEN, P. M.; VAN TASSELL, C. P.; WIGGANS, G. R.; SONSTEGARD, T.  
194 S.; SCHNABEL, R. D.; TAYLOR, J. F.; SCHENKEL, F. S. Invited review: Reliability of  
195 genomic predictions for North American Holstein bulls. **Journal of Dairy Science**, v. 92, n.  
196 1, p. 16–24, 2009.
- 197  
198 YUAN, X.; MILLER, D. J.; ZHANG, J.; HERRINGTON, D.; WANG, Y. An overview of  
199 population genetic data simulation. **Journal of Computational Biology**, v. 19, n. 1, p. 42–  
200 54, 2012.
- 201

## 202 2 REVISÃO DE LITERATURA

### 203 2.1 DESEQUILÍBRIO DE LIGAÇÃO

204 O Desequilíbrio de Ligação, do inglês Linkage Disequilibrium (LD), corresponde à  
205 associação não randômica entre alelos em dois ou mais *loci* (MEUWISSEN; GODDARD,  
206 2000). Tal fenômeno se deve ao fato de que os alelos não são passados para a progênie  
207 de forma independente, mas sim em blocos - também chamados de haplótipos (ARDLIE;  
208 KRUGLYAK; SEIELSTAD, 2002). Quanto menor a distância física entre os alelos, maior a  
209 probabilidade de estes serem herdados em conjunto, uma vez que há uma menor chance  
210 de ocorrer recombinação ou Crossing over - processo celular durante a meiose que  
211 ocasiona na troca de partes entre os cromossomos paterno e materno (WANG et al., 2015).

212 O LD é um fenômeno populacional, ou seja, não pode ser medido em apenas um  
213 animal, pois leva em conta as frequências alélicas observadas em um conjunto de  
214 indivíduos pertencentes a uma mesma população. Os níveis de LD resultam de uma série  
215 de processos ocorridos ao longo do tempo. Segundo Ardlie, Kruglyak e Seielstad (2002),  
216 as taxas de mutação e recombinação gênicas são os fatores de efeito mais evidentes, uma  
217 vez que ocasionam na quebra de grandes blocos de alelos ao longo das gerações (REICH  
218 et al., 2001), no entanto aspectos demográficos da população também afetam a extensão  
219 e distribuição do LD no genoma, conforme descrito a seguir.

220 A deriva genética é um processo que ocorre geralmente em populações pequenas,  
221 de baixo tamanho efetivo ( $N_e$ ), e corresponde a alterações nas frequências gênicas e  
222 haplotípicas a cada geração devido ao processo aleatório de segregação mendeliana. Este  
223 fenômeno tende a causar um aumento do LD, uma vez que haplótipos são perdidos da  
224 população (OLDENBROEK; VAN DER WAAIJ, 2014). Por outro lado, o rápido crescimento  
225 populacional – e aumento do  $N_e$  – tende a reduzir a deriva genética e o LD (ARDLIE;  
226 KRUGLYAK; SEIELSTAD, 2002). Desta forma, alterações no tamanho efetivo da  
227 população ao longo de sua história, tais quais expansões ou reduções, também chamadas  
228 de eventos bottleneck, alteram os níveis de LD na população (REICH et al., 2001).

229           Eventos de Admixture (cruzamentos entre duas ou mais populações previamente  
 230 isoladas) ou migração (fluxo de gametas ou indivíduos para uma determinada população),  
 231 fazem com que o LD a princípio seja proporcional às diferenças entre as frequências  
 232 alélicas das populações misturadas, de forma independente das distâncias entre  
 233 marcadores. Porém este LD rapidamente se dissipa, enquanto que o LD entre alelos  
 234 próximos persiste por mais gerações (ARDLIE; KRUGLYAK; SEIELSTAD, 2002). A seleção  
 235 natural ou artificial também altera os níveis de LD entre certos alelos, aumentando as  
 236 frequências alélicas em regiões vizinhas aos locais alvo da seleção - efeito conhecido como  
 237 hitchhiking - podendo levar à fixação de tais alelos, ou através da redução da frequência de  
 238 alelos deletérios (ARDLIE; KRUGLYAK; SEIELSTAD, 2002; TEO et al., 2009).

239           Conhecer os níveis de LD de uma população apresenta diversas aplicações, desde  
 240 inferir eventos demográficos com base em estimativas de tamanhos populacionais efetivos  
 241 ao longo das gerações (FREEMAN et al., 2006; MCEVOY et al., 2011; REICH et al., 2001)  
 242 e estimar taxas de recombinação (LI; STEPHENS, 2003), até para fins de seleção genômica  
 243 (MEUWISSEN; HAYES; GODDARD, 2001) e estimação de efeitos de marcadores  
 244 moleculares por meio de estudos de ampla associação genômica (do inglês Genome-Wide  
 245 Association Studies - GWAS) (ARDLIE; KRUGLYAK; SEIELSTAD, 2002). O nível de LD  
 246 influencia diretamente no número de marcadores necessários nestes últimos  
 247 procedimentos, uma vez que quanto mais fraca a associação entre marcadores, mais  
 248 próximos estes devem estar dos locais de interesse e maior deverá ser a densidade do  
 249 painel (PÉREZ O'BRIEN et al., 2014).

250           Diferentes métricas podem ser utilizadas para descrever os níveis de LD. Conforme  
 251 revisado por Lu et al. (LU et al., 2012), a métrica  $D$  corresponde à diferença entre as  
 252 frequências haplotípicas observadas e esperadas:

$$253 \qquad D = f(AB) - f(A)f(B) ,$$

254           Onde  $f(AB)$  é a frequência estimada do haplótipo AB utilizando a frequência  
 255 genotípica observada e assumindo equilíbrio de Hardy-Weinberg, enquanto que  $f(A)$  e  $f(B)$   
 256 são as frequências dos alelos A e B, respectivamente. Devido a alta dependência da  
 257 métrica  $D$  às frequências alélicas, esta não é utilizada para comparar LD entre múltiplos

258 pares de loci (LU et al., 2012). Outra métrica, introduzida por Lewontin (1964), é a  $D'$ , cujo  
 259 valor é determinado dividindo  $D$  por seu maior valor possível, dado a as frequências alélicas  
 260 nos dois loci. Assim, quando  $D'$  é igual a 1, os dois loci são completamente correlacionados.  
 261 No entanto, valores de  $D' < 1$  não apresentam interpretação clara, devido a sua magnitude  
 262 relativa. Além disso, as estimativas de  $D'$  são fortemente infladas em amostras pequenas,  
 263 especialmente para SNPs raros (ARDLIE; KRUGLYAK; SEIELSTAD, 2002).

264 Hill e Roberson (1968) desenvolveram uma nova métrica,  $r^2$ , que calcula-se como:

$$265 \quad r^2 = \frac{D^2}{f(A)f(a)f(B)f(b)}$$

266

267 Onde  $f(A)$ ,  $f(a)$ ,  $f(B)$  e  $f(b)$  são as frequências observadas dos alelos A, a, B e b,  
 268 respectivamente. O  $r^2$  varia de 0 a 1, onde 0 significa que os alelos são completamente não  
 269 correlacionados e 1 significa que os alelos são completamente correlacionados. Esta  
 270 métrica apresenta a vantagem de ser menos dependente das frequências alélicas  
 271 (BOHMANOVA; SARGOLZAEI; SCHENKEL, 2010; ZHAO; NETTLETON; DEKKERS,  
 272 2007) e mais adequada para a estimativa de LD entre marcadores bialélicos, como os SNPs  
 273 (BRITO et al., 2011).

## 274 2.2 CORRIDA DE HOMOZIGOSIDADE E SUA RELAÇÃO COM A ENDOGAMIA

### 275 2.2.1 Coeficiente de endogamia

276 Endogamia é o termo utilizado para descrever o acasalamento de indivíduos  
 277 relacionados entre si por ancestralidade (FALCONER; MACKAY, 1996). A principal  
 278 consequência da endogamia é que indivíduos aparentados podem carregar a mesma cópia  
 279 alélica de seu ancestral em comum, e quando estes são acasalados, ambas as cópias  
 280 podem ser passadas para sua progênie, gerando o que chamamos de homozigidade  
 281 (FALCONER & MACKAY, 1996).

282 Segundo Wright (WRIGHT, 1922), existem duas classes de efeito para o fenômeno  
 283 da endogamia: a depressão endogâmica e a uniformidade da progênie. A primeira se refere  
 284 ao decaimento de vigor do indivíduo, com efeitos no peso, fertilidade, saúde, entre outros.

285 Este evento pode ser justificado pelo fato de que alelos desfavoráveis são frequentemente  
286 recessivos, uma vez que a seleção natural tende a excluir alelos de efeitos negativos da  
287 população, e quando estes são recessivos há uma maior chance de 'sobreviverem' na  
288 população em um estado heterozigoto. Já a uniformidade da progênie é diretamente  
289 proporcional à porcentagem de homozigosidade, e por conseguinte leva à fixação de alelos  
290 e ao aumento da prepotência (a capacidade de um indivíduo de passar suas características  
291 para a progênie). O autor também afirma que os efeitos descritos são observados  
292 experimentalmente como consequências médias da endogamia, porém não são  
293 necessariamente inevitáveis.

294 O objetivo da seleção genética é aumentar a frequência de certos alelos na  
295 população, o que ocorre por meio da escolha de um número limitado de indivíduos para  
296 serem progenitores da próxima geração – aqueles que apresentam os alelos desejados.  
297 Assim, há uma maior chance de que os reprodutores sejam aparentados (ROBERTSON,  
298 1961), fazendo com que a endogamia ao longo das gerações seja inevitável (LEROY,  
299 2014). No entanto, o acasalamento de indivíduos semelhantes entre si pode reduzir a  
300 variabilidade genética da população, o que ao longo das gerações leva a uma redução da  
301 resposta à seleção para a característica sob seleção e as demais características  
302 (KRISTENSEN; SØRENSEN, 2005). Assim, faz-se necessário a caracterização e o controle  
303 eficiente da endogamia, de modo a garantir a possibilidade de adaptação dos indivíduos  
304 para objetivos de seleção do futuro, mantendo a diversidade genética e evitando o acúmulo  
305 de efeitos prejudiciais associados à endogamia (HOWARD et al., 2017a).

## 306 **2.2.2 Corridas de homozigosidade**

307 Com o surgimento e a popularização dos painéis de genotipagem que fazem uso  
308 de marcadores moleculares do tipo SNP, é possível quantificar o nível de endogamia de  
309 um indivíduo fazendo-se uso de informações genômicas, sem a dependência de registros  
310 de pedigree. Uma das formas de utilizar tais informações consiste em calcular a matriz  
311 genômica de relacionamentos (do inglês Genomic Relationship Matrix - GRM) entre  
312 indivíduos de um grupo (VANRADEN, 2008). Diferente do método baseado em pedigree,  
313 estimativas genômicas de endogamia não são limitadas pela profundidade do pedigree  
314 (número de gerações que se tem registro), nem por erros de registro, além de captar

315 oscilações entre animais com o mesmo pedigree, como por exemplo irmãos completos  
316 (PRYCE et al., 2014).

317 Além do método que faz o uso da GRM, também é possível calcular a proporção  
318 de SNPs que são homozigotos para estimar o nível de endogamia de um indivíduo. No  
319 entanto, ambos os métodos não diferenciam quais homozigotos são IBD ou IBS (PRYCE  
320 et al., 2014). Uma alternativa a tal limitação pode ser o uso do conceito de Corridas de  
321 Homozigosidade (do inglês Runs of Homozygosity, ROH), que corresponde a regiões no  
322 genoma de um indivíduo que apresentam haplótipos homozigotos herdados de um mesmo  
323 ancestral (BROMAN; WEBER, 1999). Tal conceito surge do entendimento de que  
324 homozigotos que são IBD (também chamados de autozigotos) não se encontram  
325 distribuídos uniformemente pelo genoma, mas sim em segmentos, ou corridas  
326 (MCQUILLAN et al., 2008).

327 O comprimento das ROH pode ser utilizado para inferir o número de gerações  
328 passadas em que ocorreu a endogamia (BJELLAND et al., 2013), visto que as ROH são  
329 mais curtas quando o ancestral em comum que originou a endogamia surgiu há muitas  
330 gerações atrás, e mais longas quando a endogamia ocorreu mais recentemente  
331 (PURFIELD et al., 2012). Isto se deve ao fato de que a cada geração, eventos de  
332 recombinação gênica podem acontecer (MCQUILLAN et al., 2008). Assim, a identificação  
333 e caracterização do perfil das ROH pode revelar informações sobre o histórico, estrutura e  
334 demografia de uma população, uma vez que a formação das ROH sofre influência de  
335 processos como deriva genética, bottlenecks e intensa seleção (PERIPOLLI et al., 2017),  
336 que levam a um aumento da endogamia na população (FALCONER; MACKAY, 1996).

337 O estudo das ROH também pode revelar assinaturas de seleção, caracterizadas  
338 pela fixação de alelos sob pressão de seleção em uma população (MUIR et al., 2008;  
339 WALSH, 2021) e subsequente aumento de homozigosidade em regiões vizinhas (STELLA  
340 et al., 2010; URBINATI et al., 2016). Quando muitos indivíduos em uma população  
341 apresentam uma mesma região genômica em ROH, tal região é chamada de ilha de ROH  
342 (PERIPOLLI et al., 2018). Um dos fatores de formação de tais ilhas pode ser a presença de  
343 alto LD na região, conforme já observado por Purfield et al., (PURFIELD et al., 2012), Bosse  
344 et al. (BOSSE et al., 2012), Herrero-Medrano et al. (HERRERO-MEDRANO et al., 2013),

345 entre outros. No entanto, a relação de causa e efeito entre LD e ROH ainda não foi  
346 completamente elucidada, uma vez que ambos os fenômenos associam-se com o aumento  
347 de endogamia da região, porém não ocorrem sempre em conjunto (AL-MAMUN et al., 2015;  
348 SELLI et al., 2021).

349 A detecção das ROH pode ser feita por meio de diferentes Softwares (PERIPOLLI  
350 et al., 2017), métodos (SANTOS et al., 2021), e adotando-se diferentes parâmetros  
351 (MEYERMANS et al., 2020). O Software PLINK (PURCELL et al., 2007) e o pacote R  
352 DetectRuns (BISCARINI et al., 2018) estão entre os programas mais utilizados atualmente  
353 (ABIED et al., 2020; CEBALLOS; HAZELHURST; RAMSAY, 2018; DZOMBA et al., 2021;  
354 SELLI et al., 2021). O primeiro faz o uso de um método baseado em janelas, que percorrem  
355 os cromossomos de um indivíduo SNP por SNP, marcando a proporção de vezes em que  
356 cada marcador aparece em uma janela de homozigosidade. Este método é melhor descrito  
357 por Bjelland et al. (2013). Já o pacote DetectRuns apresenta, de maneira análoga ao PLINK,  
358 um método baseado em janelas (sliding windows), e adicionalmente um método de 'corridas  
359 consecutivas' (consecutive runs). Este segundo método, inicialmente proposto por  
360 (MARRAS et al., 2015), não faz o uso de janelas, e analisa os marcadores de forma  
361 independente.

362 Para cada um dos métodos propostos, existe uma série de parâmetros que devem  
363 ser definidos, como: número mínimo de marcadores por corrida, intervalo máximo entre  
364 marcadores em uma corrida, número máximo de marcadores heterozigotos ou perdidos,  
365 entre outros, além de parâmetros específicos dos métodos baseados em janela. Peripolli et  
366 al. (2017) revisou vários estudos que buscaram a detecção de ROH em indivíduos de  
367 diferentes espécies, e reportou uma grande variação entre os parâmetros adotados, o que  
368 por sua vez pode apresentar impacto sobre o número e o tamanho das ROH identificadas  
369 (HOWRIGAN; SIMONSON; KELLER, 2011) e sobre estimativas de FROH  
370 (MASTRANGELO et al., 2016). Além disso, a densidade do painel de genotipagem utilizado  
371 apresenta interação com os parâmetros escolhidos (FERENČAKOVIĆ; SÖLKNER; CURIK,  
372 2013). Painéis de 50 mil marcadores tendem a superestimar o número de corridas curtas,  
373 de até 5 Mb (FERENČAKOVIĆ; SÖLKNER; CURIK, 2013; PURFIELD et al., 2012),  
374 enquanto que painéis HD, com mais de 700 mil marcadores, tendem a subestimar o número  
375 de corridas com mais de 8 Mb (FERENČAKOVIĆ; SÖLKNER; CURIK, 2013).

## 376 2.3 SELEÇÃO GENÉTICA

377 Ao falar com qualquer pessoa sobre seleção genética, a primeira ideia que vem à  
378 cabeça é provavelmente o conceito de seleção natural, "a sobrevivência daquele que  
379 melhor se adapta", descrita por Charles Darwin em seu livro "On the Origin of Species" (A  
380 Origem das Espécies) de 1859. No entanto, é possível verificar evidências de que os seres  
381 humanos já praticavam o ato de selecionar os indivíduos para gerar a próxima geração  
382 muito antes, mesmo sem conhecer os princípios da genética (OLDENBROEK; VAN DER  
383 WAAIJ, 2014).

384 Até onde sabe-se, na Europa até o século XVIII, não existia uma maneira  
385 sistemática de selecionar indivíduos. Sir Robert Bakewell foi o primeiro a introduzir o  
386 registro preciso de desempenho de animais, permitindo assim que a seleção objetiva se  
387 tornasse possível, além de utilizar a endogamia para fixar certas características de interesse,  
388 introduzir o conceito de teste de progênie, e promover a ideia 'breed the best to the best'  
389 (OLDENBROEK; VAN DER WAAIJ, 2014), acasalar os melhores com os melhores, em  
390 tradução livre.

391 Na mesma época em que Darwin viveu e publicou seus trabalhos, o monge Gregor  
392 Mendel descobria as leis básicas de hereditariedade, publicando em 1865 os resultados de  
393 seus estudos com ervilhas. Mendel demonstrou como o material genético é herdado de  
394 forma independente de ambos os genitores, e que cada indivíduo carrega duas cópias do  
395 mesmo gene, além das diferentes formas de interação entre os alelos - dominância,  
396 recessão, ou aditiva. No entanto, acredita-se que Darwin nunca teve contato com os  
397 estudos de Mendel (FAIRBANKS, 2020), que não foram reconhecidos como importantes  
398 até 1900 (OLDENBROEK; VAN DER WAAIJ, 2014).

### 399 2.3.1 Tradicional

400 A maior parte da teoria sobre melhoramento animal que utiliza-se hoje data da  
401 primeira metade do século XX (OLDENBROEK; VAN DER WAAIJ, 2014). A publicação do  
402 estatístico R. A. Fisher de 1919, na qual ele descreve o modelo infinitesimal (onde assume-  
403 se um número infinito de loci, com efeitos infinitesimalmente pequenos) e o conceito de



404 análise de variância (ANOVA), é considerada uma das contribuições que fundamentam o  
405 campo da genética quantitativa tal qual é conhecida hoje (VISSCHER; GODDARD, 2019).  
406 Sewall Wright derivou o coeficiente de endogamia (F) em seu trabalho publicado em 1922  
407 e propriedades de populações mendelianas em diversos outros estudos (GIANOLA; ROSA,  
408 2015). Juntamente com J. B. S. Haldane, estes três autores são considerados os  
409 fundadores da genética populacional teórica (OLDENBROEK; VAN DER WAAIJ, 2014).

410 J. L. Lush é conhecido como "o pai do melhoramento animal moderno", e defendia  
411 que a escolha dos animais deveria basear-se em uma combinação do uso da estatística  
412 quantitativa e informações genéticas, ao invés da aparência subjetiva dos mesmos  
413 (OLDENBROEK; VAN DER WAAIJ, 2014). Ele contribuiu de maneira excepcional para o  
414 desenvolvimento do melhoramento genético ao combinar as ideias de Fisher, Wright e  
415 Haldane em um contexto agropecuário BUSSIMAN (2021). A publicação de seu livro  
416 'Animal Breeding Plans' em 1937 inspirou L. N. Hazel a trabalhar para ele, Hazel foi o  
417 responsável por publicar a teoria do índice de seleção, que determina quais pesos devem  
418 ser colocados em diferentes características sob seleção (OLDENBROEK; VAN DER  
419 WAAIJ, 2014).

420 O estatístico C. R. Henderson, que foi aluno de Hazel, deduziu as equações de  
421 modelos mistos (MML, do inglês Mixed Model Equations) em 1949 e 1950 e desenvolveu o  
422 conceito de Estimated Breeding Value (EBV - valor genético estimado), possibilitando a  
423 classificação dos animais de acordo com o seu potencial genético estimado. Estas  
424 equações permitiram a estimação de ambos efeitos fixos (Best Linear Unbiased Estimate –  
425 BLUE) e aleatórios (Best Linear Unbiased Predictor – BLUP) para cada animal avaliado,  
426 bem como de seus parentes, a partir da integração de relacionamentos genéticos entre os  
427 indivíduos com a inclusão do pedigree. Este modelo ficou conhecido como Modelo Animal.  
428 Devido às limitações computacionais da época, apenas em 1980 estes modelos foram  
429 implementados (OLDENBROEK; VAN DER WAAIJ, 2014).

### 430 **2.3.2 Genômica**

431 Em 1990, Lande e Thompson (1990) apresentaram o conceito de Marker Assisted  
432 Selection (MAS - Seleção Assistida por Marcadores), que consiste na ideia de incluir

433 informações de marcadores com grandes efeitos em características quantitativas, ou  
434 marcadores que estivessem em desequilíbrio de ligação com estes, nas MME tradicionais.  
435 Assim, esperava-se obter predições mais acuradas. No entanto, este método apresentou  
436 certas limitações, conforme descrito pelos próprios autores: o número de marcadores  
437 moleculares necessários para captar associações significativas com os QTL, o tamanho  
438 amostral necessário das populações para detectar QTL para características com baixa  
439 herdabilidade, e erros de amostragem na estimação de pesos relativos no index de seleção  
440 combinando informações moleculares e fenotípicas (LANDE; THOMPSON, 1990).

441 Mais de uma década depois, Meuwissen, Hayes e Goddard (MEUWISSEN;  
442 HAYES; GODDARD, 2001) mostraram, através de uma simulação, ser possível estimar,  
443 com alta acurácia, os valores genéticos dos indivíduos que não apresentam fenótipos e  
444 nem progênes, ao utilizar marcadores em alta densidade em desequilíbrio de ligação com  
445 os QTL simulados. Esta abordagem compreende o que chamamos hoje de seleção  
446 genômica (LOURENCO et al., 2020).

447 Em 2008, VanRaden demonstrou como incluir informações de relacionamento  
448 genômico ao substituir a tradicional matriz de parentesco A, que utiliza-se de dados  
449 pedigree, pela matriz G, que leva em consideração os alelos compartilhados entre  
450 indivíduos. Assim, o método proposto demonstrou o potencial de aumentar a confiabilidade  
451 das predições de animais jovens de 32% para mais de 60% (VANRADEN, 2008). Esta  
452 abordagem, apesar de ter sido amplamente utilizada na época, apresenta certas limitações,  
453 como o fato de que as informações genômicas não podiam ser estendidas a animais não  
454 genotipados, de forma que a avaliação BLUP tradicional ainda se fazia necessária, e assim  
455 eram geradas diferentes métricas para cada grupo (animais genotipados e não  
456 genotipados). Também por este motivo, o método ficou conhecido como 'multistep'  
457 (múltiplos passos) (LOURENCO et al., 2020).

458 No ano seguinte, Misztal, Legarra e Aguilar apresentaram o método single-step  
459 Genomic Linear Predictor (ssGBLUP), que permite a avaliação em passo único de animais  
460 com e sem informações genômicas, ao combinar de maneira eficiente as matrizes A e G  
461 em uma única matriz, denominada matriz H (LEGARRA; AGUILAR; MISZTAL, 2009;  
462 MISZTAL; LEGARRA; AGUILAR, 2009). A partir de 2009, com o advento do

463 sequenciamento do genoma bovino, a redução dos custos de genotipagem, e  
464 disponibilização de painéis de marcadores de densidade cada vez mais elevada, a seleção  
465 genômica passou a ser cada vez mais utilizada nos rebanhos bovinos, especialmente nos  
466 leiteiros, além de demais espécies de animais de produção (LOURENCO et al., 2020) e até  
467 mesmo de companhia.

## 468 2.4 ACASALAMENTOS DIRIGIDOS

469 Uma vez que a seleção dos melhores indivíduos para gerar a próxima geração  
470 tenha sido feita, o próximo passo é decidir como tais indivíduos devem ser acasalados. A  
471 escolha de pares a serem acasalados é denominada *mate allocation* (alocação de  
472 acasalamentos ou acasalamentos dirigidos), enquanto que o processo simultâneo de  
473 seleção e alocação de machos e fêmeas em pares, para fins de acasalamento, é chamado  
474 de *mate selection* (seleção de acasalamentos) (ALLAIRE, 1980; KINGHORN, 2011;  
475 PRYCE; HAYES; GODDARD, 2012).

476 Tal processo de decisão pode levar em consideração um grande número de  
477 variáveis, incluindo o valor genético e a endogamia esperados na progênie, a  
478 coancestralidade entre os pais utilizados, o número de vezes em que um touro foi utilizado  
479 em um acasalamento e a probabilidade da progênie herdar um alelo homozigoto recessivo  
480 (COLE, 2015; KINGHORN, 2011). Pode-se ainda levar em conta restrições de grupos, ou  
481 utilizar o direcionamento de acasalamentos para aumentar-se a conectividade entre  
482 rebanhos ou a migração de genes (KINGHORN, 2011). Existem ainda casos onde a  
483 característica que se deseja melhorar é não linear ou deseja-se encontrar um ótimo  
484 intermediário (não é sempre melhor quando o valor genético é superior, como por exemplo,  
485 peso ao nascimento), nestes casos o acasalamentos dirigidos também são realizados  
486 (ALLAIRE, 1980; TORO; VARONA, 2010).

487 Neves e colaboradores (2009) desenvolveram uma simulação de dados para  
488 avaliar e comparar diferentes estratégias de acasalamentos dirigidos, com o objetivo de  
489 gerar bovinos de maior valor genético e menor variabilidade de progênie. Segundo os  
490 autores, existem duas principais metodologias para se conduzir os acasalamentos. A  
491 primeira, pode ser feita ao combinar-se touros e vacas semelhantes, denominada

492 Acasalamento Preferencial Positivo (AP). Esta estratégia ocasiona no nascimento de  
493 animais com genótipos mais extremos, o que é de interesse de programas de  
494 melhoramento genético, uma vez que deseja-se obter indivíduos com valores genéticos  
495 superiores. No entanto, caso a característica de interesse for do tipo em que se deseja obter  
496 um ótimo intermediário, tal estratégia não é recomendada. Além disso, extremos inferiores  
497 também são produzidos, o que leva a um aumento da proporção de fêmeas de valor  
498 genético inferior para a reposição do rebanho. A segunda estratégia descrita por Neves et  
499 al. (NEVES et al., 2009) é a de Acasalamento Preferencial Negativo (AN), que busca a  
500 combinação de touros e vacas não-similares. A vantagem desta estratégia é a maior  
501 uniformidade da progênie obtida, o que é de interesse na produção de animais destinados  
502 ao abate. Além disso, esta estratégia também é recomendada quando se busca um ótimo  
503 intermediário na característica sob seleção.

## 504 2.5 FERRAMENTAS COMPUTACIONAIS

### 505 2.5.1 Simulação de dados

506 Conforme discutido nos tópicos anteriores, existe uma gama de decisões que  
507 devem ser tomadas em um programa de melhoramento genético: número de machos e  
508 fêmeas a serem mantidos ou descartados, intensidade de seleção, quais informações serão  
509 incluídas nos critérios de seleção, quais modelos utilizar, decisões acerca de como formar  
510 os acasalamentos, entre outras. Uma alternativa eficiente e a baixo custo de se avaliar os  
511 efeitos de tais decisões, especialmente a longo prazo, é fazendo-se o uso da simulação de  
512 dados. Ela permite a rápida replicação e o teste de diferentes hipóteses, o que se tornaria  
513 inviável utilizando animais reais devido aos custos e tempo envolvidos (DAETWYLER et  
514 al., 2013). Outra importante vantagem da simulação de dados diz respeito à possibilidade  
515 de simplificação de sistemas complexos, permitindo, em alguma medida, o isolamento de  
516 certos efeitos e a observação dos mesmos de forma mais direta.

517 Diversos programas ou softwares já foram desenvolvidos com o intuito de gerar  
518 simulações de dados genômicos realistas (FAUX et al., 2016; GAYNOR; GORJANC;  
519 HICKEY, 2021; HOWARD et al., 2017b; SARGOLZAEI; SCHENKEL, 2009). Entre eles,  
520 existem diferentes abordagens de simulação. Segundo Yuan et al. (2012), os três principais

521 métodos de simulação disponíveis são o backward-time, forward-time e a reamostragem  
522 (resampling). A primeira abordagem, também conhecida como coalescente, consiste em  
523 "retroceder no tempo". A simulação se inicia a partir de uma amostra da geração atual e  
524 busca rastrear todos os alelos da amostra populacional até um único ancestral, chamado  
525 de ancestral comum mais recente (MRCA, do inglês Most Recent Common Ancestor). Em  
526 seguida, ela avança até a geração atual, introduzindo mutações e outros eventos na  
527 genealogia gerada (YUAN et al., 2012).

528 O método forward-time, por sua vez, é um método simples muito utilizado nos  
529 estudos com populações animais (DAETWYLER et al., 2013). Este tipo de simulação parte  
530 de uma população inicial e acompanha a sua evolução por múltiplas gerações seguindo  
531 diversos modelos genéticos, e a amostragem é feita geralmente na última ou em uma das  
532 últimas gerações simuladas. Segundo Daetwyler e colaboradores (2013), este método de  
533 simulação é capaz de replicar de maneira precisa a trajetória evolutiva, mesmo que  
534 complexa, de qualquer população real. Simulações baseadas em reamostragem consistem  
535 em utilizar uma amostra de sequências genômicas ou haplótipos reais para formar uma  
536 base de indivíduos e gerar os genomas presentes em uma população, seja com um  
537 pedigree real ou simulado. Este método é um dos melhores quando se tem como objetivo  
538 replicar com exatidão as frequências alélicas e desequilíbrio de ligação presentes nas  
539 populações de interesse, no entanto apresenta maiores limitações na introdução de novas  
540 características genéticas e mutações (DAETWYLER et al., 2013).

## 541 **2.5.2 Linguagens de programação**

542 Com o avanço das inovações tecnológicas, nos aproximamos de uma realidade em  
543 que as habilidades básicas de programação e análise de dados são quase tão requisitadas  
544 quanto o saber do inglês. Na área do agronegócio, isso não é diferente. O estudo do  
545 melhoramento genético, por exemplo, exige uma base sólida de conhecimentos em  
546 estatística (FALCONER; MACKAY, 1996) e a implementação de modelos matemáticos  
547 através de programas específicos. Entre os programas mais utilizados no melhoramento  
548 genético, encontram-se o QMSim (SARGOLZAEI; SCHENKEL, 2009) para a simulação de  
549 dados, o FIMPUTE (SARGOLZAEI; CHESNAIS; SCHENKEL, 2014) para a imputação de  
550 genótipos, e os programas da família BLUPF90 (MISZTAL et al., 2014), todos

551 desenvolvidos por pesquisadores da área, utilizando-se linguagens de programação de  
552 baixo nível (mais próximas do computador e mais distantes do usuário), como Fortran e  
553 C++. Em áreas mais abrangentes da medicina veterinária e zootecnia, é imprescindível a  
554 realização de estudos científicos com as devidas análises e testes estatísticos, e a sua  
555 execução não é viável sem o uso de Softwares ou linguagens de programação como SAS  
556 (SAS Institute. 2011), R (R CORE TEAM, 2020), Python (VAN ROSSUM; DRAKE JR,  
557 1995), ou similares.

558 As grandes fazendas, e por muitas vezes as pequenas propriedades também  
559 (PINTO et al., 2023), caminham cada vez mais em direção à automatização, seja com o  
560 objetivo de executar tarefas rotineiras (BEWLEY, 2010), coletar fenótipos (EGGER-  
561 DANNER et al., 2014; ESPOSITO et al., 2020), detectar doenças (NEETHIRAJAN, 2017;  
562 RUTTEN et al., 2013; VAN NUFFEL et al., 2015), ou ainda, buscando-se aumentar a  
563 eficiência produtiva do rebanho por meio de múltiplas medições e análises em conjunto  
564 (CURTI et al., 2023; DUTTA et al., 2015; KAMILARIS; KARTAKOULLIS; PRENAFETA-  
565 BOLDÚ, 2017; TULLO; FINZI; GUARINO, 2019). Para este fim, alguns dos conhecimentos  
566 necessários envolvem o Aprendizado de Máquinas, Aprendizagem Profunda, Inteligência  
567 Artificial, e Visão Computacional (FUENTES et al., 2022), geralmente implementadas por  
568 meio da linguagem Python.

569 Uma outra área da programação de interesse do agronegócio é o desenvolvimento  
570 de aplicativos. Estes aplicativos podem apresentar aplicações desde o gerenciamento de  
571 rebanhos, reunindo em uma interface os diversos dispositivos de monitoramento de uma  
572 fazenda ou performando análises em bancos de dados (CURTI et al., 2023; DUTTA et al.,  
573 2015; KAMILARIS; KARTAKOULLIS; PRENAFETA-BOLDÚ, 2017; TULLO; FINZI;  
574 GUARINO, 2019); (FUENTES et al., 2022)), até o diagnóstico de imagens em uma clínica  
575 veterinária (BORTOLUZZI et al., 2023; ZURAW; AEFNER, 2022) ou a avaliação de  
576 animais em um programa de melhoramento genético (COLE; VANRADEN, 2010; SELLI;  
577 MILLER; VENTURA, 2024). Para atingir este objetivo, diversas linguagens de programação  
578 podem ser utilizadas, a depender do grau de complexidade do aplicativo. É possível criar  
579 interfaces atraentes com ferramentas acessíveis como a linguagem R e o pacote Shiny  
580 (CHANG et al., 2024), Software de Business Intelligence como o Power BI e Tableau, ou  
581 criar aplicações mais complexas com o uso de linguagens tais quais Python, Java

582 (ARNOLD; GOSLING; HOLMES, 2005), JavaScript, Swift (para dispositivos da Apple)  
583 (SCHALLER et al., 2018), C# (para aplicativos Windows), entre outras.

### 584 **2.5.3 Visualização de dados**

585 Uma das maneiras mais eficientes de se passar uma mensagem é através do uso  
586 de imagens ou gráficos. Visualizações bem elaboradas são capazes de apresentar mais  
587 informações em uma mesma área quando comparados a textos ou tabelas, além de facilitar  
588 o entendimento sobre os dados. Desta forma, é mais fácil perceber diferenças sutis ao  
589 observar gráficos do que outras formas de representações, o que possibilita a apresentação  
590 de informações de maneira mais compacta (COLE; VANRADEN, 2010). Tal eficiência das  
591 visualizações, tanto na transmissão de ideias quanto na capacidade de serem retidas pelo  
592 cérebro humano, se deve em grande parte à habilidade do sistema visual humano em  
593 captar, com baixo esforço e de forma eficiente, uma quantidade imensa de informações  
594 (VAN ESSEN; ANDERSON; FELLEMAN, 1992).

595 A tarefa de representar dados através de gráficos, no entanto, não é  
596 necessariamente intuitiva. Existem diversas formas de apresentar a mesma informação,  
597 como por exemplo utilizando-se gráficos de pontos, linhas, barras, ou muitos outros. Tal  
598 escolha pode ter grande impacto na forma como a informação é recebida pelo espectador  
599 (ROUGIER; DROETTBOOM; BOURNE, 2014). É possível ainda, que a adição de certos  
600 adornos, ou mesmo a escolha de cores e escalas tenha um efeito negativo na compreensão  
601 dos dados, o que por muitas vezes torna o gráfico tendencioso ou enganador (BATEMAN  
602 et al., 2018).

603 Rougier et al. (2014) publicaram "Ten Simple Rules for Better Figures" (10 regras  
604 simples para fazer figuras melhores, em tradução livre). Entre elas, destaca-se a  
605 necessidade de se conhecer o público alvo, ter clareza sobre qual a mensagem pretendida,  
606 ponderar sobre em que meio a figura será divulgada, incluir legendas descritivas, utilizar  
607 opções de customização, fazer uso das cores de forma eficiente, evitar confundir o  
608 espectador, evitar adornos desnecessários, focar na mensagem que se deseja transmitir,  
609 e fazer uso da ferramenta adequada. Durant et al. (2022), por sua vez, descreveram em  
610 seu artigo "Ten simple rules for developing visualization tools in genomics" (10 regras

611 simples para desenvolver ferramentas de visualização de dados genômicos, em tradução  
612 livre). Os autores focam no desenvolvimento de ferramentas de visualização ao invés das  
613 figuras em si, e discutem sobre o envolvimento de colaboradores, o processo de  
614 desenvolvimento e validação da aplicação, e a publicação final.

615           Dentre as diversas escolhas que devem ser feitas durante a elaboração de figuras,  
616 alguns pontos devem ser ressaltados. Alguns tipos de gráficos costumam ser associados  
617 com certos tipos de dados, como por exemplo o uso de gráficos de barras para a  
618 comparação de variáveis discretas, enquanto que gráficos de linhas são normalmente  
619 associados a dados temporais (ZACKS; TVERSKY, 1999). Zacks e Tversky (1999)  
620 apontam que a mistura desses tipos de representação pode causar confusões aos  
621 usuários. Segundo Rougier et al. (2014), o uso de gráficos 3D ou do tipo "torta" (pie charts)  
622 deve ser evitado para a comparação de quantidades, pois induzem uma percepção  
623 incorreta. É possível também a criação de estilos novos de gráficos, quando se pretende  
624 apresentar dados que sejam muito complexos e que não possam ser representados por  
625 meio dos tipos de gráficos convencionais (DURANT et al., 2022; SELLI et al., 2021; SELLI;  
626 MILLER; VENTURA, 2024).

627           A escolha das cores representa outro ponto de atenção. O uso excessivo de cores,  
628 como por exemplo ao adotar a paleta de cores de arco-íris, sobrecarrega a imagem e pode  
629 tirar a atenção do espectador de detalhes mais relevantes. Por este motivo, Rougier et al.  
630 (2014) recomendam que o autor considere a real necessidade da adoção de diferentes  
631 cores em um gráfico. Caso o seu uso não traga nenhum ganho relevante, ele deve ser  
632 evitado. Além disso, para diferentes tipos de dados, recomenda-se o uso de tipos de paletas  
633 de cores específicas. Conforme descrito por Rougier et al. (2014): as paletas sequenciais  
634 apresentam variação de um único tipo de cor, e são indicadas para representar dados  
635 quantitativos; Já as paletas de cores divergentes variam de uma cor para outra, e devem  
636 ser utilizadas para ressaltar o desvio de valores da mediana; Paletas de cores qualitativas  
637 apresentam grandes variações entre as cores, e são recomendadas para a representação  
638 de dados discretos ou categóricos.

639           Existe uma série de ferramentas que podem ser utilizadas para a confecção de  
640 visualizações. O pacote R ggplot2 e as bibliotecas matplotlib e seaborn, implementadas em



641 linguagem Python, são alguns exemplos de ferramentas bastante flexíveis e que permitem  
 642 a criação de uma grande variedade de gráficos estáticos, inclusive com a combinação de  
 643 diferentes tipos. O uso de gráficos dinâmicos, por outro lado, tem despertado crescente  
 644 interesse por parte de pessoas tanto da academia quanto da indústria (ALI et al., 2016;  
 645 MOROTA et al., 2021; SELLI; MILLER; VENTURA, 2024).

646 Gráficos dinâmicos permitem a apresentação de um maior volume de dados através  
 647 do uso de animações ou filtros. Com a implementação de filtros, tais gráficos tornam-se  
 648 interativos, permitindo que o usuário faça certas escolhas acerca do que deseja visualizar.  
 649 Hullman e Diakopoulos (HULLMAN; DIAKOPOULOS, 2011) sugeriram que a escolha de  
 650 quais dados serão apresentados poderia ser confiada ao usuário quando a interatividade  
 651 estiver habilitada e a inserção de elementos como botões e barras de pesquisa for possível.  
 652 O mesmo pode ser aplicado à escala de cores, período referente aos dados, categorias,  
 653 entre outros. Entre as ferramentas que permitem a elaboração de tais visualizações, estão  
 654 o pacote R Shiny e os pacotes Plotly e Dash, que podem ser implementados em R ou  
 655 Python, e correspondem a ferramentas de código aberto. Existem também ferramentas  
 656 proprietárias, como por exemplo o Tableau e o Power BI, que não exigem o conhecimento  
 657 de programação, porém apresentam certas limitações, conforme discutido por (SELLI;  
 658 MILLER; VENTURA, 2024).

## 659 2.6 REFERÊNCIAS

- 660 ABIED, A.; XU, L.; SAHLU, B. W.; XING, F.; AHBARA, A.; PU, Y.; LIN, J.; BERIHULAY,  
 661 H.; ISLAM, R.; HE, X.; MWACHARO, J. M.; ZHAO, Q.; MA, Y. Genome-Wide Analysis  
 662 Revealed Homozygosity and Demographic History of Five Chinese Sheep Breeds  
 663 Adapted to Different Environments. **Genes**, v. 11, n. 12, p. 1480, 9 dez. 2020.  
 664  
 665 ALI, S. M.; GUPTA, N.; NAYAK, G. K.; LENKA, R. K. Big data visualization: Tools and  
 666 challenges. Em: Proceedings of the 2016 2nd International Conference on Contemporary  
 667 Computing and Informatics, IC3I 2016, 2016, [...]. Institute of Electrical and Electronics  
 668 Engineers Inc., 2016. p. 656–660.  
 669  
 670 ALLAIRE, F. R. Mate selection by selection index theory. **Theoretical and Applied**  
 671 **Genetics**, v. 57, n. 6, p. 267–272, 1980.  
 672  
 673 AL-MAMUN, H. A.; A CLARK, S.; KWAN, P.; GONDRO, C. Genome-wide linkage  
 674 disequilibrium and genetic diversity in five populations of Australian domestic sheep.  
 675 **Genetics Selection Evolution**, v. 47, n. 1, p. 1–14, 2015.

- 676  
677 ARDLIE, K. G.; KRUGLYAK, L.; SEIELSTAD, M. Patterns of linkage disequilibrium in the  
678 human genome. **Nature Reviews Genetics** 2002 3:4, v. 3, n. 4, p. 299–309, 2002.
- 679  
680 ARNOLD, K., GOSLING, J. & HOLMES, D. **The Java programming language**, Addison  
681 Wesley Professional, 2005.
- 682  
683 BATEMAN, S.; LEPROU, J.; JULIAN, A.; WEIR, A. Atypical cases of Zinc poisoning in  
684 Taranaki dairy herds during the spring of 2012. n. January 2013, p. 1–19, 2018.
- 685  
686 BEWLEY, J. Precision dairy farming: Advanced analysis solutions for future profitability.  
687 **The First North American Conference on Precision Dairy Management**, n. November,  
688 p. 16, 2010.
- 689  
690 BISCARINI, F.; COZZI, P.; GASPA, G.; MARRAS, G. **detectRUNS: an R package to**  
691 **detect runs of homozygosity and heterozygosity in diploid genomes**CRAN (The  
692 **Comprehensive R Archive Network)**2018. Disponível em: <[https://cran.r-](https://cran.r-project.org/web/packages/detectRUNS/vignettes/detectRUNS.vignette.html#references)  
693 [project.org/web/packages/detectRUNS/vignettes/detectRUNS.vignette.html#references](https://cran.r-project.org/web/packages/detectRUNS/vignettes/detectRUNS.vignette.html#references)>.
- 694  
695 BJELLAND, D. W.; WEIGEL, K. A.; VUKASINOVIC, N.; NKRUMAH, J. D. Evaluation of  
696 inbreeding depression in Holstein cattle using whole-genome SNP markers and alternative  
697 measures of genomic inbreeding. **Journal of Dairy Science**, v. 96, n. 7, p. 4697–4706,  
698 2013.
- 699  
700 BOHMANOVA, J.; SARGOLZAEI, M.; SCHENKEL, F. S. Characteristics of linkage  
701 disequilibrium in North American Holsteins. **BMC Genomics**, v. 11, n. 1, 2010.
- 702  
703 BORTOLUZZI, E.; SCHMIDT, P.; BROWN, R.; JENSEN, M.; MANCKE, M.; LARSON, R.;  
704 LANCASTER, P.; WHITE, B. Image Classification and Automated Machine Learning to  
705 Classify Lung Pathologies in Deceased Feedlot Cattle. **Veterinary Sciences**, v. 10, n. 2,  
706 p. 113, 3 fev. 2023.
- 707  
708 BOSSE, M.; MEGENS, H.-J.; MADSEN, O.; PAUDEL, Y.; FRANTZ, L. A. F.; SCHOOK, L.  
709 B.; CROOIJMANS, R. P. M. A.; GROENEN, M. A. M. Regions of Homozygosity in the  
710 Porcine Genome: Consequence of Demography and the Recombination Landscape.  
711 **PLoS Genetics**, v. 8, n. 11, p. e1003100, 29 nov. 2012.
- 712  
713 BRITO, F. V.; NETO, J. B.; SARGOLZAEI, M.; COBUCI, J. A.; SCHENKEL, F. S.  
714 Accuracy of genomic selection in simulated populations mimicking the extent of linkage  
715 disequilibrium in beef cattle. **BMC Genetics**, v. 12, 2011.
- 716  
717 BROMAN, K. W.; WEBER, J. L. Long homozygous chromosomal segments in reference  
718 families from the Centre d'Etude du Polymorphisme Humain. **American Journal of**  
719 **Human Genetics**, v. 65, n. 6, p. 1493–1500, 1999.
- 720  
721 BUSSIMAN, F. O. **Avaliação genética de cavalos Campolina: do pedigree à**  
722 **possibilidade de seleção genômica** [Genetic evaluation of Campolina horses: from

- 723 pedigree to the genomic selection possibility]. 2021. 233 f. Tese (Doutorado em Ciências)  
724 – Faculdade de Medicina Veterinária e Zootecnia, Universidade de São Paulo, 2021.  
725
- 726 CEBALLOS, F. C.; HAZELHURST, S.; RAMSAY, M. Assessing runs of Homozygosity: a  
727 comparison of SNP Array and whole genome sequence low coverage data. **BMC**  
728 **Genomics**, v. 19, n. 1, p. 106, 30 dez. 2018.  
729
- 730 CHANG, W.; CHENG, J.; ALLAIRE, J. J.; SIEVERT, C.; SCHLOERKE, B.; XIE, Y.;  
731 ALLEN, J.; MCPHERSON, J.; DIPERT, A.; BORGES, B. **shiny: Web Application**  
732 **Framework for R**. [s.l: s.n.]  
733
- 734 COLE, J. B. A Simple Strategy for Managing Many Recessive Disorders in a Dairy Cattle  
735 Breeding Program. **Genetics Selection Evolution**, v. 47, n. 1, p. 94, dez. 2015.  
736
- 737 COLE, J. B.; VANRADEN, P. M. Visualization of results from genomic evaluations.  
738 **Journal of Dairy Science**, v. 93, n. 6, p. 2727–2740, 1 jun. 2010.  
739
- 740 CURTI, P. D. F.; SELLI, A.; PINTO, D. L.; MERLOS-RUIZ, A.; BALIEIRO, J. C. D. C.;  
741 VENTURA, R. V. Applications of Livestock Monitoring Devices and Machine Learning  
742 Algorithms in Animal Production and Reproduction: An Overview. **Animal Reproduction**,  
743 v. 20, n. 2, p. e20230077, 2023.  
744
- 745 DAETWYLER, H. D.; CALUS, M. P. L.; PONG-WONG, R.; DE LOS CAMPOS, G.;  
746 HICKEY, J. M. Genomic Prediction in Animals and Plants: Simulation of Data, Validation,  
747 Reporting, and Benchmarking. **Genetics**, v. 193, n. 2, p. 347–365, 1 fev. 2013.  
748
- 749 DURANT, E.; ROUARD, M.; GANKO, E. W.; MULLER, C.; CLEARY, A. M.; FARMER, A.  
750 D.; CONTE, M.; SABOT, F. Ten Simple Rules for Developing Visualization Tools in  
751 Genomics. **PLOS Computational Biology**, v. 18, n. 11, p. e1010622, 10 nov. 2022.  
752
- 753 DUTTA, R.; SMITH, D.; RAWNSLEY, R.; BISHOP-HURLEY, G.; HILLS, J.; TIMMS, G.;  
754 HENRY, D. Dynamic cattle behavioural classification using supervised ensemble  
755 classifiers. **Computers and Electronics in Agriculture**, v. 111, p. 18–28, fev. 2015.  
756
- 757 DZOMBA, E. F.; CHIMONYO, M.; PIERNEEF, R.; MUCHADEYI, F. C. Runs of  
758 homozygosity analysis of South African sheep breeds from various production systems  
759 investigated using OvineSNP50k data. **BMC Genomics**, v. 22, n. 1, p. 7, 6 dez. 2021.  
760
- 761 EGGER-DANNER, C.; COLE, J. B.; PRYCE, J. E.; GENGLER, N.; HERINGSTAD, B.;  
762 BRADLEY, A.; STOCK, K. F. Invited review: Overview of new traits and phenotyping  
763 strategies in dairy cattle with a focus on functional traits. **Animal**, v. 9, n. 2, p. 191–207,  
764 2014.  
765
- 766 ESPOSITO, S.; CARPUTO, D.; CARDI, T.; TRIPODI, P. Applications and trends of  
767 machine learning in genomics and phenomics for next-generation breeding. **Plants**, v. 9,  
768 n. 1, 2020.  
769

- 770 FAIRBANKS, D. J. Mendel and Darwin: Untangling a Persistent Enigma. **Heredity**, v. 124,  
771 n. 2, p. 263–273, fev. 2020.
- 772
- 773 FALCONER, D. S.; MACKAY, T. F. C. **Introduction to Quantitative Genetics Fourth**  
774 **Edition**. [s.l: s.n.]
- 775
- 776 FAUX, A.; GORJANC, G.; GAYNOR, R. C.; BATTAGIN, M.; EDWARDS, S. M.; WILSON,  
777 D. L.; HEARNE, S. J.; GONEN, S.; HICKEY, J. M. AlphaSim: Software for Breeding  
778 Program Simulation. **The Plant Genome**, v. 9, n. 3, p. 1–14, 2016.
- 779
- 780 FERENČAKOVIĆ, M.; SÖLKNER, J.; CURIK, I. Estimating autozygosity from high-  
781 throughput information: effects of SNP density and genotyping errors. **Genetics Selection**  
782 **Evolution**, v. 45, n. 1, p. 42, 2013.
- 783
- 784 FREEMAN, A. R.; HOGGART, C. J.; HANOTTE, O.; BRADLEY, D. G. Assessing the  
785 Relative Ages of Admixture in the Bovine Hybrid Zones of Africa and the Near East Using  
786 X Chromosome Haplotype Mosaicism. **Genetics**, v. 173, n. 3, p. 1503–1510, 1 jul. 2006.
- 787
- 788 FUENTES, S.; GONZALEZ VIEJO, C.; TONGSON, E.; DUNSHEA, F. R. The Livestock  
789 Farming Digital Transformation: Implementation of New and Emerging Technologies Using  
790 Artificial Intelligence. **Animal Health Research Reviews**, v. 23, n. 1, p. 59–71, jun. 2022.
- 791
- 792 GAYNOR, R. C.; GORJANC, G.; HICKEY, J. M. AlphaSimR: an R package for breeding  
793 program simulations. **G3 Genes|Genomes|Genetics**, v. 11, n. 2, 12 abr. 2021. Disponível  
794 em: <<https://academic.oup.com/g3journal/article/doi/10.1093/g3journal/jkaa017/6025179>>.  
795 Acesso em: 31 mar. 2022.
- 796
- 797 GIANOLA, D.; ROSA, G. J. M. One Hundred Years of Statistical Developments in Animal  
798 Breeding. **Annual Review of Animal Biosciences**, v. 3, n. 1, p. 19–56, 16 fev. 2015.
- 799
- 800 HERRERO-MEDRANO, J.; MEGENS, H.-J.; GROENEN, M. A.; RAMIS, G.; BOSSE, M.;  
801 PÉREZ-ENCISO, M.; CROOIJMANS, R. P. Conservation genomic analysis of domestic  
802 and wild pig populations from the Iberian Peninsula. **BMC Genetics**, v. 14, n. 1, p. 106,  
803 2013.
- 804
- 805 HILL, W. G.; ROBERTSON, A. Linkage disequilibrium in finite populations. **Theoretical**  
806 **and Applied Genetics**, v. 38, n. 6, p. 226–231, jun. 1968.
- 807
- 808 HOWARD, J. T.; PRYCE, J. E.; BAES, C.; MALTECCA, C. Invited review: Inbreeding in  
809 the genomics era: Inbreeding, inbreeding depression, and management of genomic  
810 variability. **Journal of Dairy Science**, v. 100, n. 8, p. 6009–6024, 2017a.
- 811
- 812 HOWARD, J. T.; TIEZZI, F.; PRYCE, J. E.; MALTECCA, C. Geno-Diver: A combined  
813 coalescence and forward-in-time simulator for populations undergoing selection for  
814 complex traits. **Journal of Animal Breeding and Genetics**, v. 134, n. 6, p. 553–563,  
815 2017b.
- 816

- 817 HOWRIGAN, D. P.; SIMONSON, M. A.; KELLER, M. C. Detecting autozygosity through  
818 runs of homozygosity: A comparison of three autozygosity detection algorithms. **BMC**  
819 **Genomics**, v. 12, n. 1, p. 460, 23 dez. 2011.
- 820  
821 HULLMAN, J.; DIAKOPOULOS, N. Visualization Rhetoric: Framing Effects in Narrative  
822 Visualization. **IEEE Transactions on Visualization and Computer Graphics**, v. 17, n.  
823 12, p. 2231–2240, dez. 2011.
- 824  
825 KAMILARIS, A.; KARTAKOULLIS, A.; PRENAFETA-BOLDÚ, F. X. A review on the  
826 practice of big data analysis in agriculture. **Computers and Electronics in Agriculture**, v.  
827 143, n. September, p. 23–37, dez. 2017.
- 828  
829 KINGHORN, B. P. An algorithm for efficient constrained mate selection. **Genetics**  
830 **Selection Evolution**, v. 43, n. 1, p. 4, 20 dez. 2011.
- 831  
832 KRISTENSEN, T. N.; SØRENSEN, A. C. Inbreeding – lessons from animal breeding,  
833 evolutionary biology and conservation genetics. **Animal Science**, v. 80, n. 2, p. 121–133,  
834 9 abr. 2005.
- 835  
836 LANDE, R.; THOMPSON, R. Efficiency of Marker-Assisted Selection in the Improvement  
837 of Quantitative Traits. **Genetics**, v. 124, n. 3, p. 743–756, 1 mar. 1990.
- 838  
839 LEGARRA, A.; AGUILAR, I.; MISZTAL, I. A relationship matrix including full pedigree and  
840 genomic information. **Journal of Dairy Science**, v. 92, n. 9, p. 4656–4663, 2009.
- 841  
842 LEROY, G. Inbreeding depression in livestock species: Review and meta-analysis. **Animal**  
843 **Genetics**, v. 45, n. 5, p. 618–628, 2014.
- 844  
845 LEWONTIN, R. C. THE INTERACTION OF SELECTION AND LINKAGE. I. GENERAL  
846 CONSIDERATIONS; HETEROTIC MODELS. **Genetics**, v. 49, n. 1, p. 49–67, 10 jan.  
847 1964.
- 848  
849 LI, N.; STEPHENS, M. Modeling Linkage Disequilibrium and Identifying Recombination  
850 Hotspots Using Single-Nucleotide Polymorphism Data. **Genetics**, v. 165, n. 4, p. 2213–  
851 2233, 1 dez. 2003.
- 852  
853 LOURENCO, D.; LEGARRA, A.; TSURUTA, S.; MASUDA, Y.; AGUILAR, I.; MISZTAL, I.  
854 Single-Step Genomic Evaluations from Theory to Practice: Using SNP Chips and  
855 Sequence Data in BLUPF90. **Genes**, v. 11, n. 7, p. 790, 14 jul. 2020.
- 856  
857 LU, D.; SARGOLZAEI, M.; KELLY, M.; LI, C.; VOORT, G. V.; WANG, Z.; PLASTOW, G.;  
858 MOORE, S.; MILLER, S. P. Linkage disequilibrium in Angus, Charolais, and Crossbred  
859 beef cattle. **Frontiers in Genetics**, v. 3, n. AUG, p. 1–10, 2012.
- 860  
861 MARRAS, G.; GASPA, G.; SORBOLINI, S.; DIMAURO, C.; AJMONE-MARSAN, P.;  
862 VALENTINI, A.; WILLIAMS, J. L.; MACCIOTTA, N. P. P. Analysis of runs of homozygosity  
863 and their relationship with inbreeding in five cattle breeds farmed in Italy. **Animal**  
864 **Genetics**, v. 46, n. 2, p. 110–121, abr. 2015.

- 865  
866 MASTRANGELO, S.; TOLONE, M.; DI GERLANDO, R.; FONTANESI, L.; SARDINA, M.  
867 T.; PORTOLANO, B. Genomic inbreeding estimation in small populations: Evaluation of  
868 runs of homozygosity in three local dairy cattle breeds. **Animal**, v. 10, n. 5, p. 746–754,  
869 2016.
- 870  
871 MCEVOY, B. P.; POWELL, J. E.; GODDARD, M. E.; VISSCHER, P. M. Human population  
872 dispersal “Out of Africa” estimated from linkage disequilibrium and allele frequencies of  
873 SNPs. **Genome Research**, v. 21, n. 6, p. 821–829, jun. 2011.
- 874  
875 MCQUILLAN, R.; LEUTENEGGER, A.-L.; ABDEL-RAHMAN, R.; FRANKLIN, C. S.;  
876 PERICIC, M.; BARAC-LAUC, L.; SMOLEJ-NARANCIC, N.; JANICIJEVIC, B.; POLASEK,  
877 O.; TENESA, A.; MACLEOD, A. K.; FARRINGTON, S. M.; RUDAN, P.; HAYWARD, C.;  
878 VITART, V.; RUDAN, I.; WILD, S. H.; DUNLOP, M. G.; WRIGHT, A. F.; CAMPBELL, H.;  
879 WILSON, J. F.  
880 Runs of Homozygosity in European Populations. **The American Journal of Human**  
881 **Genetics**, v. 83, n. 3, p. 359–372, set. 2008.
- 882  
883 MEUWISSEN, T. H. E.; GODDARD, M. E. Fine mapping of quantitative trait loci using  
884 linkage disequilibria with closely linked marker loci. **Genetics**, v. 155, n. 1, p. 421–430,  
885 2000.
- 886  
887 MEUWISSEN, T. H. E.; HAYES, B. J.; GODDARD, M. E. Prediction of Total Genetic Value  
888 Using Genome-Wide Dense Marker Maps. **Genetics**, v. 157, n. 4, p. 1819–1829, 1 abr.  
889 2001.
- 890  
891 MEYERMANS, R.; GORSSEN, W.; BUYS, N.; JANSSENS, S. How to study runs of  
892 homozygosity using PLINK? A guide for analyzing medium density SNP data in livestock  
893 and pet species. **BMC Genomics**, v. 21, n. 1, p. 94, 29 dez. 2020.
- 894  
895 MISZTAL, I.; LEGARRA, A.; AGUILAR, I. Computing procedures for genetic evaluation  
896 including phenotypic, full pedigree, and genomic information. **Journal of Dairy Science**, v.  
897 92, n. 9, p. 4648–4655, 2009.
- 898  
899 MISZTAL, I.; TSURUTA, S.; LOURENCO, D.A.L.; AGUILAR, I.; LEGARRA, A.; VITEZICA,  
900 Z. **Manual for BLUPF90 family of programs**. 2014. Disponível em:  
901 [http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90\\_all2.pdf](http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all2.pdf). Acesso em: 08 Mar.  
902 2024.
- 903  
904 MOROTA, G.; CHENG, H.; COOK, D.; TANAKA, E. ASAS-NANP SYMPOSIUM:  
905 Prospects for Interactive and Dynamic Graphics in the Era of Data-Rich Animal Science1.  
906 **Journal of Animal Science**, v. 99, n. 2, p. skaa402, 1 fev. 2021.
- 907  
908 MUIR, W. M.; WONG, G. K.-S.; ZHANG, Y.; WANG, J.; GROENEN, M. A. M.;  
909 CROOIJMANS, R. P. M. A.; MEGENS, H.-J.; ZHANG, H.; OKIMOTO, R.; VEREIJKEN, A.;  
910 JUNGIERIUS, A.; ALBERS, G. A. A.; LAWLEY, C. T.; DELANY, M. E.; MACEACHERN, S.;  
911 CHENG, H. H. Genome-wide assessment of worldwide chicken SNP genetic diversity

- 912 indicates significant absence of rare alleles in commercial breeds. **Proceedings of the**  
913 **National Academy of Sciences**, v. 105, n. 45, p. 17312–17317, 11 nov. 2008.
- 914  
915 NEETHIRAJAN, S. Recent advances in wearable sensors for animal health management.  
916 **Sensing and Bio-Sensing Research**, v. 12, p. 15–29, 2017.
- 917  
918 NEVES, H. H. D. R.; CARVALHEIRO, R.; CARDOSO, V.; FRIES, L. A.; QUEIROZ, S. A.  
919 D. Acasalamento dirigido para aumentar a produção de animais geneticamente superiores  
920 e reduzir a variabilidade da progênie em bovinos. **Revista Brasileira de Zootecnia**, v. 38,  
921 n. 7, p. 1201–1204, jul. 2009.
- 922  
923 OLDENBROEK, K.; VAN DER WAAIJ, L. **Textbook Animal Breeding: Animal Breeding**  
924 **and Genetics for BSc Students**. [s.l.] Centre for Genetic Resources and Animal Breeding  
925 and Genomics Group, Wageningen University and Research Centre, 2014.
- 926  
927 PÉREZ O'BRIEN, A. M.; MÉSZÁROS, G.; UTSUNOMIYA, Y. T.; SONSTEGARD, T. S.;  
928 GARCIA, J. F.; VAN TASSELL, C. P.; CARVALHEIRO, R.; DA SILVA, M. V. B.;  
929 SÖLKNER, J. Linkage disequilibrium levels in *Bos indicus* and *Bos taurus* cattle using  
930 medium and high density SNP chip data and different minor allele frequency distributions.  
931 **Livestock Science**, v. 166, n. 1, p. 121–132, ago. 2014.
- 932  
933 PERIPOLLI, E.; METZGER, J.; DE LEMOS, M. V. A.; STAFUZZA, N. B.; KLUSKA, S.;  
934 OLIVIERI, B. F.; FEITOSA, F. L. B.; BERTON, M. P.; LOPES, F. B.; MUNARI, D. P.;  
935 LÔBO, R. B.; MAGNABOSCO, C. D. U.; DI CROCE, F.; OSTERSTOCK, J.; DENISE, S.;  
936 PEREIRA, A. S. C.; BALDI, F. Autozygosity islands and ROH patterns in Nellore lineages:  
937 evidence of selection for functionally important traits. **BMC Genomics**, v. 19, n. 1, p. 680,  
938 17 dez. 2018.
- 939  
940 PERIPOLLI, E.; MUNARI, D. P.; SILVA, M. V. G. B.; LIMA, A. L. F.; IRGANG, R.; BALDI,  
941 F. Runs of homozygosity: current knowledge and applications in livestock. **Animal**  
942 **Genetics**, v. 48, n. 3, p. 255–271, jun. 2017.
- 943  
944 PINTO, D. L.; SELLI, A.; TULPAN, D.; ANDRIETTA, L. T.; GARBOSSA, P. L. M.; VOORT,  
945 G. V.; MUNRO, J.; MCMORRIS, M.; ALVES, A. A. C.; CARVALHEIRO, R.; POLETI, M. D.;  
946 BALIEIRO, J. C. de C.; VENTURA, R. V. Image feature extraction via local binary patterns  
947 for marbling score classification in beef cattle using tree-based algorithms. **Livestock**  
948 **Science**, v. 267, 1 jan. 2023. . Acesso em: 31 jan. 2023.
- 949  
950 PRYCE, J. E.; HAILE-MARIAM, M.; GODDARD, M. E.; HAYES, B. J. Identification of  
951 genomic regions associated with inbreeding depression in Holstein and Jersey dairy cattle.  
952 **Genetics Selection Evolution**, v. 46, n. 1, p. 1–14, 2014.
- 953  
954 PRYCE, J. E.; HAYES, B. J.; GODDARD, M. E. Novel strategies to minimize progeny  
955 inbreeding while maximizing genetic gain using genomic information. **Journal of Dairy**  
956 **Science**, v. 95, n. 1, p. 377–388, 2012.
- 957  
958 PURCELL, S.; NEALE, B.; TODD-BROWN, K.; THOMAS, L.; FERREIRA, M. A. R.;  
959 BENDER, D.; MALLER, J.; SKLAR, P.; DE BAKKER, P. I. W.; DALY, M. J.; SHAM, P. C.

- 960 PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage  
961 Analyses. **The American Journal of Human Genetics**, v. 81, n. 3, p. 559–575, set.  
962 2007.
- 963  
964 PURFIELD, D. C.; BERRY, D. P.; MCPARLAND, S.; BRADLEY, D. G. Runs of  
965 homozygosity and population history in cattle. **BMC Genetics**, v. 13, n. 1, p. 70, 2012.  
966
- 967 R CORE TEAM. **R: A Language and Environment for Statistical Computing** Vienna,  
968 Austria. R Foundation for Statistical Computing, 2020. Disponível em: <[https://www.r-  
969 project.org/](https://www.r-project.org/)>.
- 970  
971 REICH, D. E.; CARGILI, M.; BOIK, S.; IRELAND, J.; SABETI, P. C.; RICHTER, D. J.;  
972 LAVERY, T.; KOUYOUMJIAN, R.; FARHADIAN, S. F.; WARD, R.; LANDER, E. S. Linkage  
973 disequilibrium in the human genome. **Nature** **2001 411:6834**, v. 411, n. 6834, p. 199–204,  
974 10 maio 2001.
- 975  
976 ROBERTSON, A. Inbreeding in artificial selection programmes. **Genetical Research**, v. 2,  
977 n. 2, p. 189–194, 14 jul. 1961.
- 978 ROUGIER, N. P.; DROETTBOOM, M.; BOURNE, P. E. Ten Simple Rules for Better  
979 Figures. **PLoS Computational Biology**, v. 10, n. 9, p. e1003833, 11 set. 2014.  
980
- 981 RUTTEN, C. J.; VELTHUIS, A. G. J.; STEENEVELD, W.; HOGEVEEN, H. Invited review:  
982 Sensors to support health management on dairy farms. **Journal of Dairy Science**, v. 96,  
983 n. 4, p. 1928–1952, abr. 2013.
- 984  
985 SANTOS, W.; SCHETTINI, G.; FONSECA, M. G.; PEREIRA, G. L.; CHARDULO, L. A.;  
986 NETO, O.; BALDASSINI, W. A.; OLIVEIRA, H.; CURI, R. Fine-scale estimation of  
987 inbreeding rates, runs of homozygosity and genome-wide heterozygosity levels in the  
988 Mangalarga Marchador horse breed. **Journal of Animal Breeding and Genetics**, v. 138,  
989 n. 2, p. 161–173, 19 mar. 2021.
- 990  
991 SARGOLZAEI, M.; CHESNAIS, J. P.; SCHENKEL, F. S. A New Approach for Efficient  
992 Genotype Imputation Using Information from Relatives. **BMC Genomics**, v. 15, n. 1, p.  
993 478, 2014.
- 994  
995 SARGOLZAEI, M.; SCHENKEL, F. S. QMSim: A large-scale genome simulator for  
996 livestock. **Bioinformatics**, v. 25, n. 5, p. 680–681, 2009.
- 997  
998 SAS Institute Inc., **SAS 9.1.3 Help and Documentation**, Cary, NC: SAS Institute Inc.,  
999 2002-2004.
- 1000  
1001 SCHALLER, M.; GONNET, P.; DRAPER, P.W.; CHALK, A.B.G.; BOWER, R.G.; WILLIS,  
1002 J.; HAUSAMMANN, L. **SWIFT: SPH With Inter-dependent Fine-grained Tasking** maio  
1003 2018.
- 1004  
1005 SELLI, A.; MILLER, S. P.; VENTURA, R. V. The Use of Interactive Visualizations for  
1006 Tracking Haplotypic Inheritance in Livestock. **Ruminants**, v. 4, n. 1, p. 90–111, 21 fev.  
1007 2024.



- 1008  
1009 SELLI, A.; VENTURA, R. V.; FONSECA, P. A. S.; BUZANSKAS, M. E.; ANDRIETTA, L.  
1010 T.; BALIEIRO, J. C. C.; BRITO, L. F. Detection and Visualization of Heterozygosity-Rich  
1011 Regions and Runs of Homozygosity in Worldwide Sheep Populations. **Animals**, v. 11, n.  
1012 9, p. 2696, 15 set. 2021.
- 1013  
1014 STELLA, A.; AJMONE-MARSAN, P.; LAZZARI, B.; BOETTCHER, P. Identification of  
1015 Selection Signatures in Cattle Breeds Selected for Dairy Production. **Genetics**, v. 185, n.  
1016 4, p. 1451–1461, 1 ago. 2010.
- 1017  
1018 TEO, Y. Y.; SMALL, K. S.; FRY, A. E.; WU, Y.; KWIATKOWSKI, D. P.; CLARK, T. G.  
1019 Power consequences of linkage disequilibrium variation between populations. **Genetic**  
1020 **Epidemiology**, v. 33, n. 2, p. 128–135, fev. 2009.
- 1021  
1022 TORO, M. A.; VARONA, L. A note on mate allocation for dominance handling in genomic  
1023 selection. **Genetics Selection Evolution**, v. 42, n. 1, p. 1–9, 2010.
- 1024  
1025 TULLO, E.; FINZI, A.; GUARINO, M. Review: Environmental Impact of Livestock Farming  
1026 and Precision Livestock Farming as a Mitigation Strategy. **Science of The Total**  
1027 **Environment**, v. 650, p. 2751–2760, fev. 2019.
- 1028  
1029 URBINATI, I.; STAFUZZA, N. B.; OLIVEIRA, M. T.; CHUD, T. C. S.; HIGA, R. H.;  
1030 REGITANO, L. C. de A.; DE ALENCAR, M. M.; BUZANSKAS, M. E.; MUNARI, D. P.  
1031 Selection signatures in Canchim beef cattle. **Journal of Animal Science and**  
1032 **Biotechnology**, v. 7, n. 1, p. 29, 5 dez. 2016.
- 1033  
1034 VAN ESSEN, D. C.; ANDERSON, C. H.; FELLEMAN, D. J. Information Processing in the  
1035 Primate Visual System: An Integrated Systems Perspective. **Science**, v. 255, n. 5043, p.  
1036 419–423, 24 jan. 1992.
- 1037  
1038 VAN NUFFEL, A.; ZWERTVAEGHER, I.; VAN WEYENBERG, S.; PASTELL, M.;  
1039 THORUP, V. M.; BAHR, C.; SONCK, B.; SAEYS, W. Lameness detection in dairy cows:  
1040 Part 2. Use of sensors to automatically register changes in locomotion or behavior.  
1041 **Animals**, v. 5, n. 3, p. 861–885, 2015.
- 1042  
1043 VAN ROSSUM, G.; DRAKE JR, F. L. **Python reference manual**. [s.l.] Centrum voor  
1044 Wiskunde en Informatica Amsterdam, 1995.
- 1045  
1046 VANRADEN, P. M. Efficient methods to compute genomic predictions. **Journal of Dairy**  
1047 **Science**, v. 91, n. 11, p. 4414–4423, 2008.
- 1048  
1049 VISSCHER, P. M.; GODDARD, M. E. From R.A. Fisher's 1918 Paper to GWAS a Century  
1050 Later. **Genetics**, v. 211, n. 4, p. 1125–1130, 1 abr. 2019.
- 1051  
1052 WALSH, J. B. Genomic selection signatures and animal breeding. **Journal of Animal**  
1053 **Breeding and Genetics**, v. 138, n. 1, p. 1–3, 12 jan. 2021.
- 1054

- 1055 WANG, S.; ZICKLER, D.; KLECKNER, N.; ZHANG, L. Meiotic crossover patterns:  
1056 Obligatory crossover, interference and homeostasis in a single process. **Cell Cycle**, v. 14,  
1057 n. 3, p. 305, 1 fev. 2015.
- 1058  
1059 WRIGHT, S. Coefficients of Inbreeding and Relationship. **The American Naturalist**, v. 56,  
1060 n. 645, p. 330–338, jul. 1922.
- 1061  
1062 YUAN, X.; MILLER, D. J.; ZHANG, J.; HERRINGTON, D.; WANG, Y. An overview of  
1063 population genetic data simulation. **Journal of Computational Biology**, v. 19, n. 1, p. 42–  
1064 54, 2012.
- 1065  
1066 ZACKS, J.; TVERSKY, B. Bars and Lines: A Study of Graphic Communication. **Memory &**  
1067 **Cognition**, v. 27, n. 6, p. 1073–1079, nov. 1999.
- 1068  
1069 ZHAO, H.; NETTLETON, D.; DEKKERS, J. C. M. Evaluation of linkage disequilibrium  
1070 measures between multi-allelic markers as predictors of linkage disequilibrium between  
1071 single nucleotide polymorphisms. **Genetics Research**, v. 89, n. 1, p. 1–6, fev. 2007.
- 1072  
1073 ZURAW, A.; AEFNER, F. Whole-Slide Imaging, Tissue Image Analysis, and Artificial  
1074 Intelligence in Veterinary Pathology: An Updated Introduction and Review. **Veterinary**  
1075 **Pathology**, v. 59, n. 1, p. 6–25, jan. 2022.

1076

1077 **3 INTEGRAÇÃO DE SIMULAÇÃO DE DADOS GENÔMICOS E FERRAMENTAS DE**  
1078 **ANÁLISES GENÔMICAS EM BOVINOS DE CORTE: UM PIPELINE PARA**  
1079 **ESTUDOS DE SELEÇÃO GENÉTICA E ACASALAMENTOS DIRIGIDOS COM**  
1080 **FOCO NA MENSURAÇÃO DE CORRIDAS DE HOMOZIGOSIDADE.**

1081 3.1 RESUMO

1082 Este trabalho teve como objetivo o desenvolvimento de um *pipeline* envolvendo a  
1083 simulação e validação de dados genômicos, a implementação de estratégias para avaliação  
1084 dos animais com fins de seleção e direcionamento de acasalamentos, e a comparação dos  
1085 resultados finais das diferentes estratégias por meio de visualizações interativas. Este  
1086 *pipeline* caracterizou-se pela integração de múltiplas linguagens de programação,  
1087 programas utilizados no melhoramento genético e funções criadas especificamente para a  
1088 aplicação das estratégias avaliadas. Diferentes métricas de endogamia e valores genéticos  
1089 foram utilizadas em cada estratégia, e seu impacto a longo prazo, em termos de marcas de  
1090 homozigidade no genoma, foi avaliado em nosso estudo. Para isso, uma população de  
1091 bovinos de corte foi simulada com o pacote R AlphaSimR, contendo 5 cromossomos com  
1092 diferentes tamanhos e números de *loci* de interesse econômico (QTL, do inglês *Quantitative*  
1093 *Trait Loci*), correspondentes a uma característica com herdabilidade de 0,3. A partir da  
1094 simulação inicial, foram testadas três estratégias de seleção e acasalamentos dirigidos. A  
1095 primeira estratégia teve como métricas os valores genéticos estimados apenas através do  
1096 pedigree e fenótipo de cada indivíduo, e o coeficiente de endogamia baseado em Pedigree  
1097 (EBV + Fped). A segunda estratégia fez o uso de informações de marcadores genéticos do  
1098 tipo SNP (*Single Nucleotide Polymorphism*) para o cálculo dos valores genéticos, e o  
1099 coeficiente de endogamia adotado foi o genômico (GEBV + Fg). A terceira estratégia foi  
1100 semelhante à segunda, no entanto fez o uso do coeficiente de endogamia baseado em  
1101 corridas de homozigidade (GEBV + Froh). Foram observados diferentes efeitos em  
1102 relação às métricas testadas. A estratégia GEBV + Froh produziu consistentemente valores  
1103 genéticos verdadeiros e fenótipos superiores. O Fg foi mais baixo quando utilizada a  
1104 estratégia GEBV + Fg, no entanto a mesma estratégia promoveu a formação de corridas  
1105 de homozigidade maiores e mais frequentes no genoma. A estratégia mais eficiente para  
1106 minimizar a porcentagem do genoma coberta por ROHs, no entanto, foi a EBV + Fped. As  
1107 avaliações genéticas de cada animal da população, assim como os respectivos coeficientes

1108 de endogamia, foram desenvolvidos via mescla de ativações em tempo real de softwares  
1109 já existentes (BLUF90, PLINK), por meio de uma interface desenvolvida em linguagem R.  
1110 Algumas limitações foram pontuadas, tais quais o tamanho populacional utilizado e a  
1111 adequação de certos parâmetros da simulação, e possíveis soluções foram propostas.  
1112 Espera-se, com este estudo, oferecer uma base para a integração de ferramentas já  
1113 utilizadas por estudos do melhoramento animal, bem como apresentar formas eficientes de  
1114 visualização destes dados, colaborando assim com o desenvolvimento de futuros estudos  
1115 na área de melhoramento animal.

## 1116 3.2 INTRODUÇÃO

1117 Estima-se que, nas próximas décadas, a população mundial irá crescer em 33%,  
1118 chegando a quase 10 bilhões de pessoas em 2050 (UNITED NATIONS, 2022). Este  
1119 crescimento populacional irá aumentar a demanda por alimentos em, pelo menos, 50% (DE  
1120 CLERCQ; VATS; BIEL, 2018). Dentre as estratégias exploradas a fim de suprir tal  
1121 demanda, o aumento da produtividade agrícola destaca-se, uma vez que também tende a  
1122 reduzir o impacto ambiental e o uso de terras. O melhoramento genético, por sua vez,  
1123 corresponde a um dos pilares que sustentam o avanço desta produtividade.

1124 Ao buscar o progresso genético dos animais de produção, nos deparamos com um  
1125 entrave já conhecido no melhoramento genético: a dificuldade em manter-se o nível de  
1126 endogamia de uma população a níveis considerados aceitáveis (HOWARD et al., 2017).  
1127 Isto se deve ao fato de que a seleção de animais determinados como geneticamente  
1128 superiores tende a ocasionar o acasalamento de indivíduos mais aparentados que a média  
1129 da população, gerando endogamia nas gerações subsequentes (CLARK et al., 2013;  
1130 DAETWYLER et al., 2007). A endogamia, por sua vez, pode levar ao aparecimento em  
1131 maior frequência de doenças recessivas, ou ainda na piora de performance e/ou vigor do  
1132 indivíduo, fenômeno conhecido como depressão endogâmica (BOURDON, 2014);  
1133 DICKERSON, 1963). Assim, é crucial a elaboração de estratégias que permitam o ganho  
1134 genético sem que haja uma grande elevação no nível de endogamia de uma população.

1135 O coeficiente de endogamia ( $F$ ) corresponde à probabilidade de que dois alelos em  
1136 loci homólogos sejam idênticos por descendência (IBD, do inglês Identical by Descent), ou

1137 seja, herdados de um mesmo ancestral (WRIGHT, 1922). No entanto, este método não leva  
1138 em consideração a segregação mendeliana, ou seja, não identifica quais alelos foram  
1139 herdados de fato pelo indivíduo em questão. As corridas de homoziguidade (do inglês,  
1140 Runs of Homozygosity - ROH) correspondem a cópias idênticas de haplótipos que são  
1141 herdadas de um mesmo indivíduo (CEBALLOS et al., 2018), e são identificadas por meio  
1142 de análises de marcadores moleculares do tipo SNP (do inglês, Single Nucleotide  
1143 Polymorphism). Assim, ao analisar as ROH é possível não apenas calcular o coeficiente de  
1144 endogamia do indivíduo, mas também identificar quais regiões genômicas encontram-se  
1145 em homoziguidade, permitindo um maior grau de detalhamento na análise de cada  
1146 indivíduo.

1147 Uma vez definidos os critérios de seleção, a execução dos acasalamentos pode ser  
1148 feita de maneira aleatória ou com acasalamentos dirigidos, que consiste no processo de  
1149 alocação de machos e fêmeas em pares otimizados (ALLAIRE, 1980), visando o aumento  
1150 do ganho genético e restrição da endogamia de forma mais eficiente. Neste processo, é  
1151 necessário estimar os resultados esperados das progênes para cada acasalamento  
1152 possível, seja através do cálculo da média dos pais (NEVES et al., 2009), ou através da  
1153 criação de "progênes hipotéticas", que consiste na expansão do pedigree com os  
1154 acasalamentos em potencial ou na combinação dos genótipos destes. Mesmo com a  
1155 inclusão dos genótipos, no entanto, existe uma limitação técnica na predição dos resultados  
1156 dos acasalamentos, uma vez que a segregação mendeliana corresponde a 50% das  
1157 diferenças genéticas interindividuais nos valores genéticos (DAETWYLER et al., 2013).

1158 A simulação de dados é uma importante ferramenta para a investigação dos efeitos  
1159 de longo prazo da seleção genômica, uma vez que permite o teste de uma variedade de  
1160 hipóteses a baixo custo, de forma rápida e replicável (DAETWYLER et al., 2013; GAYNOR;  
1161 GORJANC; HICKEY; 2021). No entanto, deve-se notar que a simulação de genomas e  
1162 estruturas populacionais, tanto em animais quanto em plantas, é uma tarefa complexa, e  
1163 não existe um modelo ideal para a simulação de dados (DAETWYLER et al., 2013). Dentre  
1164 diversos programas de simulação genômica desenvolvidos (HICKEY; GORJANC, 2012;  
1165 HOWARD et al., 2017b; SARGOLZAEI; SCHENKEL, 2009), destaca-se a implementação  
1166 em R do Software AlphaSim (FAUX et al., 2016), AlphaSimR (GAYNOR; GORJANC;  
1167 HICKEY, 2021). Este pacote R permite um alto grau de flexibilidade na simulação de

1168 populações animais e vegetais, apresentando funções que correspondem a operações de  
1169 um programa de melhoramento, como "cruzamento" e "seleção", além de possibilitar a  
1170 integração com outras funcionalidades do R tais quais a elaboração de gráficos, loops, e o  
1171 armazenamento de informações em objetos. Através da linguagem de programação R,  
1172 ainda é possível a execução de programas externos que são frequentemente utilizados em  
1173 esquemas de avaliação genômica, como o PLINK (PURCELL et al., 2007) e o BLUPF90  
1174 (MISZTAL et al., 2014).

1175 De forma geral, após executado o processo de simulação e testados os diferentes  
1176 métodos de seleção e acasalamentos dirigidos, é necessário avaliar os resultados obtidos.  
1177 Além da aplicação dos métodos estatísticos, uma abordagem bastante eficiente  
1178 corresponde à visualização dos dados. Cole e VanRaden (2010) demonstraram como o  
1179 uso de gráficos bem elaborados para a apresentação de resultados de avaliações  
1180 genômicas é capaz de reunir uma alta densidade de informações e ainda evidenciar  
1181 sutilezas de uma forma como não é possível através do uso de tabelas e textos descritivos.  
1182 Mais recentemente, Durant e colaboradores (2022) propuseram "10 regras simples para o  
1183 desenvolvimento de ferramentas de visualização na genômica", abordando desde a  
1184 conceituação de novas ferramentas até o compartilhamento do resultado final e eventuais  
1185 atualizações.

1186 No presente trabalho, buscou-se o desenvolvimento de um pipeline completo que  
1187 envolve a simulação e validação de dados genômicos, aplicação de diferentes estratégias  
1188 para avaliação dos animais com fins de seleção e direcionamento de acasalamentos,  
1189 execução de tais acasalamentos em uma nova rodada de simulação de dados (em um  
1190 processo automatizado por diversas gerações), e finalmente o desenvolvimento de  
1191 visualizações que permitem a comparação dos resultados finais das diferentes estratégias.  
1192 Por meio da execução de tal pipeline, foram avaliados os efeitos da inclusão das ROH em  
1193 um processo de acasalamentos dirigidos para fins de seleção genômica de bovinos,  
1194 monitorando os valores genéticos e níveis de endogamia da população ao longo de  
1195 diversas gerações. Tal abordagem foi viabilizada através do uso de ferramentas de código  
1196 aberto e pode servir como ponto de partida para estudos de maior complexidade no futuro.

### 1197 3.3 MATERIAL E MÉTODOS

### 3.3.1 Simulação inicial de dados

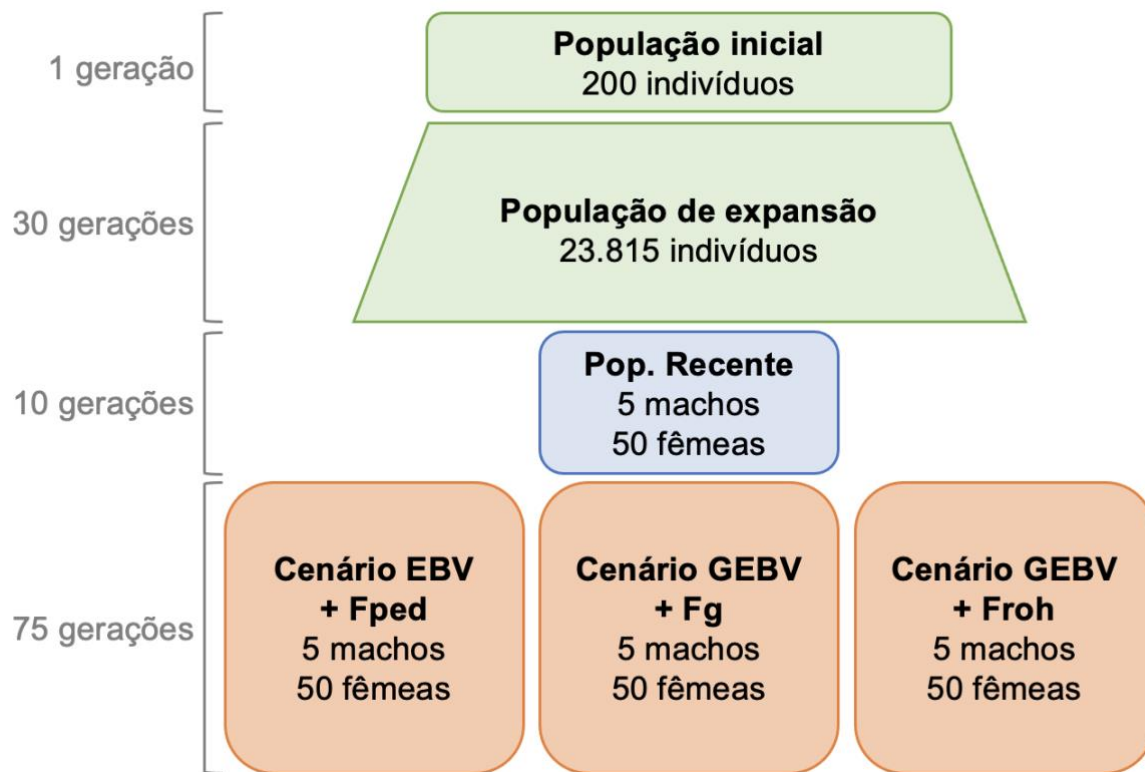
1198

1199 Os scripts desenvolvidos para a simulação de dados, bem como para o *pipeline*  
1200 descrito na seção 3.3.5. estão disponíveis através do link  
1201 [https://github.com/alanaselli/thesis\\_pipeline](https://github.com/alanaselli/thesis_pipeline).

1202 A fim de desenvolver-se uma população de bovinos de corte similar àquelas  
1203 existentes atualmente, o pacote AlphaSimR (GAYNOR; GORJANC; HICKEY, 2021) foi  
1204 implementado via linguagem R (R CORE TEAM, 2020). Os argumentos adotados em uma  
1205 primeira tentativa de desenvolvimento da simulação foram similares aos já empregados em  
1206 outros trabalhos desenvolvidos pelo grupo (SELLI et al., 2021, p. 231), e posteriormente  
1207 alterados entre uma tentativa e outra a fim de obter-se genótipos semelhantes aos de  
1208 bovinos de corte. A validação inicial de tais genótipos foi efetuada utilizando-se sempre o  
1209 decaimento de LD para comparação. Foram realizadas 5 simulações independentes a fim  
1210 de se verificar a consistência dos resultados finais deste estudo, evitando possíveis vieses  
1211 resultantes do processo estocástico de simulação.

1212 A função runMacs2 foi utilizada para gerar "haplótipos fundadores" (conforme  
1213 descrito por Gaynor, Gorjanc e Hickey, 2021), com os seguintes argumentos: nInd = 200,  
1214 Ne = 200, nChr = 5, histNe = c(1000, 2000, 4000, 7500, 10000), e histGen = c(50, 100, 500,  
1215 1000, 1500). Uma característica de efeitos aditivos foi simulada, com herdabilidade de 0,3,  
1216 contendo 100, 75, 50, 10 e 0 QTL nos cromossomos de 1 a 5, respectivamente, afim de  
1217 verificar-se como a presença de quantidades diferentes de QTL poderia afetar os resultados  
1218 observados. Também foi simulado um painel de marcadores do tipo Single Nucleotide  
1219 Polymorphisms (SNPs) com 3000, 2500, 2200, 2000, 1900 marcadores nos cromossomos  
1220 de 1 a 5, respectivamente. O esquema de gerações e populações adotado seguiu a  
1221 estrutura descrita por Brito et al. (2011), e pode ser visualizado na Figura 1.

Figura 3-1 – Estrutura populacional de uma simulação de bovinos de corte.



Fonte: Autoria própria.

1222

1223 A partir da população inicial com 200 indivíduos, iniciou-se a população de  
 1224 expansão. Na primeira geração, 100 machos e 100 fêmeas foram acasalados  
 1225 aleatoriamente em um total de 100 acasalamentos (cada indivíduo foi acasalado uma vez),  
 1226 e a cada geração o número de acasalamentos foi multiplicado por 1,2, por 30 gerações. Na  
 1227 geração 31, foram geradas 23.815 progênie, e a partir destas iniciou-se a população  
 1228 recente.

1229 A população recente foi gerada a partir da última geração da população de  
 1230 expansão. Os 5 machos e 50 fêmeas com maior fenótipo foram selecionados, e a cada  
 1231 indivíduo foi atribuído um ano de nascimento. Aos 3 melhores machos foi atribuído o ano  
 1232 atual (G) -1, e aos outros 2 machos, G - 2. O mesmo processo foi repetido com as fêmeas,  
 1233 sendo os grupos com os seguintes tamanhos: 20, 15, 10 e 5. Os anos atribuídos foram de  
 1234 G-1 a G-4. A partir de 50 acasalamentos aleatórios entre estes animais, onde cada fêmea  
 1235 foi acasalada apenas uma vez, foi gerada a geração atual, com 25 machos e 25 fêmeas.  
 1236 Os grupos mais antigos de machos (G-2) e fêmeas (G-4) foram descartados, e um novo



1237 processo de seleção foi iniciado. Entre os indivíduos da geração atual G, 20 fêmeas e 3  
 1238 machos foram selecionados com base no fenótipo, e cada um dos demais grupos passou  
 1239 pelo mesmo processo de seleção, considerando as respectivas taxas de descarte conforme  
 1240 ilustrado na Figura 2. Assim, obteve-se novamente um total de 50 fêmeas e 5 machos, que  
 1241 foram acasalados aleatoriamente para compor a próxima geração. Este processo repetiu-  
 1242 se por 10 gerações.

1243 Figura 3-2 – Número de machos e fêmeas selecionados a cada geração nos diferentes grupos.

	Animais por geração		Taxa de descarte	
	Fêmeas	Machos	Fêmeas	Machos
<b>Geração G</b>	25 fêmeas	25 machos	20% fêmeas	88% machos
<b>Geração G-1</b>	20 fêmeas	3 machos	25% fêmeas	33% machos
<b>Geração G-2</b>	15 fêmeas	2 machos	33% fêmeas	100% machos
<b>Geração G-3</b>	10 fêmeas		50% fêmeas	
<b>Geração G-4</b>	5 fêmeas		100% fêmeas	
<b>Total selecionado</b>	<b>50 fêmeas</b>	<b>5 machos</b>		

1244

1245 Fonte: Autoria própria.

1246 A última geração da população recente, incluindo os indivíduos dos grupos G-1, G-  
 1247 2 e G-3, tiveram seus genótipos exportados em formato PLINK (.ped e .map), para validar  
 1248 a simulação e dar início às "populações experimentais", etapas que serão descritas a  
 1249 seguir. O arquivo .map gerado pela função writePlink, por padrão, apresenta as posições  
 1250 dos marcadores em cM (de 0 a 10). Assim, este arquivo foi modificado para apresentar as  
 1251 posições dos marcadores simulados em pares de bases, de forma que outros programas  
 1252 pudessem utilizar posteriormente os arquivos gerados. Assim, as distâncias em cM foram  
 1253 divididas por 10, resultando em valores de 0 a 1 que representam as posições dos

1254 marcadores ao longo de cada cromossomo. Tais valores foram então multiplicados pelos  
1255 respectivos comprimentos de cada cromossomo em pares de bases, obtendo-se assim a  
1256 posição de cada marcador em pares de bases.

1257 A seguir, iniciou-se a simulação das populações experimentais, onde foram  
1258 testados 3 cenários de seleção e acasalamentos dirigidos por 75 gerações. Cada uma das  
1259 3 populações experimentais teve início a partir de uma mesma população, a última geração  
1260 da população recente. No primeiro cenário, não foram utilizadas informações genômicas.  
1261 O EBV foi calculado apenas com informações de fenótipo e pedigree, e o coeficiente de  
1262 endogamia baseou-se no pedigree ( $F_{ped}$ ). O segundo cenário incluiu dados de marcadores  
1263 genéticos do tipo SNP para o cálculo do GEBV e do coeficiente de endogamia genômico  
1264 ( $F_g$ ), além de incluir as informações de fenótipos e pedigree. O terceiro cenário diferenciou-  
1265 se do segundo apenas por substituir o  $F_g$  pelo  $F_{roh}$ . As próximas seções descrevem as  
1266 análises e pipelines utilizados nesta última etapa da simulação de dados.

### 1267 **3.3.2 Progênies Hipotéticas com Ausência de Segregação Mendeliana** 1268 **(PHASM)**

1269 Com o intuito de se obter as predições de endogamia e valor genético de cada  
1270 combinação possível entre machos e fêmeas, porém sem levar em conta os efeitos de  
1271 segregação mendeliana ou recombinação gênica, foram geradas as Progênies Hipotéticas  
1272 com Ausência de Segregação Mendeliana (PHASM). A etapa de seleção e direcionamento  
1273 de acasalamentos utilizou métricas tanto dos indivíduos candidatos quanto de suas  
1274 PHASM. Assim buscou-se prever, dentre os animais pré-selecionados, quais combinações  
1275 iriam gerar as melhores progênies.

1276 No primeiro cenário, que não envolve informações genômicas, as PHASM foram  
1277 geradas como uma extensão do pedigree: elas foram inseridas no arquivo de pedigree com  
1278 as informações de ID ('sire\_dam'), pai ('sire') e mãe ('dam'). No arquivo de informações para  
1279 o BLUPF90 (ranf90.dat), as colunas de sexo e fenótipo foram representadas com o 0 (valor  
1280 faltante).

1281           Nos outros dois cenários, a criação das PHASM incluiu, além da geração do  
1282 pedigree, a combinação dos genótipos dos pais. Assim, cada dupla gerou 4 combinações  
1283 genotípicas diferentes, uma vez que cada indivíduo apresenta duas fitas de DNA. É  
1284 importante ressaltar que não foram considerados os efeitos da segregação mendeliana e  
1285 nem de crossover, devido à complexidade computacional envolvida em simular ou tentar  
1286 replicar tais efeitos. Assim como no cenário sem genômica, as informações de sexo e  
1287 fenótipo das PHASM foram consideradas como dados faltantes.

### 1288           **3.3.3 Avaliações genômicas**

1289           As avaliações genômicas geradas para a última etapa da simulação foram  
1290 desenvolvidas através da família de programas BLUPF90. Um modelo animal considerando  
1291 apenas uma característica foi adotado, uma vez que não foram simulados efeitos fixos nem  
1292 outras características. O pedigree utilizado foi mantido a uma profundidade de 10 gerações,  
1293 para evitar o aumento da complexidade computacional ao passar das gerações. A variância  
1294 residual (porcentagem da variação da característica que não se deve a fatores genéticos)  
1295 foi definida em 0,7. A opção thrStopCorAG -1 foi utilizada para evitar que o programa  
1296 parasse ao encontrar uma baixa correlação entre as matrizes A e G, passo que foi  
1297 necessário durante o desenvolvimento do pipeline quando um número muito pequeno de  
1298 animais foi utilizado. A opção use\_yams foi adotada para que o Software rodasse mais  
1299 rápido, o que se mostrou necessário quando foram incluídas as PHASM. Os mesmos  
1300 componentes de variância foram utilizados a cada geração, e em todos os cenários.

1301           Dois cartões de parâmetros com as informações descritas acima foram criados,  
1302 sendo que um deles inclui dados genômicos, e o outro não. No cartão que inclui  
1303 informações genômicas, as opções saveDiagGOrig e origID foram utilizadas para se obter  
1304 o Fg dos animais. Os cartões de parâmetro foram utilizados no programa RENUMF90, que  
1305 então gerou um novo cartão, de acordo com as especificações passadas em cada cenário  
1306 (com ou sem genômica), que foi utilizado no programa BLUPF90. Assim, obteve-se o Fped  
1307 e EBV (sem genômica) ou Fped, Fg e GEBV (com genômica) para os candidatos de cada  
1308 geração das populações experimentais.

1309

### 1310 3.3.4 Corridas de homozigosidade

1311 Para a última geração da população recente, bem como para cada geração das  
 1312 populações experimentais, as ROH foram detectadas. A detecção das ROH foi realizada  
 1313 utilizando-se o pacote R detectRUNS (BISCARINI et al., 2018), com o método 'consecutive  
 1314 runs'. Os argumentos adotados foram: número mínimo de SNPs igual a 20, comprimento  
 1315 mínimo de ROH igual a 1.000 kb, máxima distância entre SNPs em uma ROH igual a 500  
 1316 kb, e número máximo de heterozigotos e/ou marcadores perdidos igual a 1. Foram obtidas  
 1317 as posições de começo e fim de cada corrida, comprimento e número de marcadores. A  
 1318 frequência de cada SNP em ROH também foi registrada.

1319 O FROH para cada indivíduo analisado foi obtido também através do pacote  
 1320 detectRUNS, que implementa a seguinte fórmula:

$$1321 F_{ROH} = \frac{\Sigma L_{ROH}}{L_{genoma}}$$

1322 Onde  $\Sigma L_{ROH}$  é a soma do comprimento de todas as ROH detectadas no indivíduo,  
 1323 e  $L_{genoma}$  é o comprimento total do genoma, dado pela soma das distância entre os  
 1324 marcadores extremos de cada cromossomo.  
 1325

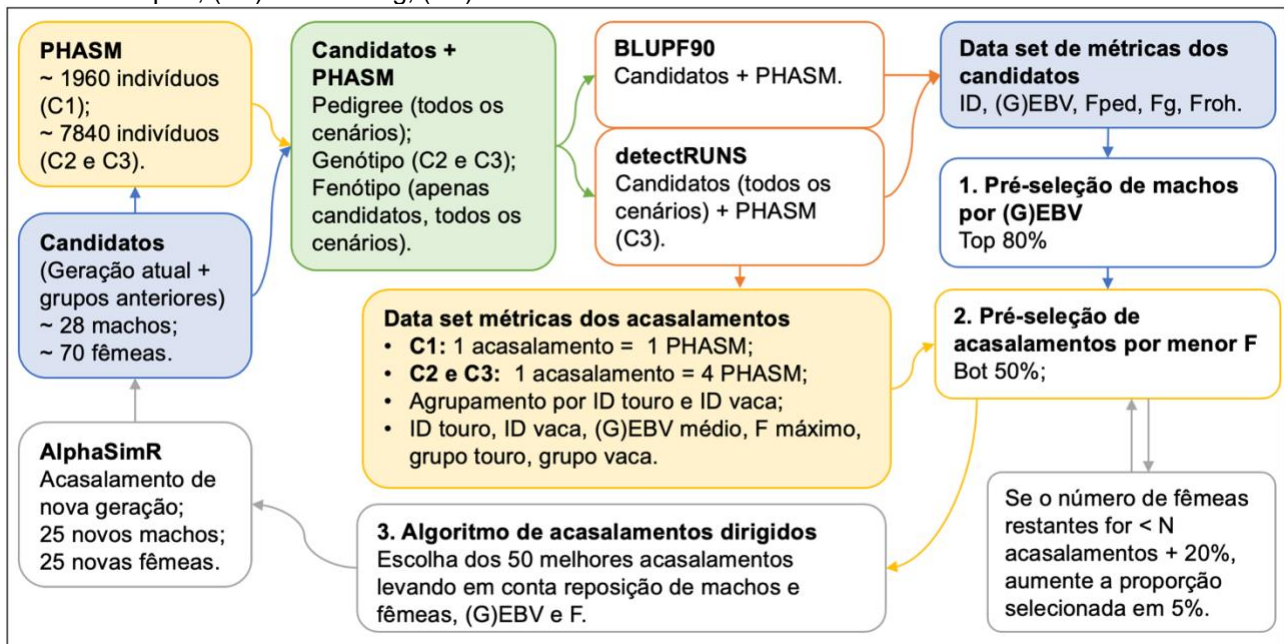
1326 Na última geração da população recente (replicata 1), informações como número e  
 1327 tamanho médio de ROHs por indivíduo, tamanho máximo de ROH e porcentagem de ROHs  
 1328 em cada cromossomo, além do número e proporção de ROHs, número e proporção de  
 1329 indivíduos e Froh médio para as diferentes classes de tamanho de ROH, foram organizadas  
 1330 em tabelas. Também foram criados gráficos com as posições das corridas de cada  
 1331 indivíduo da população para os 5 cromossomos, e um gráfico com a proporção de  
 1332 marcadores em ROHs nos 5 cromossomos para a população. Estes passos foram  
 1333 executados no R, com os pacotes detectRUNS (BISCARINI et al., 2018), dplyr (WICKHAM  
 1334 et al., 2023), e ggplot2 (WICKHAM, 2016).

1335

### 3.3.5 Pipeline de análises e seleção

Para cada um dos 3 cenários de seleção e acasalamentos dirigidos, foram desenvolvidos diferentes pipelines com os programas necessários para cada etapa, porém mantendo-se a mesma metodologia nos 3 cenários (Figura 3). Na primeira geração de cada cenário, haviam 98 indivíduos provenientes da população recente (25 machos e 25 fêmeas da última geração, 3 machos e 45 fêmeas de gerações anteriores, conforme o esquema representado na Figura 2) - exceto pelos grupos de descarte. A partir destes indivíduos, foram geradas as PHASM, que corresponderam a 1960 indivíduos no primeiro cenário, e 7840 nos demais cenários.

Figura 3-3 – Esquema de seleção e direcionamento de acasalamentos para os três cenários: (C1) EBV + Fped, (C2) GEBV + Fg, (C3) GEBV + Froh.



Fonte: Autoria própria.

Os dados dos candidatos e das PHASM foram utilizados em conjunto para gerar previsões através do BLUPF90. Assim, foram obtidos os valores genéticos estimados e coeficientes de endogamia Fped e Fg para todos os indivíduos. O Froh dos candidatos também foi obtido através do pacote detectRUNS, mesmo nos cenários 1 e 2, para fins de comparação dos cenários ao final do processo.

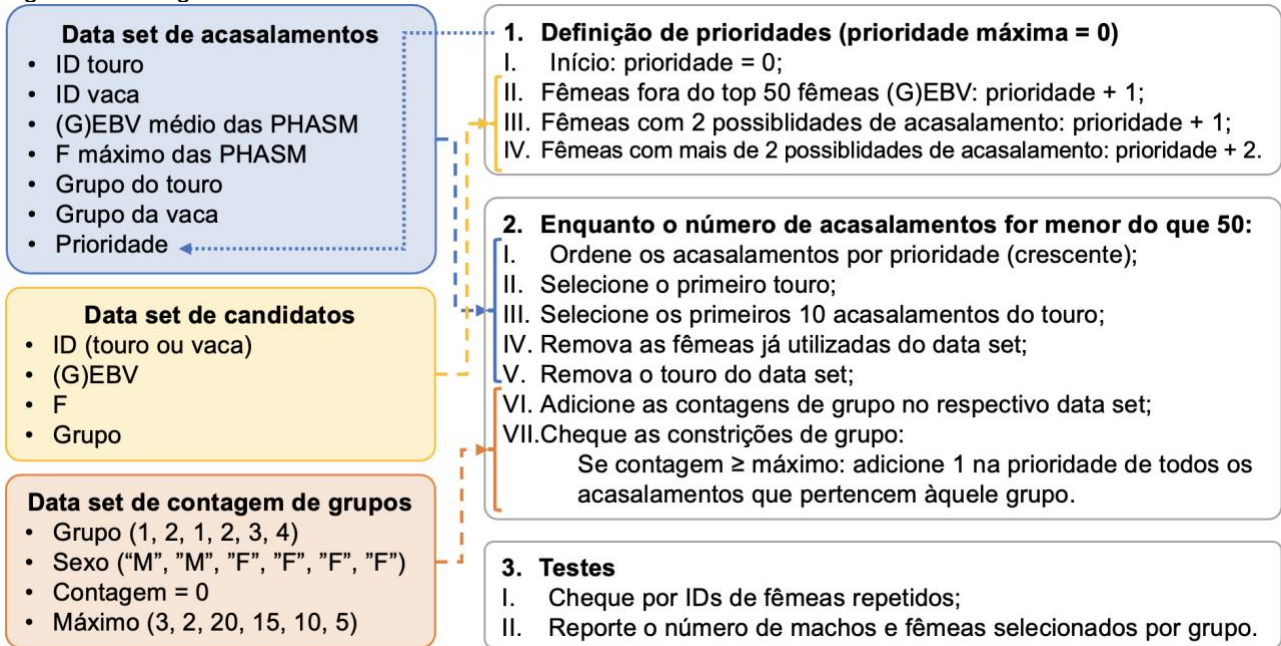
1354 O processo de seleção inicia-se com a exclusão dos 20% candidatos machos com  
1355 os menores (G)EBVs, bem como suas PHASM. Assim, foram mantidos 24 machos e seus  
1356 respectivos acasalamentos. Em seguida, apenas os dados das PHASM foram utilizados no  
1357 processo. Estes dados consistem em ID e grupo do pai e da mãe, (G)EBV e coeficiente de  
1358 endogamia estimados para as PHASM. Assim, são excluídos 50% dos acasalamentos  
1359 (PHASM) com o maior F.

1360 Devido à possibilidade de flutuação no número de acasalamentos gerados de uma  
1361 geração para a outra (que será explicado posteriormente), foi necessário adicionar uma  
1362 etapa onde verifica-se se o número de acasalamentos restantes é igual ou superior ao  
1363 número de acasalamentos necessários mais 20%. Caso esta condição não tenha sido  
1364 atingida, a proporção dos acasalamentos selecionados é aumentada em 5%, até que se  
1365 obtenha o número mínimo necessário. A necessidade de se ter 20% a mais de  
1366 acasalamentos se deve ao fato de que ao selecionar os primeiros acasalamentos, outros  
1367 tornam-se inviáveis - seja por restrições de grupos animais, ou porque uma mesma vaca  
1368 aparece em mais de um acasalamento, apesar de só poder ser acasalada uma vez.  
1369 Algumas vacas só aparecem em um acasalamento, e caso o macho correspondente já  
1370 tenha sido selecionado com outras vacas, este acasalamento também torna-se inválido.

1371 Após excluídos os acasalamentos que geram maior endogamia, inicia-se o  
1372 processo de escolha de quais acasalamentos serão de fato implementados e mantidos na  
1373 próxima geração, conforme ilustrado na Figura 4. Uma nova coluna é adicionada no dataset  
1374 de acasalamentos, com o objetivo de definir a prioridade dos mesmos. Quanto menor for o  
1375 valor, maior é a prioridade do acasalamento. No início, todos os acasalamentos recebem a  
1376 prioridade 0. Em seguida, os acasalamentos ranqueados fora dos top 50 para (G)EBV  
1377 recebem 1 ponto. As fêmeas que apresentam 2 possibilidades de acasalamento recebem  
1378 mais um ponto em seus acasalamentos, e aquelas com mais de 2 possibilidades, recebem  
1379 mais 2.

1380

1381 Figura 3-4 – Algoritmo de direcionamento de acasalamentos.



1382

1383 Fonte: Autoria própria.

1384 A seguir, o dataset de acasalamentos é ordenado pela coluna de prioridade, e o  
 1385 touro presente no primeiro acasalamento é selecionado. Os 10 primeiros acasalamentos  
 1386 deste touro são alocados para um novo dataset, "matings". Todos os outros acasalamentos  
 1387 que contêm aquele touro e vacas presentes nos acasalamentos selecionados são excluídos  
 1388 do dataset de acasalamentos. Um dataset contendo a contagem de indivíduos em cada  
 1389 grupo de machos e fêmeas é atualizado, e caso a contagem em algum dos grupos exceda  
 1390 o valor estabelecido, mais um ponto é adicionado à coluna de prioridade para os indivíduos  
 1391 daquele grupo. Desta forma, evita-se o problema de ficar sem acasalamentos disponíveis,  
 1392 que poderia ocorrer caso todos os acasalamentos que não atendem às restrições de grupos  
 1393 fossem excluídos. Este processo é repetido até que se obtenha um número igual ou  
 1394 superior a 50 acasalamentos no dataset matings.

1395 Após concluída a etapa de direcionamento de acasalamentos, alguns testes são  
 1396 executados para se evitar que hajam fêmeas repetidas, e o número de machos e fêmeas  
 1397 selecionados em cada grupo é reportado. Os acasalamentos são então executados, e os  
 1398 novos indivíduos gerados (25 machos e 25 fêmeas) têm seus genótipos e fenótipos  
 1399 registrados. Estes indivíduos, juntamente com os grupos anteriores (exceto pelo grupo de

1400 descarte) iniciam a próxima geração, em um novo ciclo de seleção e direcionamento de  
1401 acasalamentos.

1402 É possível que o número de indivíduos que iniciam uma geração apresente variação  
1403 de uma geração para a outra, uma vez que o algoritmo de direcionamento de  
1404 acasalamentos não é exato ao selecionar os acasalamentos disponíveis para cada macho.  
1405 Caso o primeiro macho selecionado apresente um número de acasalamentos disponíveis  
1406 inferior a 10, torna-se necessário selecionar um número superior a 5 machos para se obter  
1407 os 50 acasalamentos desejados. Assim, ao selecionar o 6º macho, novamente busca-se  
1408 pelos primeiros 10 acasalamentos, e não pelo número necessário para chegar a 50, o que  
1409 pode fazer com que um número superior a 50 acasalamentos seja gerado naquela geração.

### 1410 **3.3.6 Validação da simulação inicial**

1411 Os genótipos em formato PLINK obtidos das últimas gerações da população  
1412 recente (replicata 1) foram utilizados para o cálculo da Frequência Alélica Menor (MAF, do  
1413 inglês Minor Allele Frequency) e do decaimento de desequilíbrio de ligação, através do  
1414 Software PLINK V1.9 (PURCELL et al., 2007). Para o cálculo da MAF o argumento --freq  
1415 foi utilizado, e para o cálculo de LD os seguintes argumentos foram adotados: --ld-window-  
1416  $r^2$  igual a 0, --ld-window igual a 100.000 e --ld-window-kb igual a 1.000.000.

1417 A partir dos dados gerados, algumas métricas de MAF e LD foram obtidas, como  
1418 média, mediana e desvio padrão. Para cada cromossomo, obteve-se o tamanho do  
1419 cromossomo (dado pela posição do seu último marcador), o número de marcadores  
1420 presentes no cromossomo, a distância média e o desvio padrão em kb entre pares de  
1421 marcadores adjacentes, e o  $r^2$  médio, o desvio padrão e a mediana entre os pares de  
1422 marcadores adjacentes.

1423 Também foram geradas tabelas contendo o número de pares de marcadores, a  
1424 média, o desvio padrão e a mediana de  $r^2$ , e a porcentagem de pares de marcadores com  
1425  $r^2$  acima de 0.01, 0.03 e 0.05 para diferentes distâncias entre os pares de marcadores. Uma  
1426 tabela foi gerada incluindo os marcadores de todos os cromossomos simultaneamente, e  
1427 outras tabelas foram geradas individualmente para cada cromossomo. Por fim, o  $r^2$  médio



1428 para pares de marcadores separados por diferentes distâncias foi plotado em função de  
1429 suas distâncias, para cada cromossomo, representando o decaimento do desequilíbrio de  
1430 ligação para esta população. A manipulação de dados e a criação de gráficos e tabelas  
1431 foram executadas no Software R, com os pacotes dplyr (WICKHAM et al., 2023), tidyr  
1432 (WICKHAM, VAUGHAN & GIRLICH, 2023) e ggplot2 (WICKHAM, 2016).

### 1433 **3.3.7 Avaliação dos cenários**

1434 A cada geração dos cenários de seleção e direcionamento de acasalamentos,  
1435 foram obtidos os fenótipos, valores genéticos verdadeiros, GEBVs calculados pelo  
1436 AlphaSimR, (G)EBVs calculados pelo BLUPF90 (os valores genéticos estimados pelo  
1437 BLUPF90 no primeiro cenário não incluíram dados genômicos), Fped, Fg, e Froh. No  
1438 primeiro cenário, os dados das PHASM incluíram apenas EBVs (BLUPF90) e Fped. No  
1439 segundo cenário, as PHASM apresentaram genótipos, e as informações obtidas foram  
1440 GEBVs (BLUPF90), Fped e Fg. No terceiro cenário, além dos GEBVs (BLUPF90), Fped e  
1441 Fg, também foram obtidos os Froh das PHASM.

1442 Com o objetivo de verificar-se a evolução das métricas avaliadas em cada cenário,  
1443 as médias das novas 50 progênies de cada geração foram plotadas em gráficos de linha,  
1444 onde cada linha corresponde a um cenário. Foi gerado um gráfico para cada métrica e  
1445 replicata da simulação, a fim de se averiguar possíveis vieses de simulação sobre os  
1446 resultados observados em cada métrica.

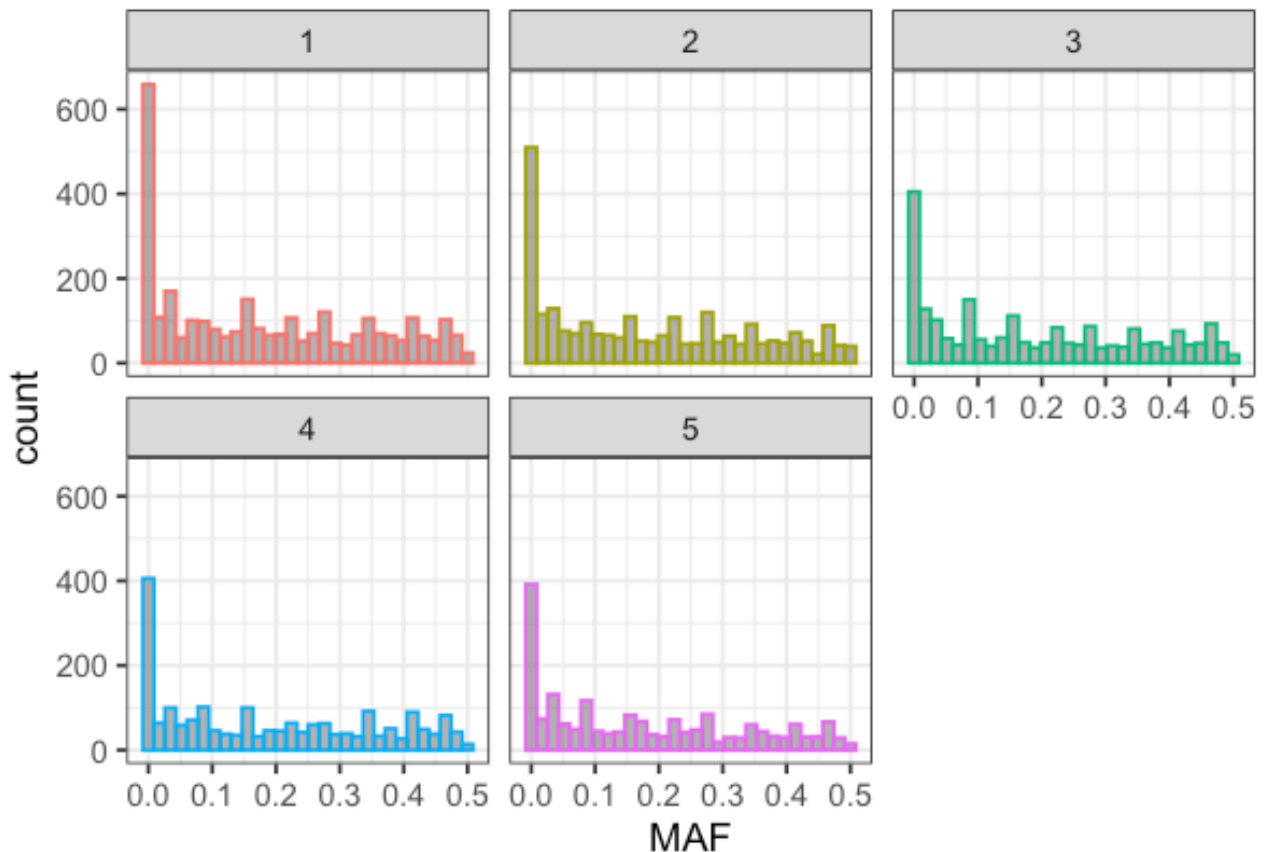
1447 Também foram contrastadas algumas das métricas das PHASM e reais nos  
1448 diferentes cenários, para verificar se as predições feitas sobre as PHASM se aproximaram  
1449 dos resultados obtidos com os acasalamentos dirigidos escolhidos. Por fim, a cada geração  
1450 foi registrado o número de indivíduos com genótipo, fenótipo e pedigree que compuseram  
1451 a população de treino das predições genômicas, e o número de PHASM que compuseram  
1452 a população teste, contendo apenas genótipos e pedigree. Estas informações foram  
1453 utilizadas na criação de um gráfico, que mostra o número de animais em cada população  
1454 ao longo das gerações. A manipulação de dados e a criação de gráficos foram executadas  
1455 no Software R, com os pacotes dplyr (WICKHAM et al., 2023), tidyr (WICKHAM, VAUGHAN  
1456 & GIRLICH, 2023) e ggplot2 (WICKHAM, 2016).

## 1457 3.4 RESULTADOS E DISCUSSÃO

1458 **3.4.1 Simulação e validação de dados**

1459 A partir da última geração da população recente, a MAF foi calculada. A MAF média  
 1460 obtida foi de 0,18, com desvio padrão de 0,16 e mediana igual a 0,16. Entre os 11.600  
 1461 marcadores simulados, 2372 (20%) apresentaram MAF igual a 0, o que pode ser devido ao  
 1462 baixo  $N_e$  da população recente. A Figura 5 apresenta a distribuição dos marcadores em  
 1463 relação à MAF. Estes resultados demonstram uma maior fixação dos alelos em comparação  
 1464 com os resultados obtidos por Neves et al. (2014), que reportaram MAF média e desvio  
 1465 padrão de 0,226 e 0,144 para uma população de gado Nelore ( $N=685$ ) genotipado com  
 1466 aproximadamente 300 mil marcadores do tipo SNP.

1467 Figura 3-5 – Histograma de Minor Allele Frequencies (MAF) nos 5 cromossomos de indivíduos de uma  
 1468 população bovina simulada.



1469

1470 Fonte: Autoria própria.

1471 A partir dos dados genotípicos obtidos da população recente, foi calculado o  $r^2$  entre  
 1472 todos os pares de marcadores moleculares pertencentes a um mesmo cromossomo. Pérez  
 1473 O'Brien et al. (2014) calcularam o LD ( $r^2$ ) médio a diferentes distâncias em bovinos taurinos  
 1474 e zebuínos, utilizando um painel HD com 777 mil marcadores e dois subconjuntos deste  
 1475 painel, com 50 mil e 259 mil marcadores cada. Os autores reportaram um comportamento  
 1476 errático a curtas distâncias (até 20 kb) quando utilizou-se o painel 50K, devido ao baixo  
 1477 número de pares de marcadores a tais distâncias. Este comportamento também foi  
 1478 observado em relação ao painel simulado, que buscou mimetizar um painel de mesma  
 1479 densidade.

1480 A tabela 1 mostra estatísticas obtidas para cada cromossomo, como tamanho em  
 1481 MB, número de pares de marcadores adjacentes, distância média e desvio padrão entre  
 1482 marcadores adjacentes, bem como média, desvio padrão e mediana do  $r^2$  entre marcadores  
 1483 adjacentes. A distância média variou de 62,3 a 78,3 Mb, próximo ao reportado por Lu et al.  
 1484 (2012), que foi de aproximadamente 70 KB em todos os cromossomos, ao utilizar um painel  
 1485 com 51.620 marcadores. Os  $r^2$  médios obtidos para os cromossomos de 1 a 5 variaram  
 1486 entre 0,31 e 0,36, com alto desvio padrão (de 0,34 a 0,37). Apesar da variação sutil do  $r^2$ ,  
 1487 este resultado vai ao encontro dos números de QTL simulados para cada cromossomo  
 1488 (100, 75, 50, 10 e 0), uma vez que quanto maior o número de QTL, maior foi o  $r^2$  observado.

1489 Tabela 3-1 – Sumarização dos cromossomos analisados e estatísticas de  $r^2$ , levando-se em conta apenas  
 1490 marcadores adjacentes

Cromossomo	Tamanho (Mb)	N pares SNPs	Distância média (kb)	Desvio padrão da Distância	$r^2$ médio	Desvio Padrão $r^2$	$r^2$ mediano
1	145,9	2341	62,3	67,17	0,36	0,37	0,17
2	125	1990	62,8	67,48	0,33	0,36	0,16
3	115,9	1795	64,6	70,09	0,32	0,36	0,14
4	109,9	1594	69	74,14	0,31	0,34	0,14
5	118	1508	78,3	78,23	0,31	0,35	0,14

1491 Fonte: Autoria própria.

1492 Os valores médios e medianos de  $r^2$  entre marcadores adjacentes obtidos nesta  
 1493 simulação (Tabela 1) foram superiores aos reportados na literatura. Neves et al. (2014)  
 1494 utilizaram um painel de marcadores de alta densidade para analisar uma população de  
 1495 bovinos da raça Nelore. Os autores reportaram  $r^2$  médio de 0,293, e mediano de 0,164,  
 1496 entre pares de marcadores adjacentes em todo o genoma. Brito et al. (2011) simularam  
 1497 duas populações de bovinos de corte, com uma característica aditiva de herdabilidade 0,25

1498 e um painel de genotipagem de média densidade (50 mil marcadores). Os valores médios  
 1499 de  $r^2$  obtidos no cromossomo 1 foram de 0,25 e 0,33, com desvio padrão de 0,25 e 0,30,  
 1500 para populações de maior ( $N_e = 272$ ) e menor ( $N_e = 118$ ) tamanho efetivo,  
 1501 respectivamente. Os valores de  $r^2$  reportados em outros cromossomos foram similares. Lu  
 1502 et al. (2012) utilizaram painéis de 30 mil marcadores com uma distância média de 70 kb  
 1503 entre marcadores, e reportaram  $r^2$  médios de 0,25, 0,17 e 0,17 entre marcadores adjacentes  
 1504 do cromossomo 1 nas raças Angus, Charolês e cruzados, respectivamente.

1505 A Tabela 2 mostra estatísticas dos  $r^2$  obtidos entre pares de marcadores com  
 1506 diferentes distâncias. Conforme esperado, quanto maior a distância entre os marcadores,  
 1507 menor o  $r^2$  observado. O maior  $r^2$  médio foi de 0,36, em distâncias de até 50 kb, enquanto  
 1508 que o menor valor foi de 0,16, em distâncias de de 500 a 1000 kb. O  $r^2$  mediano foi inferior,  
 1509 variando de 0,18 a 0,08, evidenciando juntamente com o alto desvio padrão (em média 0,3)  
 1510 a existência de pares em alto desequilíbrio que se destacam em relação a maioria dos  
 1511 pares.

1512 Tabela 3-2 – Desequilíbrio de ligação ( $r^2$ ) entre pares de marcadores separados por diferentes distâncias.

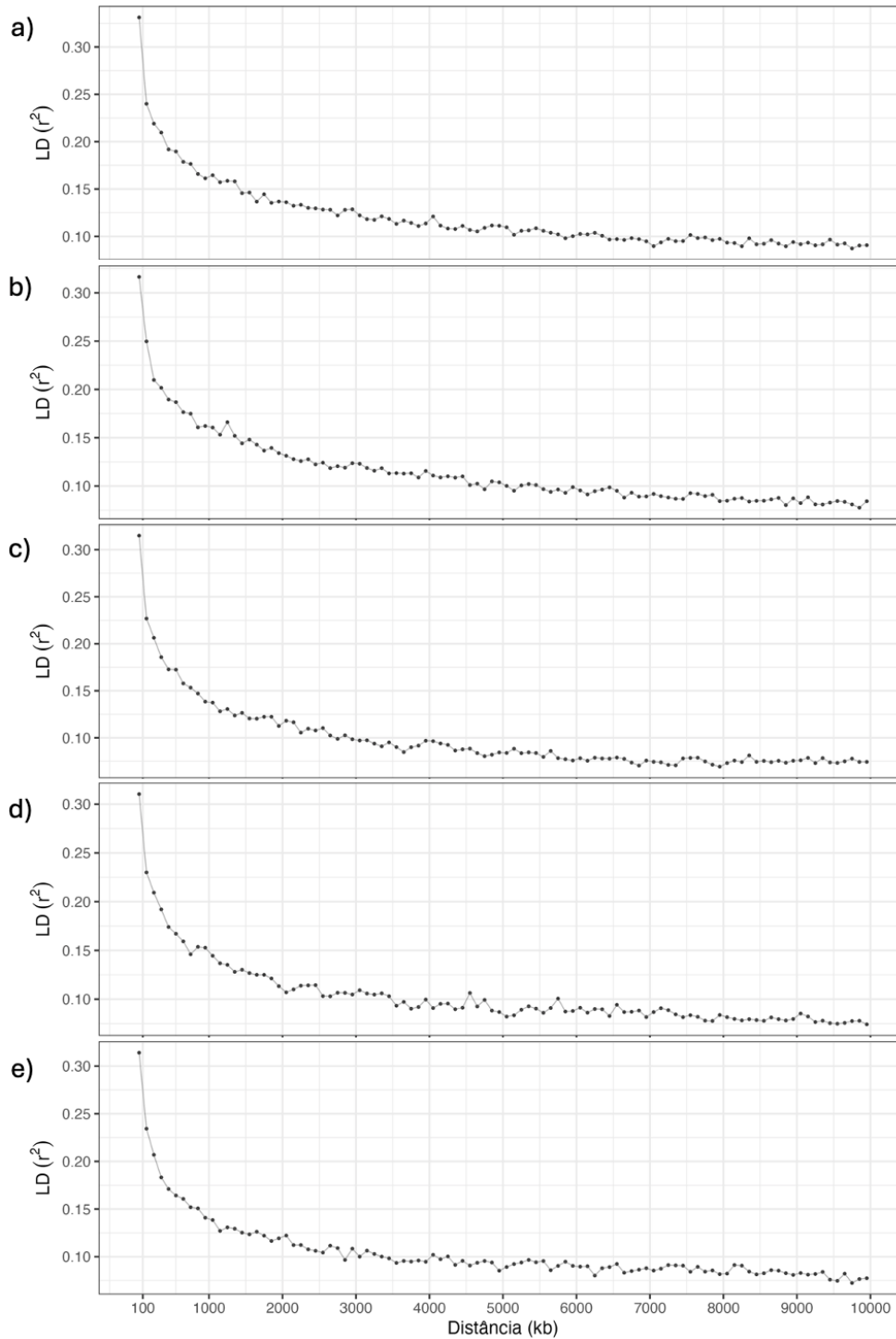
Distância (kb)	N pares SNPs	$r^2$ médio	Desvio Padrão do $r^2$	$r^2$ mediano	% $r^2 > 0,1$	% $r^2 > 0,3$	% $r^2 > 0,5$
0-50	7443	0,36	0,38	0,18	0,62	0,4	0,31
50-100	7190	0,28	0,32	0,13	0,55	0,32	0,22
100-200	14.125	0,24	0,29	0,1	0,51	0,28	0,18
200-300	14.021	0,21	0,27	0,09	0,48	0,25	0,14
300-400	14.095	0,2	0,25	0,08	0,47	0,23	0,13
400-500	14.079	0,18	0,23	0,08	0,46	0,21	0,11
500-1000	69.393	0,16	0,22	0,07	0,43	0,18	0,09

1513 Fonte: Autoria própria.

1514 A Figura 6 mostra a curva de decaimento de desequilíbrio de ligação nos  
 1515 cromossomos simulados. De forma geral, o  $r^2$  inicia-se em aproximadamente 0,35 entre  
 1516 pares de marcadores com até 100 kb de distância, caindo rapidamente abaixo de 0,2 e  
 1517 chegando próximo a 0,15 a distâncias de 1000 kb, finalizando a curva próximo a 0,1 no  
 1518 primeiro cromossomo e próximo a 0 no cromossomo 5, ao considerar-se distâncias de 10  
 1519 Mb. O primeiro cromossomo apresentou uma curva levemente mais alta que os demais  
 1520 cromossomos, como já observado pelo maior  $r^2$  reportado nas Tabelas 1 e 2. Ao comparar-

1521 se distâncias similares, Pérez O'Brien et al. (2014) reportaram  $r^2$  inferiores a 0,1 a distâncias  
1522 de 1000 kb utilizando janelas de 10 kb, e  $r^2$  entre 0,1 e 0,03 a distâncias de 1 a 10 Mb  
1523 utilizando janelas de 100 kb.

1524 Figura 3-6 – Decaimento de desequilíbrio de Ligação em uma população de bovinos de corte simulada.  
1525 Cromossomos (a) 1, (b) 2, (c) 3, (d) 4 e (e) 5.



1526  
1527

Fonte: Autoria própria.

1528 Os valores de  $r^2$  observados nesta simulação foram consistentemente superiores  
 1529 aos reportados na literatura. Isto provavelmente ocorreu devido ao baixo número de  
 1530 animais utilizados por 10 gerações na população recente, com 50 fêmeas e 5 machos  
 1531 apenas ( $N_e = (4*5*50)/(5+50) = 18,18$ ). O baixo tamanho efetivo das populações relaciona-  
 1532 se diretamente com o aumento no desequilíbrio de ligação observado (BRITO et al., 2011;  
 1533 PORTO-NETO; KIJAS; REVERTER, 2014; SARGOLZAEI et al., 2008). Um segundo ponto  
 1534 que deve ser notado é a alta dispersão dos valores observados, evidenciada pelo elevado  
 1535 desvio padrão do  $r^2$  (Tabela 2), o que indica que entre certos pares de marcadores o  $r^2$  é  
 1536 bem mais alto ou baixo do que a média observada. No entanto, altos desvios padrões  
 1537 também foram reportados por Brito et al. (2011).

### 1538 3.4.2 Perfil de corridas de homozigidade

1539 Foram identificadas 5251 corridas de homozigidade na última geração da  
 1540 população recente, com tamanho médio de 3,763 Mb ( $\pm 6,685$ ), e máximo de 117,923 Mb  
 1541 (encontrada no cromossomo 5). É importante notar que o número de cromossomos  
 1542 incluídos na análise impacta diretamente no número de corridas identificadas por animal.  
 1543 No presente estudo, apenas 5 cromossomos foram simulados. Ao analisar o número de  
 1544 corridas por cromossomo, foram identificadas 1187 (22,6%), 1023 (19,5%), 1003 (19,1%),  
 1545 984 (18,7%) e 1054 (20,1%) corridas nos cromossomos de 1 a 5, respectivamente (Tabela  
 1546 1). Quanto maior o cromossomo, maior foi a proporção de corridas identificadas, com  
 1547 exceção do cromossomo 5, que apesar de ser o menor cromossomo, apresentou um  
 1548 número maior de corridas do que o cromossomo 2.

1549 Tabela 3-3 – Número médio e porcentagem de ROHs em cada cromossomo de indivíduos de uma  
 1550 população simulada de bovinos de corte.

Cromossomo	N médio de ROHs por indivíduo	Tamanho médio de ROHs por indivíduo (Mb)	Tamanho máximo (Mb)	% de ROHs
1	12,11	4,54	80,4	0,226
2	10,44	4,97	81,8	0,195
3	10,23	3,81	45,3	0,191
4	10,04	3,56	58,2	0,187
5	10,76	5,39	118	0,201

1551 Fonte: Autoria própria.

1552 O FROH médio obtido na população foi de 0,328 ( $\pm$  0,079). As corridas foram  
 1553 divididas em 5 classes: até 2 Mb, 2 a 4 Mb, 4 a 8 Mb, 8 a 16 Mb e superior a 16 Mb (Tabela  
 1554 4). O número de corridas identificadas foi inversamente proporcional à classe de tamanho:  
 1555 a maioria das corridas (54%) apresentou entre 1 e 2 Mb de comprimento, com 2803 corridas  
 1556 (aproximadamente 28,6 por indivíduo), enquanto que apenas 212 (4%) das corridas  
 1557 identificadas apresentaram mais do que 16 Mb. No entanto, todos os indivíduos analisados  
 1558 apresentaram pelo menos uma corrida nas 3 primeiras classes de tamanho, apenas 2  
 1559 indivíduos não apresentaram nenhuma corrida maior que 8 Mb, e 16 indivíduos não  
 1560 apresentaram nenhuma corrida maior que 16 Mb (Tabela 4).

1561 Tabela 3-4 – Perfil de corridas de homozigidade e FROH por diferentes classes de tamanho (Mb).

Classe (Mb)	Número de ROHs	Proporção de ROHs	Número de indivíduos (% da população)	FROH médio
1-2	2803	53,38%	98 (100%)	0,33
2-4	1493	28,43%	98 (100%)	0,26
4-8	548	10,44%	98 (100%)	0,19
8-16	195	3,71%	96 (98%)	0,15
>16	212	4,04%	82 (84%)	0,11

1562 Fonte: Autoria própria.

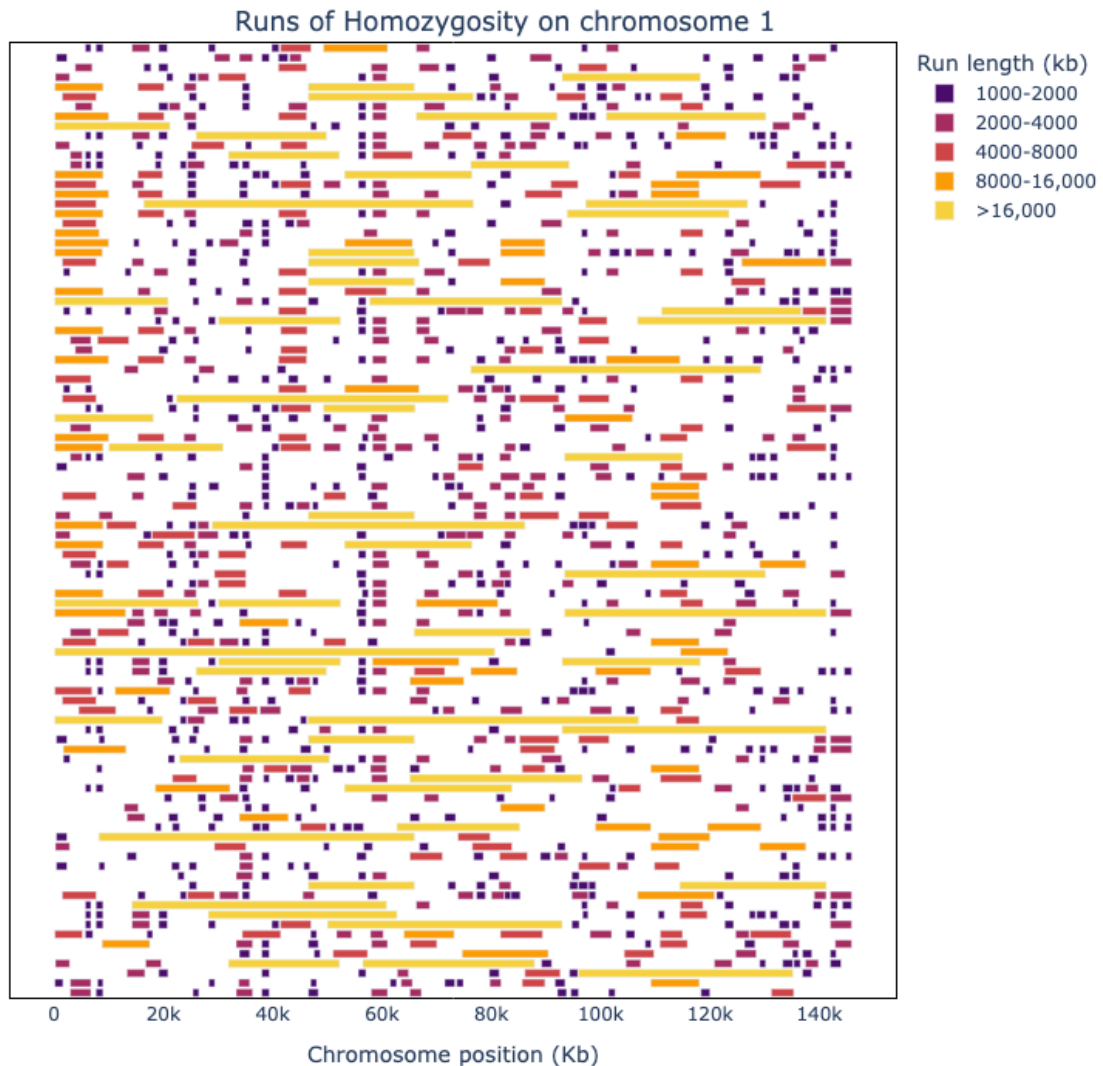
1563 Peripolli et al. (2018) analisaram um total de 9.386 bovinos da raça Nelore, com  
 1564 informações de 735.044 marcadores obtidos por meio de imputing. Foram identificadas em  
 1565 média 55,15 ROH por animal, com tamanho médio de 3,24 Mb, sendo a maior ROH de 99,3  
 1566 Mb. Assim como no presente estudo, os autores identificaram a prevalência de segmentos  
 1567 curtos (de 1 a 2 Mb e 2 a 4 Mb), que corresponderam a 78% de todas as ROH identificadas  
 1568 e 43% da cobertura genômica por ROH. O FROH médio obtido foi de 7,15%, bastante  
 1569 inferior ao obtido no presente estudo (32,8%). Ao calcular o FROH por classes de tamanho,  
 1570 Peripolli et al. (2018) identificaram valores de 0,016, 0,014, 0,015, e 0,021 nas 4 primeiras  
 1571 classes reportadas na Tabela 4.

1572 De Souza Fonseca et al. (2016) utilizaram 32.806 marcadores moleculares (após  
 1573 realizado controle de qualidade) para identificar ROH em 210 bovinos da raça Guzerá, uma  
 1574 raça bovina de dupla aptidão. Os autores observaram um número médio de 389,81 ROH  
 1575 por indivíduo, com tamanho médio de 1,980 Mb, mediano de 1,578 Mb e máximo de 85,986  
 1576 Mb. O número de corridas identificadas foi superior ao reportado no presente estudo, no

1577 entanto o comprimento das corridas foi menor. Os valores de FROH obtidos foram de 0,307,  
 1578 0,148, 0,033 e 0,024 para as classes de ROH >1 Mb, >2 Mb, >8 Mb e >16 Mb,  
 1579 respectivamente.

1580 O perfil de ROHs observado nos cromossomos de 1 a 5 estão ilustrados nas Figuras  
 1581 de 7 a 11, respectivamente. Nota-se o grande número de corridas curtas (até 4 Mb), com  
 1582 certa concentração em determinados locais, como a exemplo das posições próximas a 60  
 1583 Mb no primeiro cromossomo e próxima a 20 Mb no segundo cromossomo. Os  
 1584 cromossomos 3 e 4 apresentam visualmente um número maior de corridas pequenas, já o  
 1585 5º cromossomo apresenta novamente um número elevado de corridas grandes.

1586 Figura 3-7 – Perfil de corridas de homozigosidade no cromossomo 1 em uma população simulada de bovinos  
 1587 de corte.



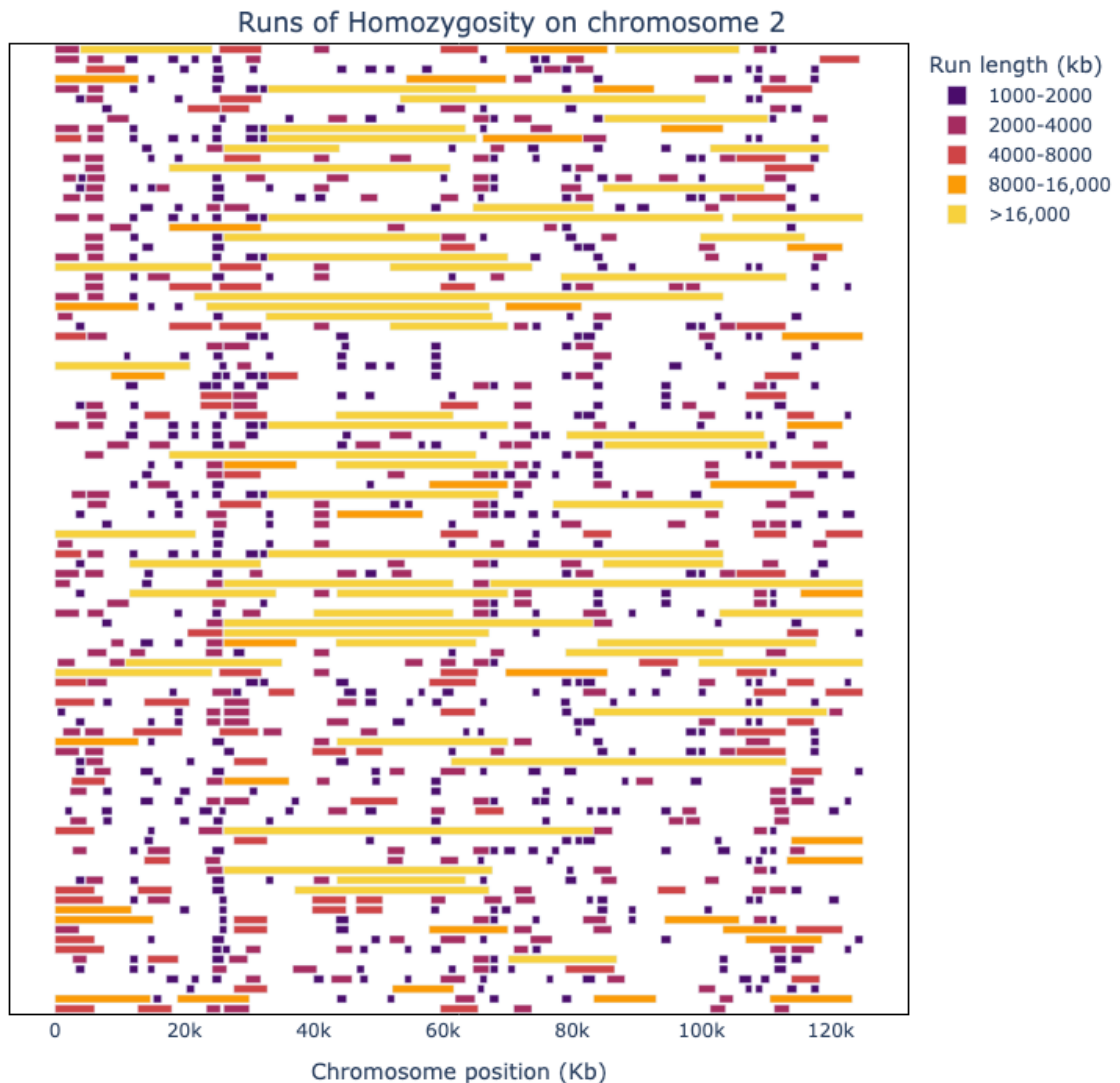
1588

1589 Fonte: Autoria própria.



1590 Na Figura 7, observa-se uma concentração de corridas de tamanho superior a 4 Mb  
 1591 no início do cromossomo, o que é de certa forma inesperado, uma vez que eventos de  
 1592 recombinação tendem a ocorrer com maior frequência nas extremidades dos cromossomos  
 1593 e assim, quebrar as corridas. Isto pode estar relacionado ao posicionamento e efeitos de  
 1594 QTL, e deverá ser investigado em estudos futuros sobre o tema. Na Figura 8, é possível  
 1595 visualizar como as corridas maiores tendem a ser localizadas mais próximas ao centro do  
 1596 cromossomo. Na figura 11, observa-se uma corrida que se estende ao longo de todo o  
 1597 cromossomo, apontando a possibilidade de um acasalamento endogâmico recente.

1598 Figura 3-8 – Perfil de corridas de homozigidade no cromossomo 2 em uma população simulada de bovinos  
 1599 de corte.



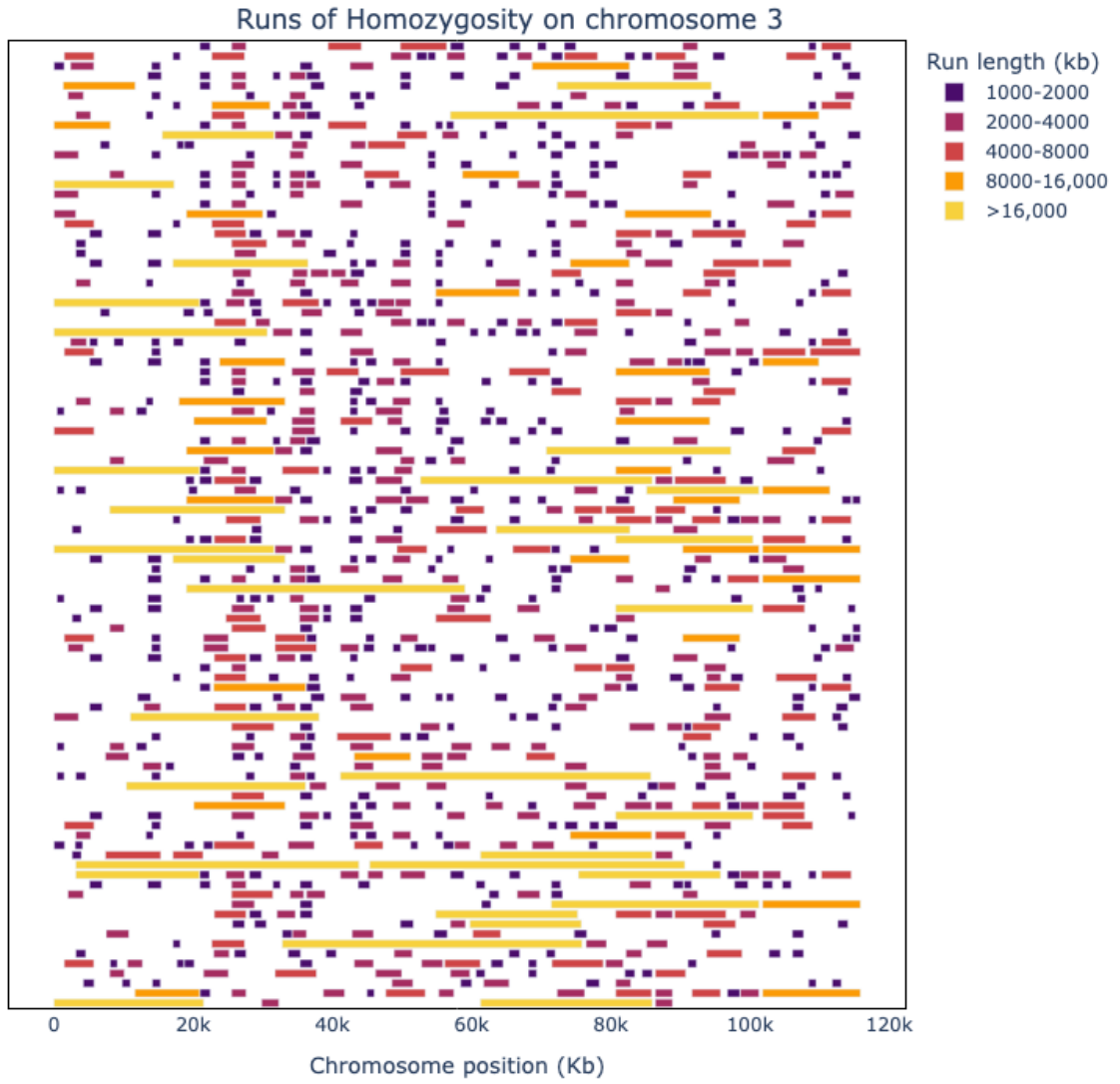
1600

1601 Fonte: Autoria própria.

1602

1603  
1604

Figura 3-9 – Perfil de corridas de homozigosidade no cromossomo 3 em uma população simulada de bovinos de corte.

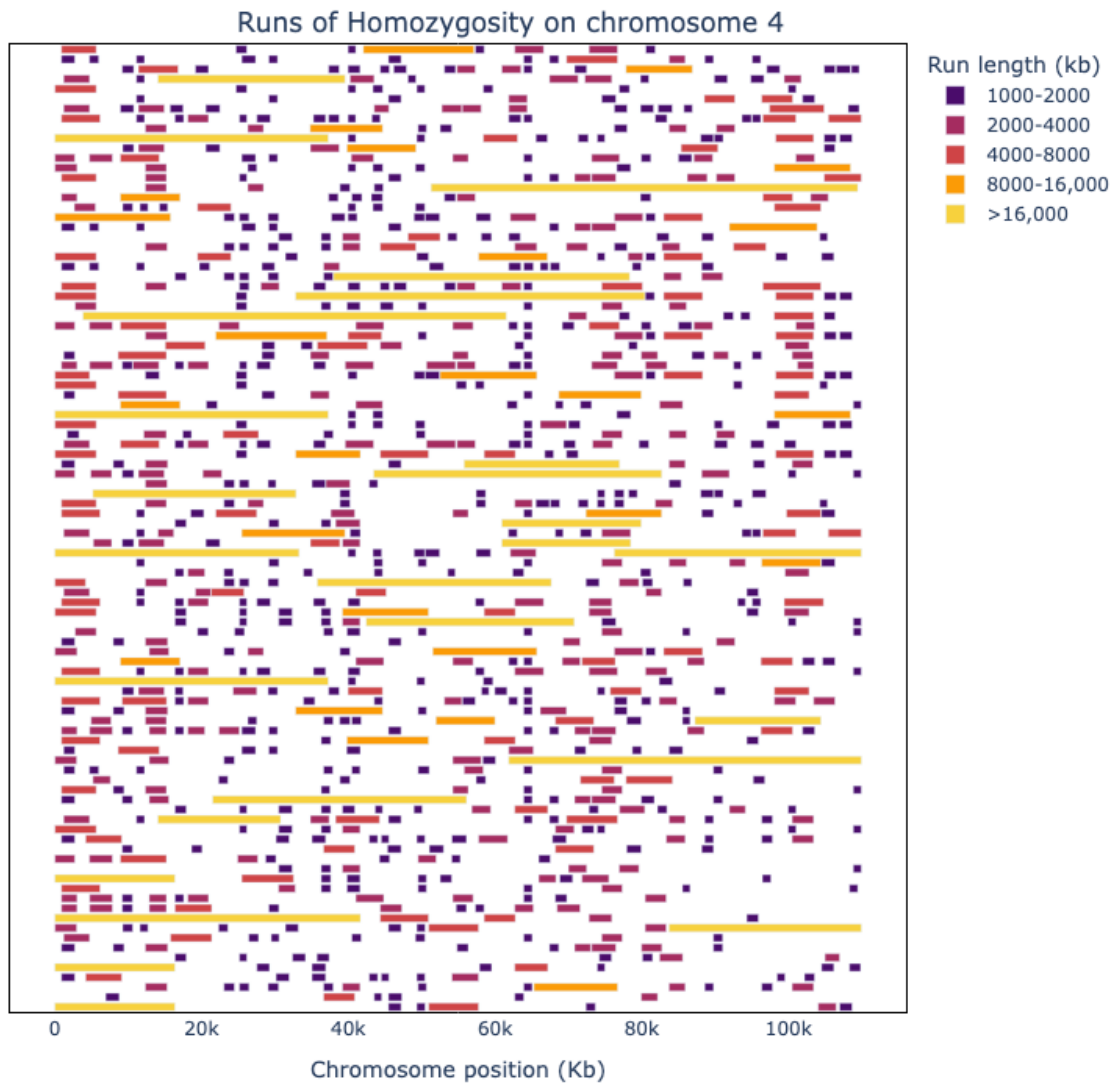


1605  
1606  
1607

Fonte: Autoria própria.

1608  
1609

Figura 3-10 – Perfil de corridas de homozigosidade no cromossomo 4 em uma população simulada de bovinos de corte.



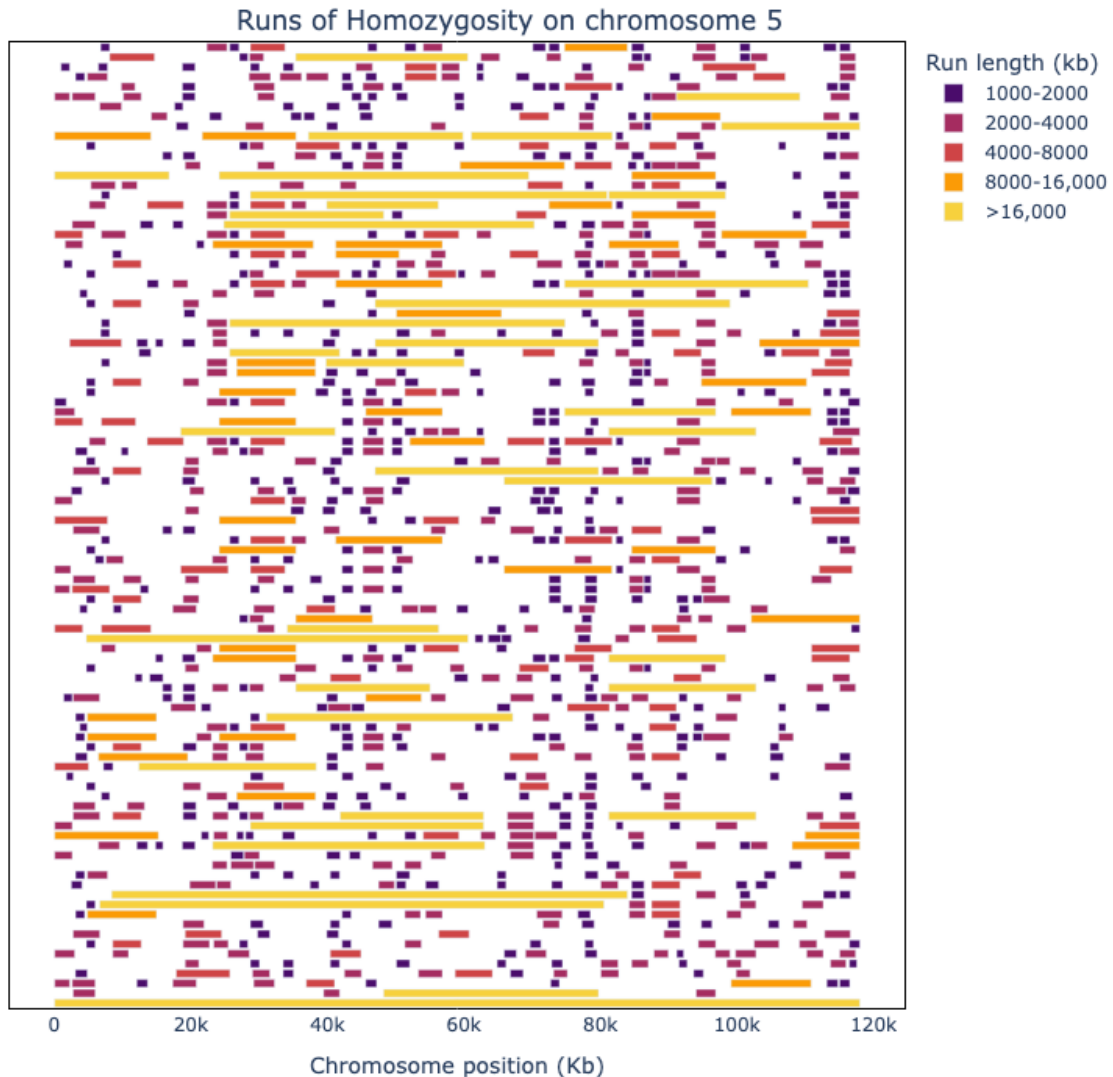
1610

1611 Fonte: Autoria própria.

1612

1613  
1614

Figura 3-11 – Perfil de corridas de homozigosidade no cromossomo 5 em uma população simulada de bovinos de corte.



1615

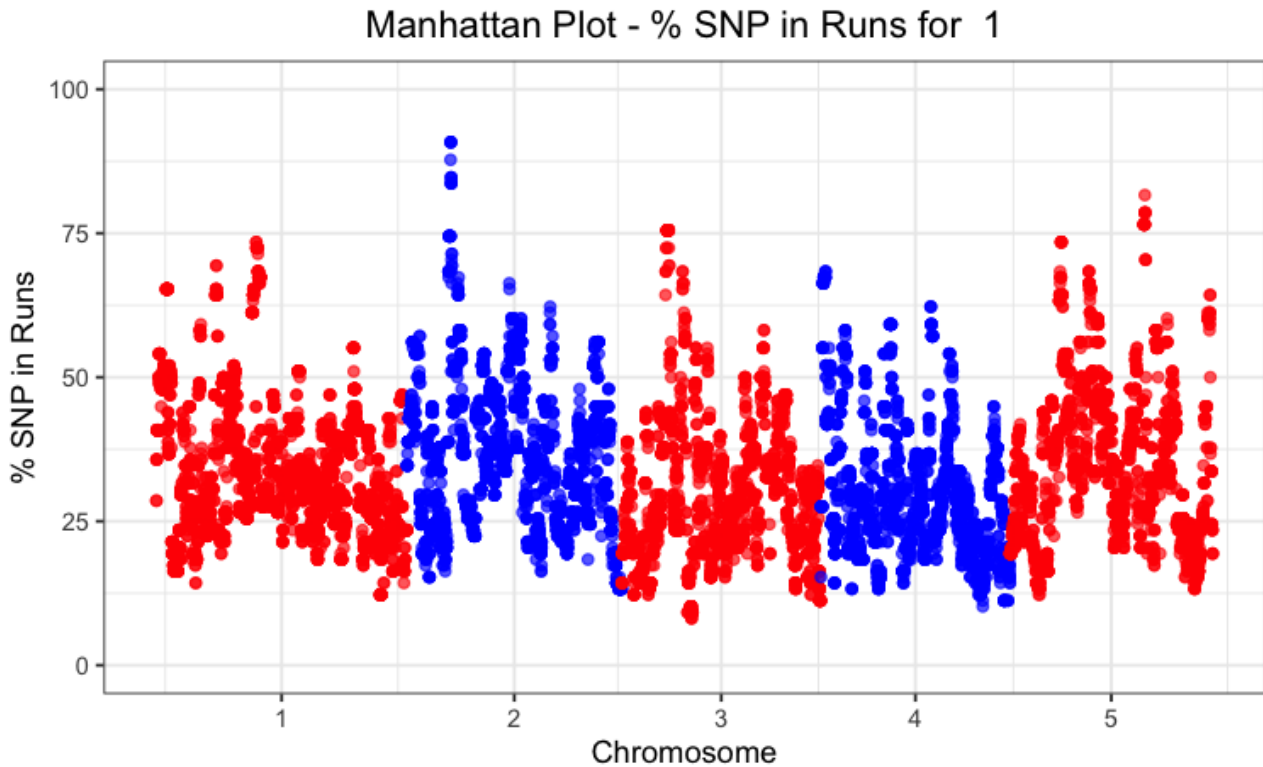
1616 Fonte: Autoria própria.

1617 Na Figura 12 é possível observar a frequência em que cada marcador aparece em  
 1618 uma ROH, onde 0 indica que nenhum dos 98 animais apresentou uma ROH na posição  
 1619 daquele marcador e 100 indica que todos os 98 animais avaliados apresentaram ROH. De  
 1620 forma geral, os 5 cromossomos apresentaram regiões em alta homozigosidade, com  
 1621 marcadores presentes em ROHs em mais de 50% da população. O segundo cromossomo  
 1622 apresentou o maior pico, seguido pelo cromossomo 5. O quarto cromossomo apresentou  
 1623 os picos mais baixos, porém ainda bastante expressivos, acima de 50%.

1624

1625  
1626

Figura 3-12 – Proporção de marcadores em corridas de homoziguidade em 5 cromossomos de uma população simulada de bovinos de corte.



1627

1628

Fonte: Autoria própria.

1629

1630

1631

1632

1633

1634

1635

1636

1637

Os motivos para estes resultados não são claros, uma vez que seria esperado observar-se uma maior homoziguidade nos cromossomos com maior número de QTL, uma vez que estas regiões encontram-se sob maior pressão de seleção. Neste estudo, a localização e os efeitos dos QTL não foram investigados, no entanto, é possível que os efeitos dos QTL no cromossomo 2 tenham sido mais elevados que nos demais cromossomos, causando o aumento da homoziguidade devido à seleção por tais regiões. Além disso, o tamanho populacional reduzido também pode ter ocasionado um aumento geral de endogamia, uma vez que torna-se mais difícil evitar acasalamentos consanguíneos em populações pequenas.

1638

1639

### 3.4.3 Comparação dos cenários de seleção e direcionamento de acasalamentos

1640

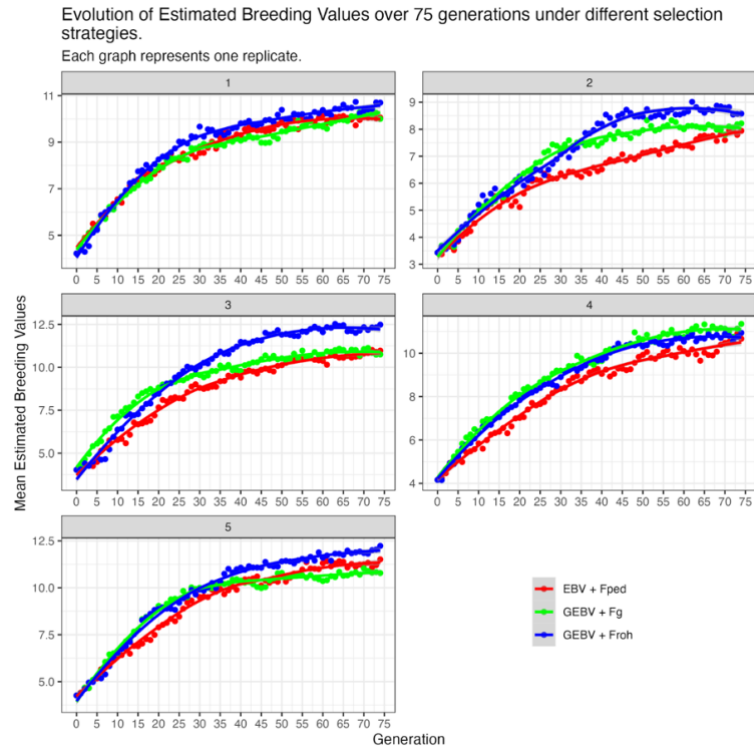
1641

Foram comparados 3 cenários com diferentes métricas adotadas para realizar-se a seleção e o direcionamento dos acasalamentos, por 75 gerações. No primeiro cenário os

1642 valores genéticos foram estimados utilizando-se dados fenotípicos e de pedigree, e o  
1643 coeficiente de endogamia utilizado baseou-se no pedigree (EBV + Fped). O segundo  
1644 cenário incluiu dados genotípicos no cálculo de valores genéticos e o coeficiente de  
1645 endogamia utilizado foi o genômico (GEBV + Fg). O terceiro cenário substituiu o coeficiente  
1646 de endogamia genômico pelo coeficiente de endogamia de corridas de homozigidade  
1647 (GEBV + Froh).

1648 Em cada um dos cenários avaliados, foram obtidos os valores genéticos estimados  
1649 pelo AlphaSimR e pelo BLUPF90, os valores genéticos verdadeiros, fenótipos, Fped, Fg e  
1650 Froh de todos os animais. A simulação foi realizada 5 vezes, com o objetivo de verificar a  
1651 consistência dos resultados. A Figura 13 apresenta as médias dos GEBVs calculados pelo  
1652 AlphaSimR para cada geração, nos 3 cenários e nas 5 replicatas. As 3 estratégias  
1653 apresentaram desempenho similar. Era esperado que a estratégia sem genômica (EBV +  
1654 Fped) seria menos eficaz na seleção de animais de valor genético superior, uma vez que a  
1655 inclusão de informações genômicas aumenta a acurácia das estimativas, levando ao  
1656 aumento mais acentuado do (G)EBV a cada geração (HAYES et al., 2009; VANRADEN et  
1657 al., 2009). No entanto, não é possível afirmar que tal estratégia foi consistentemente inferior  
1658 nas 5 replicatas quando considera-se apenas os GEBVs calculados pelo AlphaSimR  
1659 (Figura 13).

1660 Figura 3-13 – Médias dos valores genéticos estimados para cada geração em um processo de seleção com  
1661 3 diferentes estratégias, partindo-se de uma mesma população. Cada gráfico representa uma  
1662 simulação de processo estocástico gerada.



1663

1664 Fonte: Autoria própria.

1665

1666

1667

1668

1669

1670

1671

1672

1673

1674

A Figura 14 apresenta as médias dos valores genéticos verdadeiros dos indivíduos de cada geração. É possível observar uma tendência de aumento dos valores genéticos verdadeiros para as estratégias que fazem o uso de informações genômicas, conforme esperado. Apenas na replicata 5, a estratégia EBV + Fped superou a estratégia GEBV + Fg. É possível que durante o desenvolvimento do segundo cenário da replicata 5, alguns alelos desfavoráveis tenham sido fixados por deriva genética, uma vez que trata-se de uma população pequena. Isto pode ter reduzido o progresso genético da população, mesmo que a estratégia de seleção tenha sido favorável. Ao comparar as duas estratégias com genômica, apenas em uma das replicatas o cenário 2 (GEBV + Fg) superou a estratégia com Froh.

1675

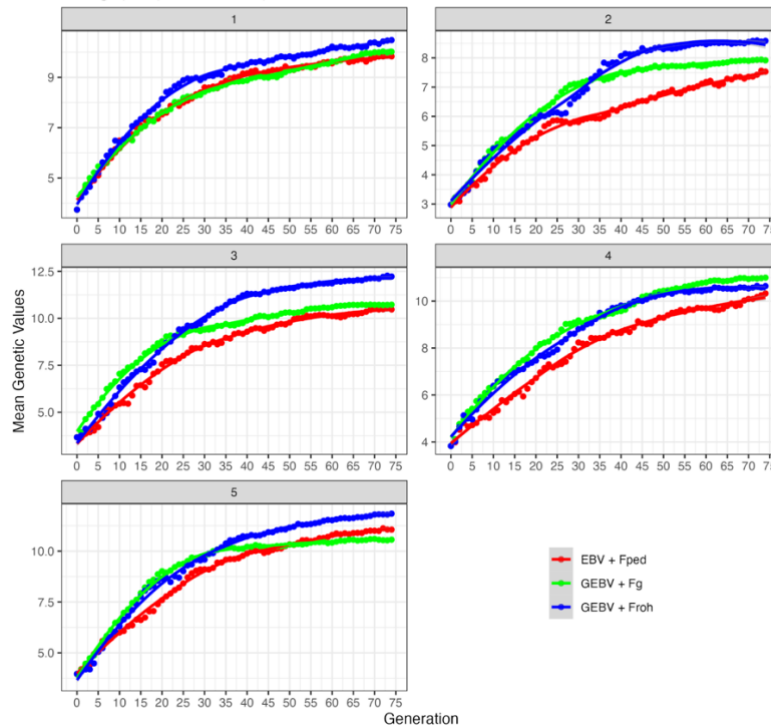
1676

Figura 3-14 – Médias dos valores genéticos verdadeiros para cada geração em um processo de seleção com 3 diferentes estratégias, partindo-se de uma mesma população. Cada gráfico representa uma

1677

simulação de processo estocástico gerada.

Evolution of Genetic Values over 75 generations under different selection strategies.  
Each graph represents one replicate.



1678

1679

Fonte: Autoria própria.

1680

1681

1682

1683

A Figura 15 apresenta a média dos fenótipos a cada geração. Os resultados são consistentes com os observados em relação aos valores genéticos verdadeiros e estimados. Percebe-se uma tendência das estratégias genômicas resultarem em fenótipos superiores, especialmente quando se faz o uso do Froh.

1684

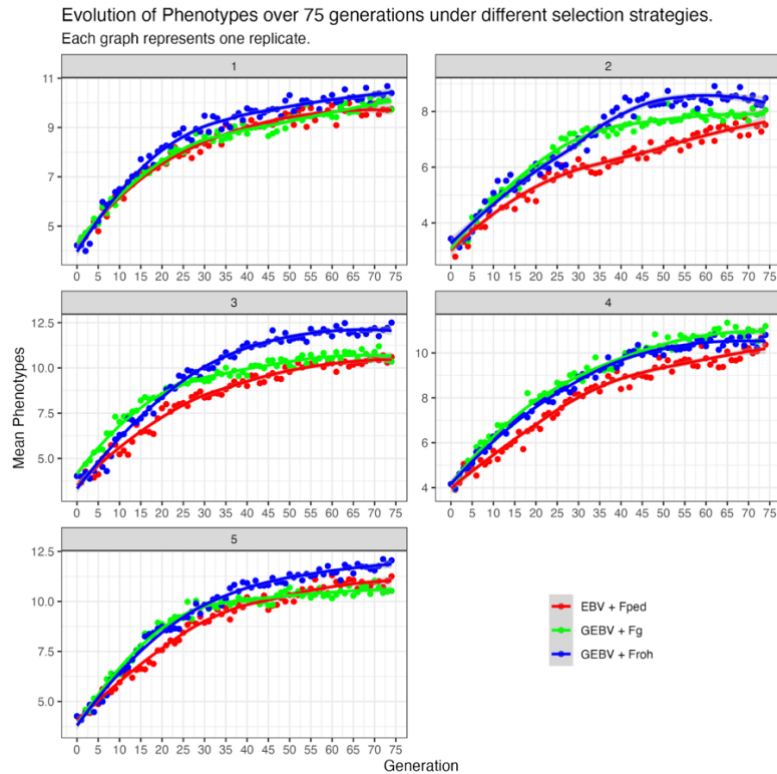
1685

Figura 3-15 – Médias dos fenótipos para cada geração em um processo de seleção com 3 diferentes estratégias, partindo-se de uma mesma população. Cada gráfico representa uma simulação de



1686

processo estocástico gerada.



1687

1688 Fonte: Autoria própria.

1689

1690

1691

1692

1693

1694

1695

1696

1697

1698

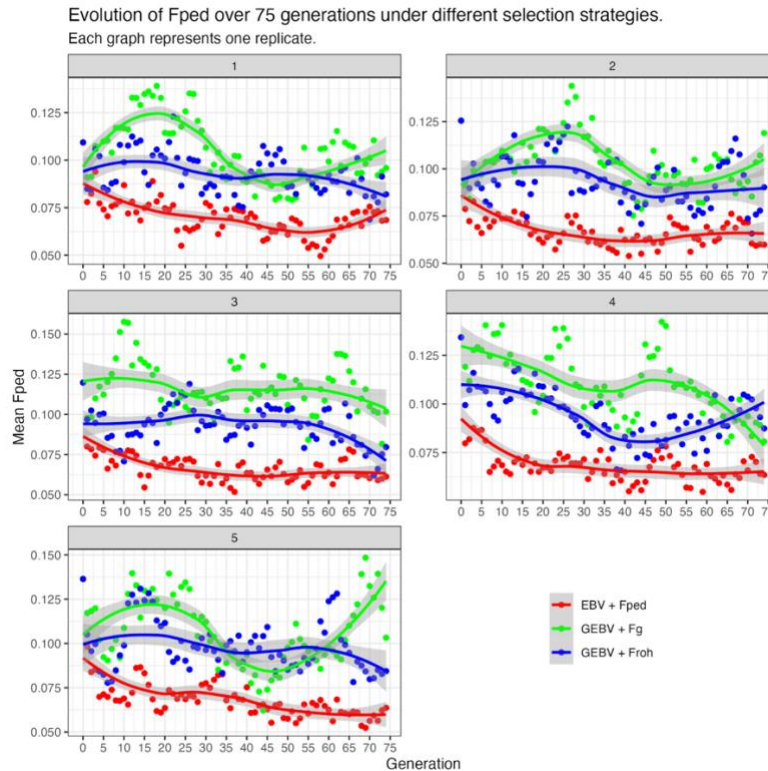
1699

A Figura 16 apresenta as médias de Fped ao longo das gerações. Conforme esperado, a estratégia que gerou o Fped mais baixo ao longo das gerações foi a EBV + Fped. É interessante observar que nas 5 replicatas houve redução do Fped ao longo das gerações, mesmo se tratando de uma população pequena. A recomendação atual é que a endogamia não aumente mais do que 1% por geração (HOWARD et al., 2017a), e pode-se afirmar que a estratégia atendeu a esta recomendação. As duas estratégias com genômica apresentaram resultados menos consistentes em relação ao Fped. Ao utilizar o Fg como métrica de seleção e direcionamento, observou-se um aumento do Fped nas primeiras 15-25 gerações em algumas das replicatas. A estratégia com Froh apresentou um comportamento menos errático, e de forma geral manteve o Fped igual ou inferior ao início do processo de seleção, porém apresentou grande variação entre as replicatas.

1700

1701  
1702  
1703

Figura 3-16 – Médias dos coeficientes de endogamia por pedigree para cada geração em um processo de seleção com 3 diferentes estratégias, partindo-se de uma mesma população. Cada gráfico representa uma simulação de processo estocástico gerada.



1704  
1705

Fonte: Autoria própria.

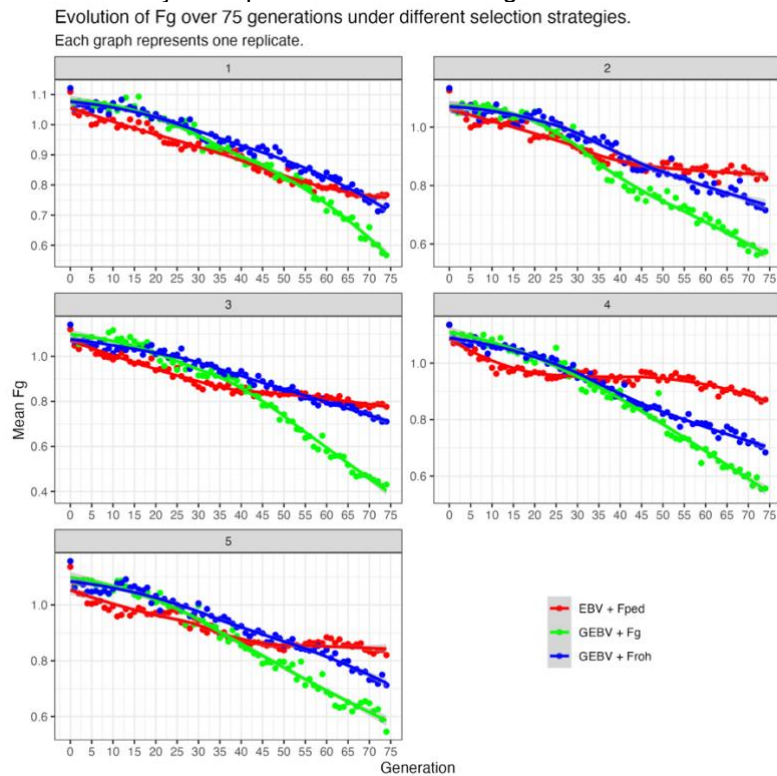
1706  
1707  
1708  
1709  
1710

A Figura 17 apresenta as médias de Fg ao longo das gerações. As duas estratégias genômicas foram mais efetivas na redução do Fg, especialmente a que usa este coeficiente como métrica de seleção e direcionamento de acasalamentos. É interessante observar que, apesar de ter sido menos eficiente, a estratégia EBV + Fped também promoveu a redução do Fg ao longo das gerações.

1711  
1712

Figura 3-17 – Médias dos coeficientes de endogamia genômicos para cada geração em um processo de seleção com 3 diferentes estratégias, partindo-se de uma mesma população. Cada gráfico

1713 representa uma simulação de processo estocástico gerada.



1714

1715 Fonte: Autoria própria.

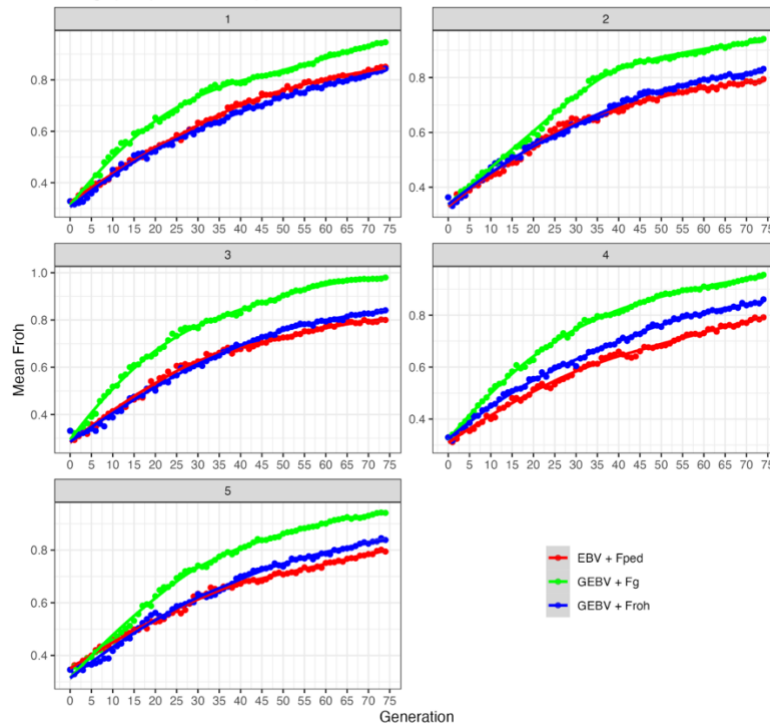
1716 A Figura 18 apresenta as médias do Froh ao longo das gerações. Diferente dos  
 1717 demais coeficientes de endogamia avaliados, o Froh apresentou aumento ao longo das  
 1718 gerações, independentemente da estratégia ou replicata. Pode-se afirmar que a estratégia  
 1719 que promoveu o maior aumento do Froh foi a GEBV + Fg, e que a estratégia EBV + Fped  
 1720 foi igual ou superior à estratégia com Froh para evitar o aumento do mesmo. Os motivos  
 1721 para estes resultados podem estar relacionados à natureza da simulação, que tende a gerar  
 1722 diversos alelos fixados, e aos tamanhos populacionais adotados.

1723 Figura 3-18 – Médias dos coeficientes de endogamia por corridas de homoziguidade para cada geração em  
 1724 um processo de seleção com 3 diferentes estratégias, partindo-se de uma mesma população.

1725

Cada gráfico representa uma simulação de processo estocástico gerada.

Evolution of Froh over 75 generations under different selection strategies.  
Each graph represents one replicate.



1726

1727

Fonte: Autoria própria.

1728

1729

1730

1731

1732

1733

1734

1735

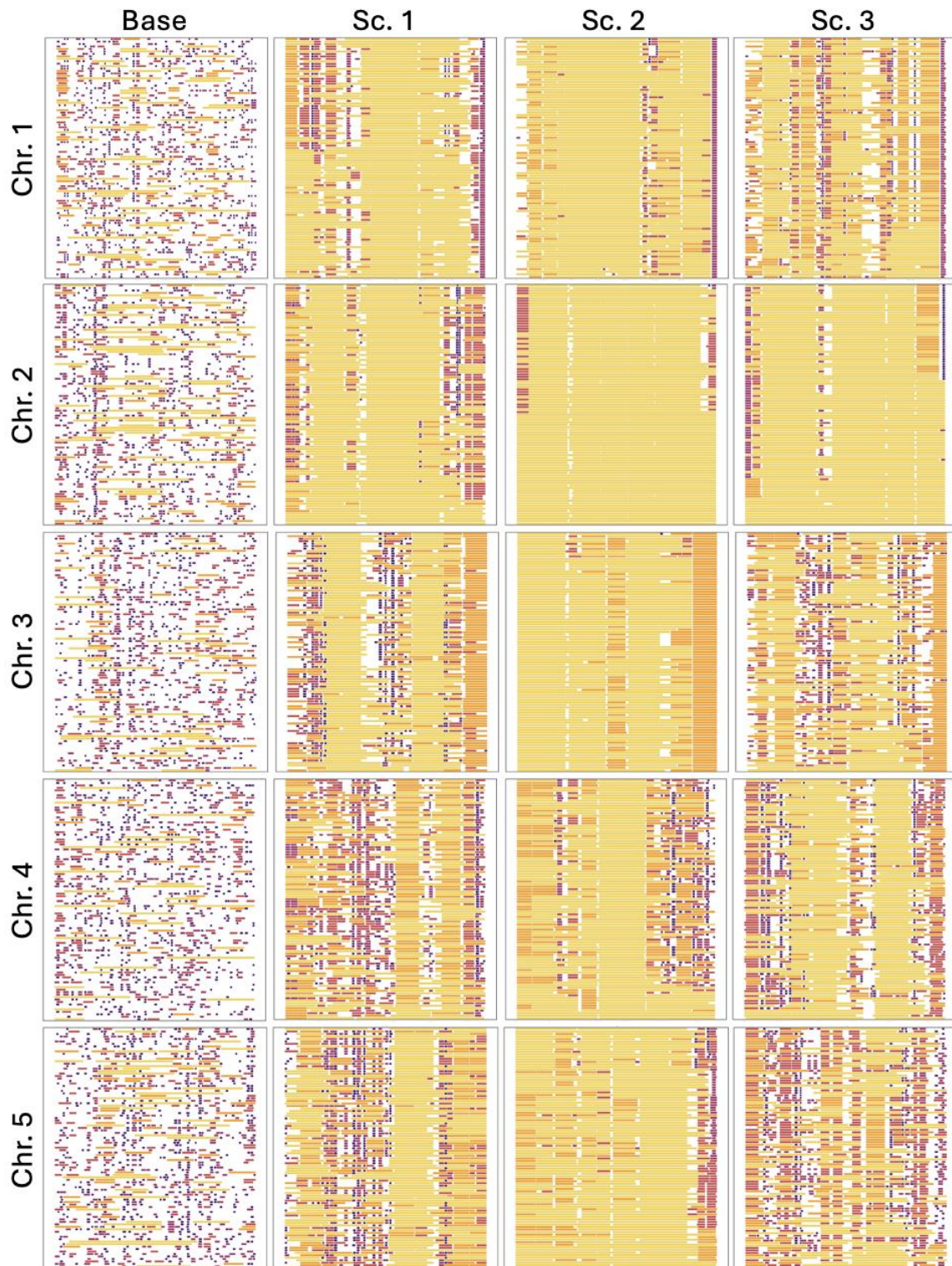
Os diferentes perfis de ROH obtidos após 75 gerações utilizando-se diferentes estratégias de seleção e acasalamentos estão ilustrados na Figura 19, bem como o perfil obtido na última geração da simulação base (replicata 1). As 3 estratégias levaram a um aumento significativo no número e comprimento das ROHs em todo o genoma simulado. É possível que, devido à simplicidade da simulação (apenas uma característica de efeito aditivo simulada, com até 100 QTL por cromossomo), e o pequeno tamanho populacional (aproximadamente 98 indivíduos e 50 acasalamentos por geração) tenham levado a uma alta taxa de fixação de alelos ao longo das 75 gerações.

1736



1737  
1738

Figura 3-19 – Perfis de corrida de homocigosidade em 5 cromossomos de uma população bovina de corte simulada (replicata 1).



1739

1740

Fonte: Autoria própria.

1741

1742

1743

A primeira coluna corresponde à última geração recente; a segunda coluna corresponde ao cenário 1 (EBV + Fped); a terceira coluna corresponde ao cenário 2 (GEBV + Fg); e a quarta coluna corresponde ao cenário 3 (GEBV + Froh). A escala de cores representa o tamanho da ROH (quanto mais escura, menor é a ROH).

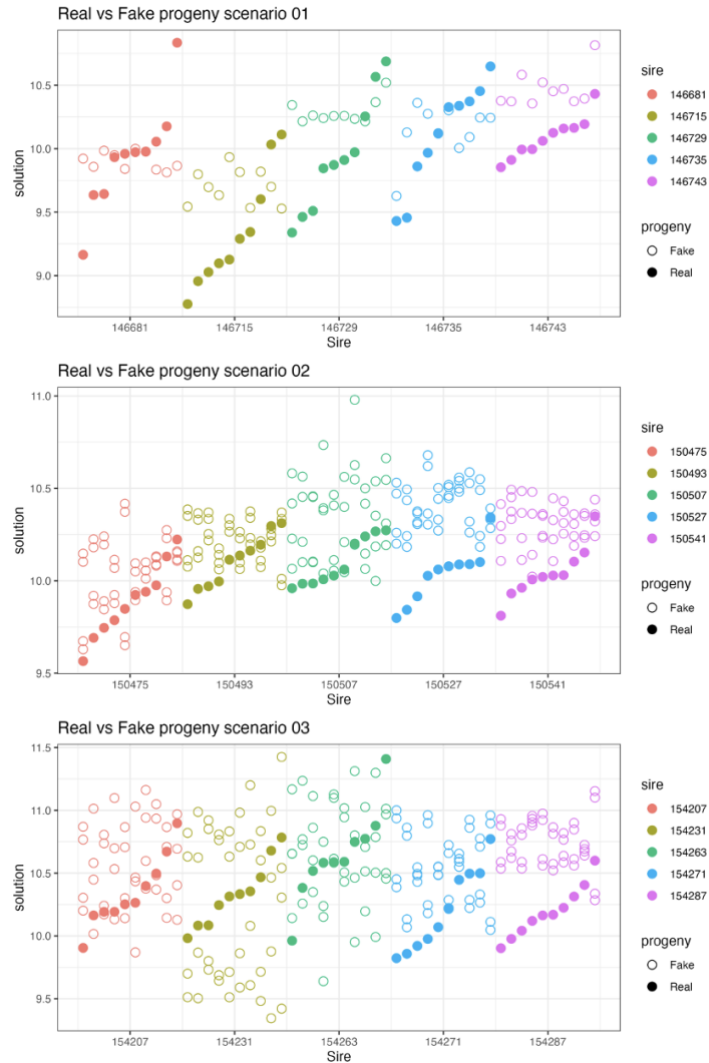
1744 No entanto, é notável que a segunda estratégia (GEBV + Fg) levou a um aumento  
1745 de homozigosidade mais extremo que as demais estratégias, conforme já indicado na  
1746 Figura 19. Ao comparar os cromossomos, evidencia-se que o segundo cromossomo  
1747 apresentou o maior aumento de homozigosidade, enquanto que o quarto e quinto  
1748 cromossomos preservaram um número maior de ROHs pequenas e algumas regiões sem  
1749 ROHs.

1750 Na Figura 20, é possível observar os (G)EBVs calculados pelo BLUPF90 para as  
1751 PHASM da penúltima geração, bem como das progênes reais geradas pelos mesmos  
1752 acasalamentos. No primeiro cenário, há apenas uma progênie falsa por acasalamento,  
1753 enquanto que nos demais cenários existem 4 PHASM para cada acasalamento. Pode-se  
1754 notar que existe uma variação entre o predito e o real, e que em muitos casos a progênie  
1755 real apresenta GEBV muito acima ou muito abaixo do máximo e mínimo preditos. Isto  
1756 ocorre, em grande parte, devido à segregação mendeliana (DAETWYLER et al., 2013), a  
1757 qual não foi levada em conta ao gerar as PHASM. É importante ressaltar que o cálculo dos  
1758 (G)EBVs das PHASM e verdadeiras foram realizados em momentos distintos, uma vez que  
1759 as PHASM foram geradas antes de realizar-se os acasalamentos que geraram as progênes  
1760 reais. Assim, é provável que existam diferentes erros para as observações falsas e reais  
1761 em um mesmo cenário, exacerbando as discrepâncias entre os valores preditos para as  
1762 PHASM e reais.

1763

1764  
1765  
1766

Figura 3-20 – Valores genéticos estimados (EBV) para as Progênes Hipotéticas com Ausência de Segregação Mendeliana e reais de cada touro selecionado em processos de seleção com 3 diferentes estratégias.



1767

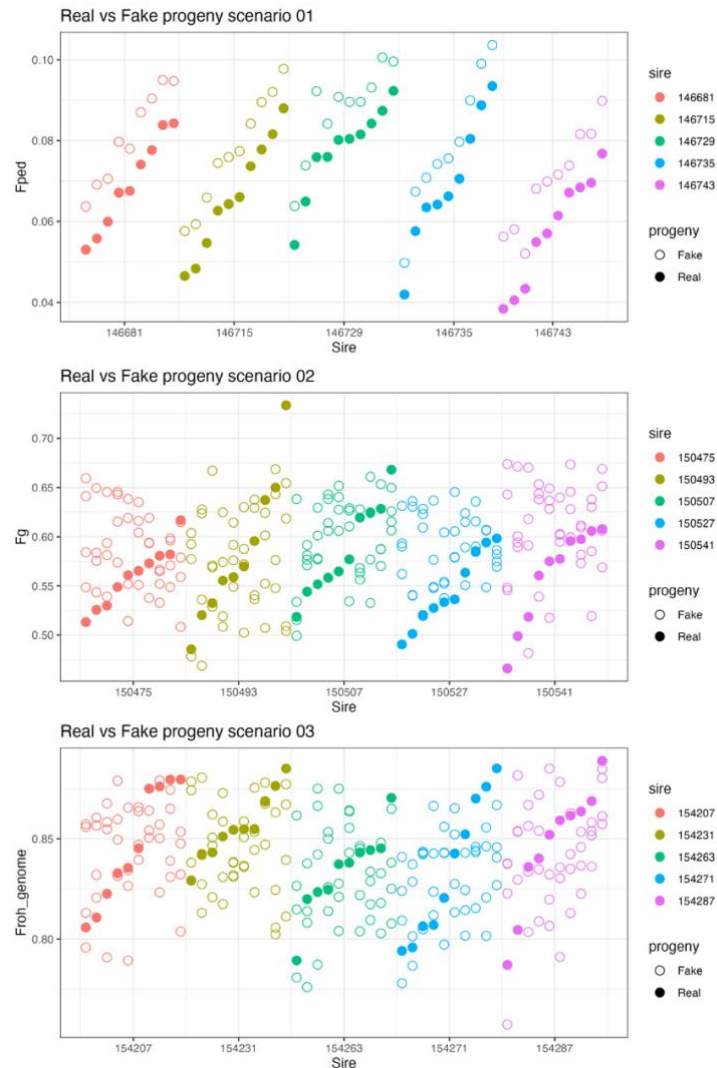
1768 Fonte: Autoria própria.

1769 Cenário 1: EBV + coeficiente de endogamia por pedigree; Cenário 2: EBV genômico + coeficiente de  
1770 endogamia genômico; Cenário 3: EBV genômico + coeficiente de endogamia por corridas de homozigidade.

1771 Na Figura 21, é possível observar os respectivos coeficientes de endogamia  
1772 utilizados para a seleção e direcionamento de acasalamentos em cada cenário, calculados  
1773 para as PHASM da penúltima geração, bem como para as progênes reais geradas pelos  
1774 mesmos acasalamentos. Novamente, pode-se verificar como o Fped e o Fg calculados para  
1775 as PHASM e reais diferem. No caso do Fped, onde não há novas informações que afetam  
1776 as previsões das progênes reais (como a adição do fenótipo), é possível que o principal  
1777 fator responsável pela diferença nas previsões é que a cada rodada do BLUPF90, utiliza-

1778 se o pedigree apenas das últimas 10 gerações. Assim, ao realizar o cálculo do Fped para  
 1779 as progênes reais, o pedigree utilizado corresponde a uma geração posterior ao pedigree  
 1780 que foi utilizado no cálculo do Fped das PHASM correspondentes ao acasalamento dos  
 1781 seus progenitores.

1782 Figura 3-21 – Coeficientes de endogamia para as Progênes Hipotéticas com Ausência de Segregação  
 1783 Mendeliana e reais de cada touro selecionado em processos de seleção com 3 diferentes  
 1784 estratégias.



1785

1786 Fonte: Autoria própria.

1787 Cenário 1: EBV + coeficiente de endogamia por pedigree; Cenário 2: EBV genômico + coeficiente de  
 1788 endogamia genômico; Cenário 3: EBV genômico + coeficiente de endogamia por corridas de homozigidade.

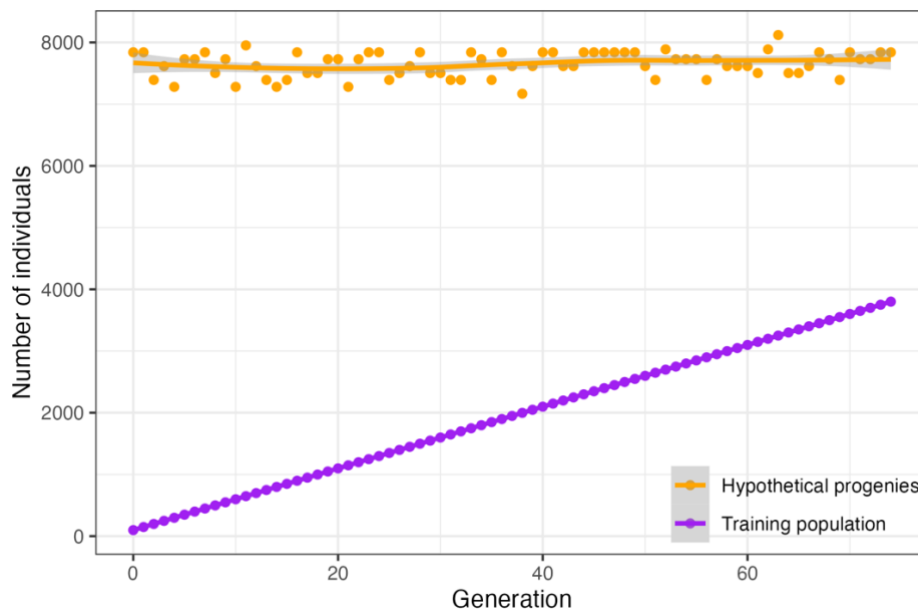
1789 No gráfico do terceiro cenário, verifica-se que o Froh das progênes reais aproxima-  
 1790 se mais da média do Froh das PHASM, quando comparado aos coeficientes de endogamia



1791 calculados nos cenários 1 e 2. O Froh é calculado através do pacote detectRUNS no R, e  
 1792 baseia-se apenas nos dados genotípicos dos próprios indivíduos (reais ou falsos).

1793 Uma das limitações deste estudo reside na proporção entre as populações de  
 1794 treino e teste nos cenários que incluem dados genômicos. A população de treino é aquela  
 1795 que contém indivíduos com genótipos e fenótipos, permitindo que os efeitos dos  
 1796 marcadores sejam estimados. A população de teste corresponde aos indivíduos que  
 1797 apresentam apenas genótipo, e seus valores genéticos são estimados através dos efeitos  
 1798 de marcadores calculados na população de treino. No início do processo de seleção e  
 1799 direcionamento de acasalamentos, apenas 98 animais compõem a população de treino,  
 1800 conforme ilustrado na Figura 2. No entanto, devido ao alto número de combinações  
 1801 possíveis entre todos os machos e fêmeas (multiplicado por 4, pois os animais são  
 1802 diplóides), a população de teste (PHASM e portanto, sem fenótipos) apresenta quase 8000  
 1803 indivíduos (Figura 22).

1804 Figura 3-22 – Número de indivíduos nas populações de treino para a predição de valores genéticos  
 1805 estimados (linha em roxo) e de Progenies Hipotéticas com Ausência de Segregação Mendeliana,  
 1806 indivíduos apenas com informações de genótipo e pedigree (linha em laranja).



1807

1808 Fonte: Autoria própria.

1809 A cada geração, aproximadamente 50 novos indivíduos são incluídos na população  
 1810 de treino, uma vez que são gerados os acasalamentos definidos pela estratégia (pelo  
 1811 menos 50, no entanto o valor pode variar caso seja necessário a inclusão de mais touros,

1812 conforme explicado ao final da seção 3.3.5). Ao final das 75 gerações, a população de treino  
1813 é composta por aproximadamente 4000 indivíduos. O número de indivíduos da população  
1814 de teste é relativamente constante, uma vez que utiliza-se apenas as PHASM geradas para  
1815 a geração atual.

#### 1816 **3.4.4 Limitações e sugestões para estudos futuros**

1817 Neste estudo, foi desenvolvido um pipeline de simulação de dados genômicos e  
1818 análises, onde diferentes estratégias de seleção e direcionamento de acasalamentos foram  
1819 propostos e testados. Até onde se sabe, não houve ainda outro trabalho na literatura que  
1820 tenha desenvolvido um pipeline com tal amplitude de ferramentas, onde o processo de  
1821 simulação é interrompido e iniciado novamente após as análises, e automatizado de forma  
1822 que os parâmetros possam ser facilmente alterados, e assim o processo possa ser repetido  
1823 pelo número desejado de vezes com mínima intervenção do usuário ou pesquisador. No  
1824 entanto, algumas limitações e pontos de melhoria devem ser considerados.

1825 O primeiro ponto de melhoria diz respeito à simulação de dados. Apesar das  
1826 diversas tentativas de se encontrar os melhores parâmetros para se desenvolver uma  
1827 população bovina de corte fidedigna – isto é, com perfis de desequilíbrio de ligação e de  
1828 corridas de homozigosidade similares aos encontrados na literatura – os resultados da  
1829 validação não se aproximaram tanto ao desejado. Para que tal objetivo seja atingido, é  
1830 essencial a simulação de um número maior de animais, sobretudo na população recente.  
1831 No presente estudo, o número de animais nas populações recentes e experimentais foi  
1832 limitado devido a uma restrição de uso do Software BLUPF90, e uma alternativa para  
1833 contornar tal limitação será proposta a seguir.

1834 Conforme ilustrado e discutido ao final da seção 3.4.3. (Figura 22), um grande  
1835 número de combinações é gerado ao computar as PHASM de todos os acasalamentos  
1836 possíveis nas populações experimentais. Ao considerar um total de 50 acasalamentos,  
1837 onde 50 novos indivíduos são gerados a cada geração, e mais 48 indivíduos de gerações  
1838 anteriores são considerados como candidatos, um número próximo a 8000 PHASM é  
1839 gerado. No início das populações experimentais (geração 0), a avaliação genômica é feita  
1840 com apenas 98 animais na população de referência, e 8000 indivíduos são alocados à

1841 população de teste. Ao final das 75 gerações, obtém-se aproximadamente 12000  
1842 indivíduos, onde 4000 compõem a população referência.

1843 Com a rápida progressão do número de PHASM geradas para cada indivíduo  
1844 adicional da população experimental, rapidamente atinge-se o limite de animais  
1845 genotipados permitido pelo Software BLUPF90, que é disponível gratuitamente para fins  
1846 acadêmicos, porém limita a inserção de até 25.000 genótipos. Como forma de contornar tal  
1847 limitação, seria possível realizar a estimação dos efeitos dos marcadores a partir dos  
1848 animais candidatos, e aplicar tais efeitos na estimação dos valores genéticos das PHASM.  
1849 Além disso, é necessário o cálculo da matriz genômica para obter-se o coeficiente de  
1850 endogamia genômico das PHASM.

1851 Com esta alteração nos passos de predição dos valores genéticos e a inclusão do  
1852 cálculo da matriz G, seria possível aumentar o número de indivíduos na população  
1853 experimental, aproximando-se mais do número de indivíduos presentes em rebanhos  
1854 comerciais de gado de corte no Brasil. O aumento do tamanho populacional por sua vez  
1855 interfere no número de combinações possíveis de acasalamentos, o que possibilita uma  
1856 maior intensidade de seleção e a execução de acasalamentos entre animais menos  
1857 aparentados entre si.

1858 Além do tamanho populacional, algumas outras modificações poderiam ser feitas  
1859 na simulação de dados de forma a aproximar ainda mais os dados de um cenário realista.  
1860 No presente estudo, apenas uma característica foi simulada, uma vez que pretendeu-se  
1861 focar na execução e viabilidade do pipeline completo de simulação e análises, evitando-se  
1862 complexidades adicionais. Ao inserir um número maior de características, possivelmente  
1863 com efeitos aditivos e não aditivos, seria possível avaliar a introdução de um index de  
1864 seleção e a viabilidade da implementação de tal pipeline em um cenário mais próximo da  
1865 realidade da pecuária de corte do Brasil.

1866 Por fim, a simulação de dados oferece uma gama quase infinita de possibilidades.  
1867 Muitos outros cenários podem ser testados, e o desenvolvimento de estudos futuros se  
1868 beneficia em larga escala ao serem iniciados a partir de um pipeline básico porém bem  
1869 estruturado, e que comporta modificações e adições com baixa dificuldade associada, ao

1870 comparar-se com o desenvolvimento de tal estudo sem uma base pré-estabelecida.  
1871 Ademais, alguns elementos adicionais podem ser explorados, tais quais as posições e  
1872 efeitos dos QTL fornecidos pelo simulador, a criação de grupos (como fazendas), e a  
1873 utilização dos perfis de corridas de homozigidade de maneira mais pormenorizada.

### 1874 3.5 CONCLUSÃO

1875 Neste trabalho, foi desenvolvido um pipeline de simulação de dados e análises  
1876 genômicas, permitindo a implementação de diferentes estratégias de seleção e  
1877 direcionamento de acasalamentos ao longo de diversas gerações, sem a necessidade de  
1878 monitoramento ou interferência do pesquisador durante o seu desenvolvimento. A  
1879 população simulada apresentou características semelhantes às das populações bovinas  
1880 reais, no entanto com a necessidade de alguns ajustes em relação ao nível de desequilíbrio  
1881 de ligação e homozigidade.

1882 Entre as estratégias de seleção e direcionamento de acasalamentos testadas,  
1883 foram observados diferentes efeitos em relação às métricas observadas. Em relação à  
1884 característica simulada, observou-se que os valores genéticos verdadeiros e os fenótipos  
1885 foram consistentemente superiores quando a estratégia GEBV + Froh foi adotada. O Fg foi  
1886 mais baixo quando utilizada a estratégia GEBV + Fg, no entanto a mesma estratégia  
1887 promoveu a formação de corridas de homozigidade maiores e mais frequentes no  
1888 genoma. A fim de minimizar-se a porcentagem do genoma coberta por ROHs, a estratégia  
1889 mais eficiente foi a EBV + Fped, seguida pela GEBV + Froh.

1890 Um dos grandes fatores limitantes deste estudo foi o tamanho populacional  
1891 utilizado, que não reflete o tamanho dos rebanhos comerciais do Brasil, e ainda impõe uma  
1892 maior dificuldade de se evitar a realização de acasalamentos endogâmicos na população.  
1893 Esta limitação foi imposta devido ao uso das PHASM, que correspondem a um número  
1894 muito grande de combinações mesmo para uma pequena população. Tal limitação poderia  
1895 ser contornada através da estimação dos efeitos dos marcadores genéticos para uma  
1896 população referência e aplicação de tais valores nas PHASM, além do cálculo manual da  
1897 matriz G, de forma que o número de PHASM não atinja o limite permitido pelo Software  
1898 BLUPF90.

1899 Por fim, o pipeline aqui apresentado serve como uma base para próximos estudos  
 1900 cujo objetivo seja mimetizar uma população bovina e aplicar diferentes estratégias de  
 1901 seleção, não contando apenas com as funcionalidades básicas do simulador, mas fazendo-  
 1902 se o uso de múltiplas ferramentas e software utilizados na indústria.

### 1903 3.6 REFERÊNCIAS

- 1904 ALLAIRE, F. R. Mate selection by selection index theory. **Theoretical and Applied**  
 1905 **Genetics**, v. 57, n. 6, p. 267–272, 1980.
- 1906
- 1907 BISCARINI, F.; COZZI, P.; GASPA, G.; MARRAS, G. **detectRUNS: an R package to**  
 1908 **detect runs of homozygosity and heterozygosity in diploid genomes**CRAN (The  
 1909 **Comprehensive R Archive Network) 2018. Disponível em: <[https://cran.r-](https://cran.r-project.org/web/packages/detectRUNS/vignettes/detectRUNS.vignette.html#references)**  
 1910 **project.org/web/packages/detectRUNS/vignettes/detectRUNS.vignette.html#references**>.
- 1911
- 1912 BOURDON, R. M. **Understanding Animal Breeding**. Harlow: Pearson, 2014.
- 1913
- 1914 BRITO, F. V.; NETO, J. B.; SARGOLZAEI, M.; COBUCI, J. A.; SCHENKEL, F. S.  
 1915 Accuracy of genomic selection in simulated populations mimicking the extent of linkage  
 1916 disequilibrium in beef cattle. **BMC Genetics**, v. 12, 2011.
- 1917
- 1918 CEBALLOS, F. C.; JOSHI, P. K.; CLARK, D. W.; RAMSAY, M.; WILSON, J. F. Runs of  
 1919 homozygosity: Windows into population history and trait architecture. **Nature Reviews**  
 1920 **Genetics**, v. 19, n. 4, p. 220–234, 2018.
- 1921
- 1922 CLARK, S. A.; KINGHORN, B. P.; HICKEY, J. M.; VAN DER WERF, J. H. J. The effect of  
 1923 genomic information on optimal contribution selection in livestock breeding programs.  
 1924 **Genetics Selection Evolution**, v. 45, n. 1, p. 1–8, 2013.
- 1925
- 1926 DAETWYLER, H. D.; VILLANUEVA, B.; BIJMA, P.; WOOLLIAMS, J. A. Inbreeding in  
 1927 genome-wide selection. **Journal of Animal Breeding and Genetics**, v. 124, n. 6, p. 369–  
 1928 376, 2007.
- 1929
- 1930 DE CLERCQ, Matthieu; VATS, Anshu.; BIEL, Alvaro. **Agriculture 4.0: The Future of**  
 1931 **Farming Technology**. In: World Government Summit. Feb. 2018. Disponível emkk:  
 1932 <[https://www.marshmcclennan.com/content/dam/mmc-](https://www.marshmcclennan.com/content/dam/mmc-web/insights/publications/2018/november/agriculture-4-0/Oliver-Wyman-Agriculture-4.0.pdf)  
 1933 **web/insights/publications/2018/november/agriculture-4-0/Oliver-Wyman-Agriculture-**  
 1934 **4.0.pdf**>. Acesso em: 29 Aug. 2022.
- 1935
- 1936 DE SOUZA FONSECA, P. A.; DOS SANTOS, F. C.; ROSSE, I. C.; VENTURA, R. V.;  
 1937 BRUNELLI, F. Â. T.; PENNA, V. M.; DA SILVA VERNEQUE, R.; MACHADO, M. A.; DA  
 1938 SILVA, M. V. G. B.; CARVALHO, M. R. S.; PEIXOTO, M. G. C. D. Retelling the recent  
 1939 evolution of genetic diversity for Guzerá: Inferences from LD decay, runs of homozygosity  
 1940 and Ne over the generations. **Livestock Science**, v. 193, n. October, p. 110–117, nov.  
 1941 2016.

- 1942
- 1943 DICKERSON, G. E. Experimental evaluation of selection theory in poultry. In:
- 1944 INTERNATIONAL CONGRESS OF GENETICS, 11., 1963, The Hague. **Proceedings...**
- 1945 The Hague: Pergmon, 1963. p. 747–761. (Genetics Today, v. 1)
- 1946
- 1947 FAUX, A.; GORJANC, G.; GAYNOR, R. C.; BATTAGIN, M.; EDWARDS, S. M.; WILSON,
- 1948 D. L.; HEARNE, S. J.; GONEN, S.; HICKEY, J. M. AlphaSim: Software for Breeding
- 1949 Program Simulation. **The Plant Genome**, v. 9, n. 3, p. 1–14, 2016.
- 1950
- 1951 GAYNOR, R. C.; GORJANC, G.; HICKEY, J. M. AlphaSimR: an R package for breeding
- 1952 program simulations. **G3 Genes|Genomes|Genetics**, v. 11, n. 2, 12 abr. 2021. Disponível
- 1953 em: <<https://academic.oup.com/g3journal/article/doi/10.1093/g3journal/jkaa017/6025179>>.
- 1954 Acesso em: 31 mar. 2022.
- 1955
- 1956 HAYES, B. J.; BOWMAN, P. J.; CHAMBERLAIN, A. J.; GODDARD, M. E. Invited review:
- 1957 Genomic selection in dairy cattle: Progress and challenges. **Journal of Dairy Science**, v.
- 1958 92, n. 2, p. 433–443, 2009.
- 1959
- 1960 HICKEY, J. M.; GORJANC, G. Simulated data for genomic selection and genome-wide
- 1961 association studies using a combination of coalescent and gene drop methods. **G3:**
- 1962 **Genes, Genomes, Genetics**, v. 2, n. 4, p. 425–427, 2012.
- 1963
- 1964 HOWARD, J. T.; PRYCE, J. E.; BAES, C.; MALTECCA, C. Invited review: Inbreeding in
- 1965 the genomics era: Inbreeding, inbreeding depression, and management of genomic
- 1966 variability. **Journal of Dairy Science**, v. 100, n. 8, p. 6009–6024, 2017a.
- 1967
- 1968 HOWARD, J. T.; TIEZZI, F.; PRYCE, J. E.; MALTECCA, C. Geno-Diver: A combined
- 1969 coalescence and forward-in-time simulator for populations undergoing selection for
- 1970 complex traits. **Journal of Animal Breeding and Genetics**, v. 134, n. 6, p. 553–563,
- 1971 2017b.
- 1972
- 1973 Microsoft Power BI. Data visualisation. Disponível em: [https://powerbi.microsoft.com/en-](https://powerbi.microsoft.com/en-gb/)
- 1974 [gb/](https://powerbi.microsoft.com/en-gb/) (acesso em 07 de março de 2024).
- 1975
- 1976 MISZTAL, I.; TSURUTA, S.; LOURENCO, D.A.L.; AGUILAR, I.; LEGARRA, A.; VITEZICA,
- 1977 Z. **Manual for BLUPF90 family of programs**. 2014. Disponível em:
- 1978 [http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90\\_all2.pdf](http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all2.pdf). Acesso em: 08 Mar.
- 1979 2024.
- 1980
- 1981 NEVES, H. H. R.; CARVALHEIRO, R.; O'BRIEN, A. M. P.; UTSUNOMIYA, Y. T.; DO
- 1982 CARMO, A. S.; SCHENKEL, F. S.; SÖLKNER, J.; MCEWAN, J. C.; VAN TASSELL, C. P.;
- 1983 COLE, J. B.; DA SILVA, M. V. G. B.; QUEIROZ, S. A.; SONSTEGARD, T. S.; GARCIA, J.
- 1984 F. Accuracy of genomic predictions in Bos indicus (Nellore) cattle. **Genetics Selection**
- 1985 **Evolution**, v. 46, n. 1, 27 fev. 2014.
- 1986
- 1987 PORTO-NETO, L. R.; KIJAS, J. W.; REVERTER, A. The extent of linkage disequilibrium in
- 1988 beef cattle breeds using high-density SNP genotypes. **Genetics Selection Evolution**, v.
- 1989 46, n. 1, p. 1–5, 2014.

- 1990
- 1991 R CORE TEAM. **R: A Language and Environment for Statistical Computing** Vienna, Austria. R Foundation for Statistical Computing, 2020. Disponível em: <<https://www.r-project.org/>>.
- 1992
- 1993
- 1994
- 1995 SARGOLZAEI, M.; SCHENKEL, F. S. QMSim: A large-scale genome simulator for livestock. **Bioinformatics**, v. 25, n. 5, p. 680–681, 2009.
- 1996
- 1997
- 1998 SARGOLZAEI, M.; SCENKEL, F. S.; JANSEN, G. B.; SCHAEFFER, L. R. Extent of linkage disequilibrium in Holstein cattle in North America. **Journal of Dairy Science**, v. 91, n. 5, p. 2106–2117, 2008.
- 1999
- 2000
- 2001
- 2002 SELLI, A.; BUSSIMAN, F. de O.; SOUZA, W. L. de; ANDRIETTA, L. T.; PASCHOAL, V. R.; PINTO, D. L.; BALIEIRO, J. C. de C.; VENTURA, R. V. Perfis De Corridas De Homozigosidade E Heterozigosidade Em Populações Bovinas Com Diferentes Combinações Genóticas Provenientes De Dados Simulados. *Em*: GOBESSO, A. A. DE O.; SANTOS, M. V. DOS; GAMEIRO, A. H.; GARBOSSA, C. A. P.; BALIEIRO, J. C. DE C.; BRUNETTO, M. A.; VENDRAMINI, T. H. A. **Novos Desafios Da Pesquisa Em Nutrição E Produção Animal**. Pirassununga: Editora 5D, 2021. p. 231–254.
- 2003
- 2004
- 2005
- 2006
- 2007
- 2008
- 2009
- 2010 UNITED NATIONS. **World Population Prospects 2022: Summary of Results**. 2022. Disponível em: <[https://www.un.org/development/desa/pd/sites/www.un.org.development.desa.pd/files/wp2022\\_summary\\_of\\_results.pdf](https://www.un.org/development/desa/pd/sites/www.un.org.development.desa.pd/files/wp2022_summary_of_results.pdf)>. Acesso em: 29 Aug. 2022.
- 2011
- 2012
- 2013
- 2014
- 2015 VANRADEN, P. M.; VAN TASSELL, C. P.; WIGGANS, G. R.; SONSTEGARD, T. S.; SCHNABEL, R. D.; TAYLOR, J. F.; SCHENKEL, F. S. Invited review: Reliability of genomic predictions for North American Holstein bulls. **Journal of Dairy Science**, v. 92, n. 1, p. 16–24, 2009.
- 2016
- 2017
- 2018
- 2019
- 2020 WICKHAM, H. **ggplot2: Elegant Graphics for Data Analysis**. Springer-Verlag New York, 2016.
- 2021
- 2022
- 2023 WICKHAM, H.; FRANÇOIS, R.; HENRY, L.; MÜLLER, K.; VAUGHAN, D. **dplyr: A Grammar of Data Manipulation**. R package version 1.1.3. 2023. Disponível em: <<https://CRAN.R-project.org/package=dplyr>>. Acesso em: 08 Mar. 2024.
- 2024
- 2025
- 2026
- 2027 WICKHAM, H.; VAUGHAN, D.; GIRLICH, M. **tidyr: Tidy Messy Data**. R package version 1.3.0. 2023. Disponível em: <<https://CRAN.R-project.org/package=tidyr>>. Acesso em: 08 Mar. 2024.
- 2028
- 2029
- 2030
- 2031 WRIGHT, S. Coefficients of Inbreeding and Relationship. **The American Naturalist**, v. 56, n. 645, p. 330–338, jul. 1922.
- 2032
- 2033

## 2034 4 O USO DE VISUALIZAÇÕES INTERATIVAS PARA RASTREAR HERANÇA 2035 HAPLOTÍPICA NA PECUÁRIA

2036 Os dados apresentados neste capítulo são um resumo da publicação feita na revista  
2037 internacional Ruminants (SELLI; MILLER; VENTURA, 2024), que pode ser acessada pelo  
2038 link <https://www.mdpi.com/2673-933X/4/1/6>.

### 2039 4.1 RESUMO

2040 Nosso objetivo foi utilizar o poder das visualizações interativas, através de  
2041 ferramentas de código aberto, para o desenvolvimento de uma estratégia eficiente que  
2042 permita visualizar dados de polimorfismo de nucleotídeo único (SNP, do inglês Single  
2043 Nucleotide Polymorphism) em uma população bovina, com foco no rastreamento da  
2044 transmissão de haplótipos. Assim, simulamos uma população realista de gado de corte, a  
2045 fim de obter haplótipos faseados e gerar os insumos necessários para a criação de nossas  
2046 visualizações. A ferramenta de visualização foi construída em Python, com o uso da  
2047 biblioteca Plotly, que permite a criação de visualizações interativas. Dois cenários foram  
2048 explorados: comparação de trios e visualização de avós. Esses cenários nos permitiram  
2049 rastrear a herança de segmentos genéticos, identificar eventos de crossover e descobrir  
2050 regiões comuns em animais aparentados e não aparentados. As potenciais aplicações  
2051 desta abordagem são significativas, particularmente para melhorar a seleção genômica em  
2052 programas de melhoramento e fazendas menores e fornecer informações valiosas para  
2053 orientar análises mais aprofundadas de regiões genômicas específicas. Além das suas  
2054 aplicações práticas, acreditamos que esta estratégia pode ser uma ferramenta educacional  
2055 valiosa, ajudando os educadores a esclarecer conceitos complexos como a amostragem  
2056 mendeliana e a diversidade haplotípica. Além disso, esperamos encorajar os produtores  
2057 pecuários a adotar tecnologias avançadas como a genotipagem e a seleção genômica,  
2058 contribuindo assim para o avanço da genética pecuária.

### 2059 4.2 INTRODUÇÃO

2060 Sabe-se que irmãos completos compartilham aproximadamente 50% de conteúdo  
2061 genômico entre si (OLDENBROEK; VAN DER WAAIJ, 2014). No entanto, esses genes não



2062 são distribuídos aleatoriamente em seus genomas, mas sim agrupados em segmentos ou  
2063 cromossomos inteiros que são herdados de cada genitor (BOHMANOVA; SARGOLZAEI;  
2064 SCHENKEL, 2010). Além disso, esse percentual pode variar, sendo possível que alguns  
2065 irmãos sejam mais parecidos que outros (OLDENBROEK; VAN DER WAAIJ, 2014). O fato  
2066 de muitos segmentos genômicos não expressarem características que possam ser  
2067 facilmente percebidas também contribui para a percepção da semelhança entre parentes.  
2068 Na produção pecuária, os produtores podem ter certas expectativas em relação à progênie  
2069 obtida, especialmente quando se emprega a seleção genômica. No entanto, os resultados  
2070 por vezes diferem das previsões, devido à amostragem Mendeliana e outros fatores (BRITO  
2071 et al., 2011). Portanto, rastrear segmentos genômicos compartilhados entre indivíduos seria  
2072 de interesse para criadores e produtores que desejem saber quais regiões, ou, mais  
2073 genericamente, quais segmentos cromossômicos foram passados de pais para filhos, e até  
2074 mesmo para auxiliar na seleção de potenciais parceiros.

2075           As representações visuais são comprovadamente fundamentais no processo de  
2076 aprendizagem e comunicação de conceitos científicos, uma vez que tornam visíveis os  
2077 conceitos que são abstratos e não podem ser experimentados diretamente (COOK, 2006;  
2078 SELLI et al., 2021), como herança genética e amostragem Mendeliana. O uso de  
2079 visualizações aproveita o aprendizado e experiências em torno de eventos relacionados à  
2080 exploração do sistema visual humano, que evoluiu ao longo de milhões de anos e hoje é  
2081 capaz de capturar imensas quantidades de informações sobre o mundo de forma rápida e  
2082 eficiente (VAN ESSEN; ANDERSON; FELLEMAN, 1992). Assim, as representações visuais  
2083 têm um papel importante na compreensão de dados complexos, potencializando a  
2084 capacidade humana para a identificação de padrões, relações e tendências (DOSHER; LU,  
2085 2017), que podem não ser facilmente discerníveis em formatos tabulares ou textuais.

2086           A representação visual de um conjunto de dados específico oferece vários  
2087 caminhos para aprimoramento, sendo que fatores como o público-alvo, o meio de  
2088 apresentação adotado e a mensagem que se pretende passar desempenham um papel  
2089 fundamental na determinação de elementos como seleção de cores, tipo de gráfico, nível  
2090 de detalhe e muito mais (ROUGIER; DROETTBOOM; BOURNE, 2014). Por outro lado, o  
2091 uso errado de tais elementos pode ser enganoso, podendo até mesmo atrapalhar o  
2092 julgamento dos espectadores (BATEMAN et al., 2018; HULLMAN; DIAKOPOULOS, 2011).

2093 Vários autores descreveram como diferentes aspectos da visualização podem ser  
2094 abordados, com meios para criar figuras melhores: Zhou e Hansen (ZHOU; HANSEN, 2016)  
2095 forneceram uma revisão abrangente das técnicas de geração de mapas de cores, para  
2096 ajudar os leitores nas decisões de mapeamento de cores. Zacks e Tversky (ZACKS;  
2097 TVERSKY, 1999) relataram que os gráficos de barras são geralmente associados a  
2098 comparações discretas, enquanto os gráficos de linhas são mais comumente associados a  
2099 tendências e dados temporais, e que a mistura desses tipos de representação pode causar  
2100 confusão aos usuários. Hullman e Diakopoulos (2011) sugeriram que a escolha de quais  
2101 dados serão apresentados poderia ser confiada ao usuário, habilitando a interatividade por  
2102 meio da inserção de elementos como botões e barras de pesquisa.

2103 O interesse em ferramentas de visualização interativa tem aumentado na academia  
2104 e na indústria, uma vez que elas facilitam a comunicação entre pesquisadores, analistas e  
2105 o público em geral (ALI et al., 2016; MOROTA et al., 2021; SELLI et al., 2021). Uma decisão  
2106 crítica está relacionada à utilização de ferramentas de código aberto ou proprietárias ao  
2107 desenvolver aplicativos de visualização. Exemplos de ferramentas de código aberto que  
2108 podem ser usadas para criar visualizações interativas são as bibliotecas Plotly e Dash, que  
2109 podem ser implementadas em Python (VAN ROSSUM; DRAKE JR, 1995), R (R CORE  
2110 TEAM, 2020), ou por meio de outras linguagens de programação. Alguns benefícios da  
2111 utilização de ferramentas de código aberto incluem a relação custo-benefício, a flexibilidade  
2112 e a possibilidade de contribuir e receber ajuda da comunidade. Por outro lado, existe a  
2113 desvantagem de haver extensas opções de personalização, o que pode levar a tempos de  
2114 desenvolvimento mais longos e a maior complexidade na criação e manutenção de  
2115 visualizações. Além disso, essas ferramentas geralmente apresentam uma curva de  
2116 aprendizado mais acentuada em comparação com soluções comerciais que são mais  
2117 fáceis de usar.

2118 Exemplos de tais soluções comerciais são as ferramentas Tableau e Power BI, que  
2119 fornecem interface fácil de usar e excelente integração de dados. No entanto, também  
2120 existem desvantagens, como custos de licenciamento, personalização restrita e  
2121 dependência do cronograma de atualização do fornecedor (ALI et al., 2016; MOROTA et  
2122 al., 2021; SELLI et al., 2021)). Conforme discutido por Curti et al. (2023), a integração de  
2123 dados é uma etapa crucial para análise e visualização de dados. Os desafios colocados

2124 nesta etapa dependem muito da estrutura dos dados a serem integrados, por exemplo, se  
2125 são dados estruturados, semiestruturados ou não estruturados, e se existe uma ligação  
2126 imediata entre as fontes de dados. A depender da estrutura de dados, pode haver a  
2127 necessidade de construção de procedimentos específicos para gerar os inputs necessários  
2128 para alimentar as visualizações.

2129 Muitas ferramentas foram desenvolvidas para visualizar dados genômicos, como  
2130 Jbrowse (BUELS et al., 2016) e outras ferramentas de visualização responsivas baseadas  
2131 na web (L'YI; GEHLENBORG, 2023). Uma apresentação recente no PAG XXXI (DEEB et  
2132 al., 2024) demonstrou uma ferramenta online denominada Chromosome Mating, que usa  
2133 informações genômicas para otimizar decisões de acasalamento, fazer previsões precisas  
2134 sobre valores genéticos e restringir a endogamia, entre outras restrições escolhidas pelos  
2135 usuários. Essa apresentação evidenciou como a indústria pode se beneficiar ao associar  
2136 tecnologia de ponta a uma interface interativa fornecida como ferramenta online, permitindo  
2137 aos agricultores tomar decisões de acasalamento baseadas em dados. Como continuação  
2138 destes estudos, pretendemos aplicar o conceito de Business Intelligence, por meio de  
2139 ferramentas de código aberto, para criar uma estratégia interativa e eficaz para visualizar  
2140 dados de Polimorfismo de Nucleotídeo Único (SNP) compartilhados entre indivíduos e  
2141 rastrear a transmissão de haplótipos em uma população pecuária. Larkin et al. (2012)  
2142 desenvolveram uma estratégia intitulada “haplotracking”, combinando resequenciamento  
2143 do genoma completo dos ancestrais, reconstrução de haplótipos e genotipagem de alta  
2144 densidade dos descendentes, a fim de identificar os alelos dos ancestrais que foram  
2145 submetidos à seleção artificial, além de detectar possíveis mutações causais que estão  
2146 próximas às regiões de interesse econômico – Quantitative Trait Loci (QTL).

2147 Nossa proposta neste estudo também é “rastrear haplótipos”, porém, com  
2148 abordagens e objetivos diferentes. Uma população realista de gado de corte foi simulada e  
2149 haplótipos faseados foram obtidos, com o objetivo de testar uma nova abordagem de  
2150 visualização, permitindo observar as semelhanças entre os indivíduos de uma população e  
2151 tornar evidentes as vantagens do emprego de visualizações interativas. A simulação dos  
2152 dados foi fundamental para a obtenção de famílias genotipadas com uma mesma  
2153 densidade de marcadores, além de validar os resultados obtidos em uma população na  
2154 qual a estrutura é conhecida. Além disso, também tivemos como objetivo gerar entradas

2155 específicas a partir de dados genômicos faseados, a fim de fornecer os meios para criar  
2156 tais visualizações, construindo um pipeline inexistente anteriormente para geração das  
2157 janelas de haplótipos.

2158 O uso de ferramentas de código aberto, ao contrário de estudos anteriores  
2159 desenvolvidos pelos autores (SELLI et al., 2021), proporcionou a flexibilidade necessária  
2160 para o desenvolvimento de gráficos customizáveis projetados para meios específicos. Com  
2161 essa estratégia, esperamos encorajar a adoção de tecnologias como a genotipagem e a  
2162 seleção genômica pelos produtores, tornando mais fácil a explicação de perguntas  
2163 frequentes conduzidas por eles, como "por que muitas vezes irmãos completos são tão  
2164 diferentes uns dos outros?", além de ilustrar as diferenças entre populações e linhas de  
2165 seleção. Adicionalmente, esta abordagem poderia ajudar educadores durante a explicação  
2166 de conceitos complexos, como amostragem mendeliana e diversidade haplotípica.

## 2167 4.3 MATERIAL E MÉTODOS

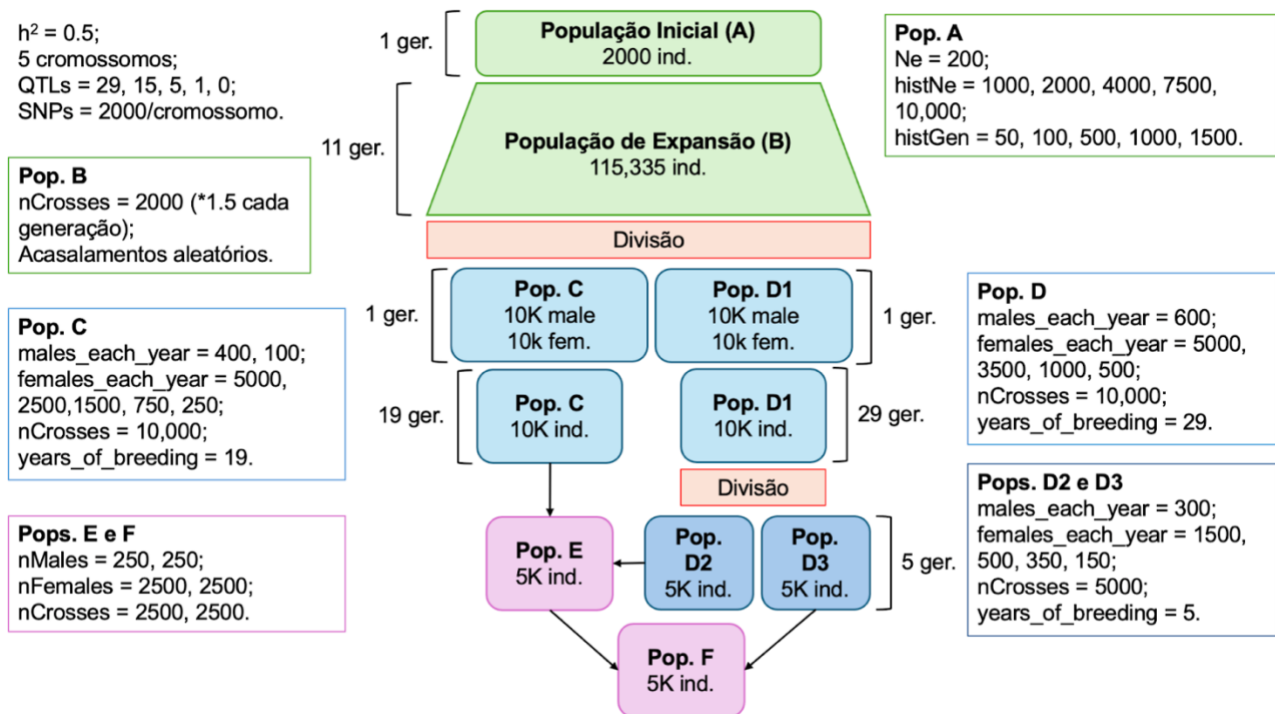
### 2168 4.3.1 Simulação e Validação dos dados

2169 Uma população de gado de corte foi simulada usando o pacote R (R CORE TEAM,  
2170 2020) AlphaSimR (GAYNOR; GORJANC; HICKEY, 2021). A função runMacs2 foi utilizada  
2171 para criar 2.000 genomas fundadores com 5 cromossomos de comprimento 100 Mb cada,  
2172 simulando uma população de tamanho efetivo de 200. Para simular a evolução histórica  
2173 desta população, os parâmetros histNe e histGen foram configurados para os respectivos  
2174 vetores: (1.000, 2.000, 4.000, 7.500 e 10.000) e (50, 100, 500, 1.000 e 1.500). Foi simulada  
2175 uma característica de herdabilidade 0,5, média 0 e variabilidade 1, com 29, 15, 5, 1 e 0 QTL  
2176 distribuídos pelos 5 cromossomos, respectivamente. Foi simulado um mapa SNP com  
2177 2.000 marcadores em cada cromossomo, com um total de 10.000 marcadores no genoma.  
2178 Os mapas QTL e SNP foram exportados para uso posterior.

2179 A simulação foi realizada conforme ilustrado na Figura 1 e serviu para estudos  
2180 adicionais além deste. Assim, nem todas as populações aqui descritas foram incluídas nos  
2181 cenários apresentados a seguir, mas constituem uma parte importante da simulação  
2182 completa. A população inicial (A), com 2.000 indivíduos, foi criada e acasalada

2183 aleatoriamente para formar a população expandida (B). A população B foi acasalada  
 2184 aleatoriamente por 11 gerações, com o número de cruzamentos aumentando em 50% a  
 2185 cada geração. As populações C e D1 foram derivadas da última geração de B,  
 2186 selecionando-se aleatoriamente 10 mil machos e 10 mil fêmeas (sem sobreposição) para  
 2187 cada uma. A população C foi selecionada por fenótipo por 19 gerações, enquanto a  
 2188 população D1 foi selecionada por 29 gerações. A população D1 foi dividida em D2 e D3,  
 2189 sendo cada uma selecionada por 5 gerações. A população E foi resultado do cruzamento  
 2190 de C e D2, e a F foi criada pelo cruzamento de E e D3. A partir da última geração de cada  
 2191 população, foram obtidos genótipos faseados, fenótipos e informações de pedigree.

2192 Figura 4-1 – Estrutura da simulação de populações de bovinos de corte.



2193

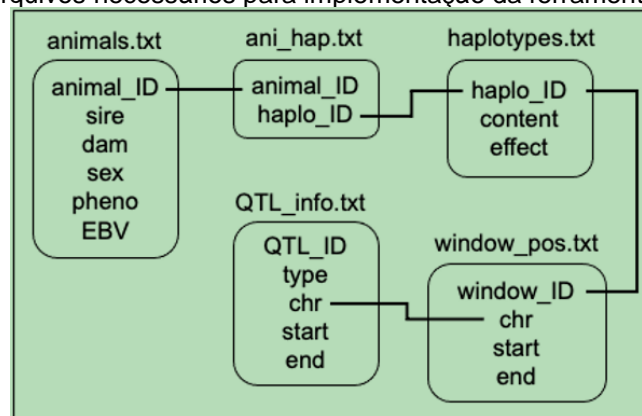
2194 Fonte: Autoria própria.

2195 A validação da simulação foi realizada analisando o desequilíbrio de ligação (LD)  
 2196 em amostras de 100 indivíduos na última geração de cada população. O  $r^2$  entre todos os  
 2197 pares de marcadores em cada cromossomo foi calculado via Plink v1.9 [26], enquanto  
 2198 estatísticas e gráficos de decaimento LD foram gerados em R. Uma Análise de  
 2199 Componentes Principais (PCA, do inglês Principal Components Analysis) também foi  
 2200 realizada nas mesmas populações em uma combinação de amostras (N = 700), usando o  
 2201 pacote R flashpcaR v2.1 (ABRAHAM; QIU; INOUE, 2017).

### 4.3.2 Regressão de SNP Único e Preparação de Dados

Para obter os efeitos dos marcadores, uma regressão utilizando apenas um SNP por vez (SSR, do inglês Single SNP Regression) foi implementada em R, adaptando o código fornecido por Gondro (GONDRO, C., 2015. pp. 111–124). Os genótipos utilizados foram obtidos de toda a última geração da população D2 (N = 5.000) e os alelos fixos foram removidos, restando um total de 9.628 SNPs (de 10.000) para o SSR. Além disso, o mapa contendo as posições e coordenadas referentes a cada SNP, os efeitos calculados dos SNPs, o pedigree e os haplótipos faseados foram processados posteriormente em Python, para a criação de cinco (5) arquivos, correspondentes à entrada de dados da ferramenta de visualização (Figura 2).

Figura 4-2 – Relação de arquivos necessários para implementação da ferramenta de visualização.



Fonte: Autoria própria.

O arquivo “QTL\_info.txt” contém os IDs de QTL e seu respectivo tipo (a característica afetada), cromossomo, posições inicial e final. O arquivo “window\_pos.txt” agrega N marcadores do mapa SNP de acordo com o número de SNPs especificado para o tamanho da janela. Para este estudo, adotamos tamanhos de janela de 25 SNPs (2.000 SNPs por cromossomo/25 = 80 janelas em cada cromossomo). Cada janela possui um ID e informações sobre seus cromossomos, posições inicial e final. Outro arquivo, “haplótipos.txt”, contém todas as combinações de alelos encontrados em cada janela dos haplótipos faseados. Cada haplótipo possui um ID, conteúdo (as combinações de alelos) e seu respectivo efeito (calculado como a soma dos efeitos do marcador com base na presença ou não do alelo com o efeito calculado). O arquivo “ani\_hap.txt” corresponde a uma matriz de tamanho  $A \times 2 \times W$ , onde A é o número de animais (vezes 2 cadeias de

2226 cromossomos) e  $W$  é o número de janelas no genoma. O conteúdo da matriz é o ID do  
2227 haplótipo para cada respectiva cadeia e janela animal. Já o arquivo “animals.txt” contém  
2228 IDs de animais e seus respectivos pais, mães, sexo, fenótipo e valor genético.

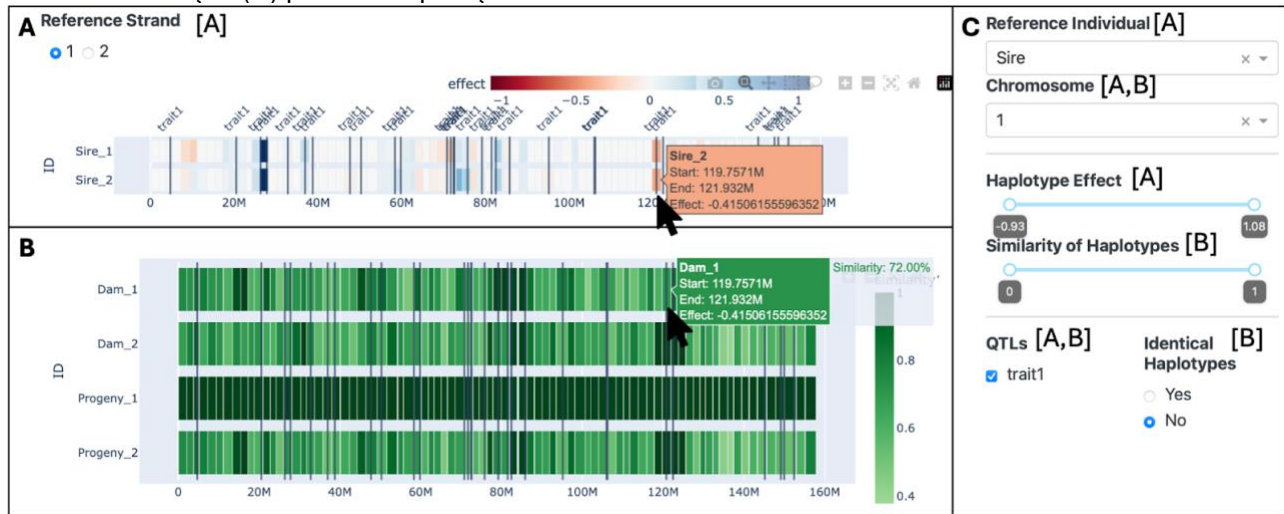
### 2229 **4.3.3 Ferramenta de Visualização**

2230 A ferramenta de visualização aqui apresentada foi construída em Python. Todos os  
2231 componentes visuais e interativos foram adicionados com as bibliotecas Plotly e Dash, e a  
2232 manipulação dos dados foi realizada por meio da biblioteca Pandas (MCKINNEY, 2010).  
2233 Conforme ilustrado no trabalho publicado, na Figura 3 (SELLI; MILLER; VENTURA, 2024),  
2234 após carregar cada um dos 5 conjuntos de dados descritos na seção anterior, um animal é  
2235 escolhido como “indivíduo de referência”, ou seja, o animal de referência a ser comparado  
2236 com todos os outros animais. Por padrão, o primeiro animal do conjunto de dados é  
2237 selecionado, assim como a primeira fita do cromossomo. A seguir, para cada uma das  
2238 janelas (haplótipos), a similaridade entre o animal e fita de referência é comparada com a  
2239 janela correspondente de cada indivíduo (ambas as fitas) da população. A similaridade é  
2240 calculada dividindo-se o número de alelos idênticos pelo número total de alelos no  
2241 haplótipo.

2242 A Figura 3 mostra a interface da ferramenta de visualização. Os haplótipos do  
2243 indivíduo de referência são plotados no topo da página (Figura 3, Seção A), coloridos por  
2244 seus efeitos – variando de vermelho escuro (valores negativos) a branco (sem efeito) e azul  
2245 escuro (valores positivos). Ao passar o mouse sobre os haplótipos do indivíduo de  
2246 referência é possível visualizar o ID do indivíduo de referência, o início, fim e efeito de cada  
2247 haplótipo. Na segunda seção da tela (Figura 3, Seção B), todos os outros haplótipos dos  
2248 animais são plotados. São coloridos de acordo com a semelhança com o respectivo  
2249 haplótipo do indivíduo de referência. O verde claro é menos semelhante e o verde escuro  
2250 é mais semelhante. Quando outro indivíduo de referência é selecionado, a cor dos  
2251 haplótipos dos demais animais é atualizada para mostrar o grau de semelhança com o  
2252 indivíduo de referência recém-selecionado, e o indivíduo de referência anterior passa a  
2253 estar presente na seção com os demais indivíduos. Cada linha de haplótipos mostrados na  
2254 visualização (seja na Seção A ou B) apresenta o número 1 ou 2 após o ID do indivíduo,  
2255 representando a fita do cromossomo (1 para paterno e 2 para materno).

2256  
2257  
2258

Figura 4-3 – Tela inicial da ferramenta de visualização. A seção (A) exibe os haplótipos do animal de referência (o pai). A seção (B) mostra todos os outros indivíduos no conjunto de dados (a mãe e a progênie). A seção (C) permite a aplicação de filtros.



2259

2260  
2261

Fonte: Autoria própria.  
As letras entre colchetes mostram em qual seção os filtros são aplicados.

2262

2263

2264

2265

2266

2267

2268

2269

2270

2271

2272

2273

2274

2275

2276

2277

2278

2279

Alguns filtros podem ser aplicados. No canto superior esquerdo da tela (Figura 3, Seção A), é possível selecionar a fita do indivíduo de referência (1 para paterna e 2 para materna) a ser utilizada como referência. No lado direito (Figura 3, Seção C), é possível selecionar o animal de referência, o cromossomo (apenas um cromossomo pode ser visualizado por vez), uma faixa de efeitos de haplótipos, uma faixa de efeitos de similaridade, os tipos de QTL observados (por característica afetada) e se haplótipos idênticos devem ser destacados ou não. Quando o filtro de efeito de haplótipo é aplicado, os haplótipos do indivíduo de referência que não passam nos critérios ficam de fora da visualização, deixando um espaço em branco onde deveriam estar. Isto, por sua vez, faz com que os haplótipos na segunda seção também sejam filtrados, refletindo os haplótipos do indivíduo de referência. Por outro lado, o filtro de similaridade é aplicado diretamente aos haplótipos na segunda seção, sem afetar os haplótipos do indivíduo de referência. Os QTL selecionados são vistos como linhas verticais na posição onde os QTL foram reportados. Quando o filtro “Haplótipos Idênticos” está habilitado, ele transforma a escala de cores dos haplótipos da segunda seção em um sistema binário. Neste sistema, um haplótipo azul significa uma combinação perfeita com o haplótipo paterno do indivíduo de referência, enquanto um haplótipo cinza indica a presença de pelo menos um alelo diferente. Se a segunda fita do indivíduo de referência for selecionada, os haplótipos do



2280 outro animal são comparados com a cadeia materna do indivíduo de referência, e os  
2281 haplótipos idênticos são coloridos em vermelho.

#### 2282 **4.3.4 Cenários para ilustrar a ferramenta de visualização**

##### 2283 4.3.4.1 Comparação de trios

2284 O primeiro cenário é a comparação Trios, em que um macho e uma fêmea foram  
2285 amostrados aleatoriamente dentre a população D2 (Figura 1) e acasalados para produzir  
2286 uma progênie. Este é um cenário simples, que visa mostrar as características básicas da  
2287 ferramenta, como a observação de quais haplótipos foram herdados de cada um dos pais  
2288 e quais trechos do genoma são semelhantes considerando-se indivíduos sem ligação direta  
2289 quanto ao pedigree. Além disso, serão aplicados filtros de efeito e similaridade para  
2290 distinguir os haplótipos de maior efeito (derivados da soma dos efeitos dos marcadores no  
2291 haplótipo) na segunda fita do segundo cromossomo do touro e determinar o quão  
2292 semelhantes eles são em relação aos respectivos haplótipos dos outros indivíduos.

##### 2293 4.3.4.2 Visualização de avós

2294 O segundo cenário consistiu a visualização de avós, em que dois machos e duas  
2295 fêmeas foram selecionados aleatoriamente da população D2 e acasalados para produzir  
2296 um casal de descendentes não aparentados. Esse casal foi acasalado para produzir 10  
2297 descendentes (irmãos completos). Neste cenário, o objetivo é investigar regiões genômicas  
2298 ancestrais transmitidas por gerações mais antigas, identificar os ancestrais responsáveis  
2299 pela transmissão dessas regiões e avaliar o impacto desses segmentos genéticos na  
2300 característica observada. Além disso, também compararemos haplótipos compartilhados  
2301 entre as três gerações e entre indivíduos não aparentados (avós).

#### 2302 4.4 RESULTADOS E DISCUSSÃO

##### 2303 4.4.1 Simulação e Validação de Dados

2304 Uma população de bovinos de corte foi simulada a fim de mostrar as vantagens do  
 2305 emprego de visualizações interativas para rastrear regiões genômicas compartilhadas entre  
 2306 indivíduos. Assim, o objetivo da simulação não foi imitar uma raça específica, mas gerar  
 2307 blocos de haplótipos semelhantes aos encontrados em uma população genérica de gado  
 2308 de corte. Sargolzaei e Schenkel (2009) descreveram as vantagens do emprego da  
 2309 simulação de dados, que incluíam os baixos custos associados, a possibilidade de observar  
 2310 – e obter dados de – várias gerações da mesma população, e poder avaliar melhor os  
 2311 resultados obtidos, uma vez que o pesquisador conhece a estrutura subjacente da  
 2312 simulação. Para validar a simulação, calculamos o LD ( $r^2$ ) entre marcadores adjacentes  
 2313 para amostras de 100 indivíduos para cada população simulada e realizamos uma análise  
 2314 de PCA em todas as amostras populacionais combinadas.

2315 A Tabela 1 mostra as estatísticas do  $r^2$  entre pares de marcadores separados por  
 2316 diferentes distâncias, calculados a partir da amostra (N = 100) da população D2, mesma  
 2317 população da qual foram obtidos os efeitos dos marcadores. O número de pares de  
 2318 marcadores analisados variou de acordo com a distância entre eles. O número de pares de  
 2319 SNP variou de 5.786 a 55.122 para distâncias mais curtas (de 0 a 50 kb) e distâncias mais  
 2320 longas (de 500 a 1.000 kb), respectivamente. Em média, para distâncias entre 100 e 500  
 2321 kb, foram observados 11.083 pares de SNPs. O  $r^2$  médio variou de 0,26 a 0,08 e o  $r^2$   
 2322 mediano, de 0,09 a 0,03, considerando as distâncias mais curta (0–50 kb) e maior (500–  
 2323 1000 kb), respectivamente. Em média, o desvio padrão em todas as distâncias foi de 0,21.  
 2324 A porcentagem de pares de marcadores com  $r^2 \geq 0,3$  diminuiu à medida que a distância  
 2325 entre os pares aumentou, variando de 0,3 a 0,06, com média de 0,15 em todas as  
 2326 distâncias.

2327 Tabela 4-1 – Desequilíbrio de ligação ( $r^2$ ) entre pares de marcadores separados por diferentes distâncias  
 2328 em uma população simulada de bovinos de corte.

Distância (kb)	Pares de SNPs	$r^2$ médio	Desvio Padrão de $r^2$	$r^2$ mediano	% de $r^2 > 0.01$	% de $r^2 > 0.03$	% de $r^2 > 0.05$
0-50	5786	0,26	0,34	0,09	0,49	0,3	0,21
050-100	5557	0,19	0,26	0,07	0,42	0,21	0,13
100-200	10943	0,15	0,22	0,05	0,38	0,17	0,09
200-300	11191	0,12	0,18	0,04	0,33	0,12	0,06
300-400	11127	0,11	0,17	0,04	0,31	0,11	0,04
400-500	11070	0,1	0,15	0,04	0,28	0,09	0,04

500-1000	55122	0,08	0,12	0,03	0,23	0,06	0,02
----------	-------	------	------	------	------	------	------

2329 Fonte: Autoria própria.

2330 As métricas de LD obtidas para a população D2 (Tabela 1) são consistentes com  
 2331 Lu et al. (2012), embora os autores tenham utilizado medidas de distância diferentes. Os  
 2332 autores relataram valores médios de  $r^2$  para três raças de gado: Angus, Charolês e  
 2333 Cruzado. Para distâncias entre 0 e 30 kb, o  $r^2$  médio foi 0,29, 0,22 e 0,21, respectivamente.  
 2334 Para distâncias entre 30 e 70 kb, o  $r^2$  médio foi de 0,23, 0,16 e 0,15, e para distâncias entre  
 2335 70 e 100 kb, o  $r^2$  médio foi de 0,19, 0,12 e 0,11. Brito et al. (2011) também relataram médias,  
 2336 desvios padrão e frequência de  $r^2$  maiores que 0,3 para duas populações simuladas. Nas  
 2337 distâncias de 0 a 100 kb, o  $r^2$  obtido em uma população de  $N_e = 272$  foi de 0,22 e o desvio  
 2338 padrão foi de 0,24. O percentual de  $r^2 > 0,3$  foi de 27,89%. Para distâncias de 100 a 200  
 2339 kb, o  $r^2$  médio e o desvio padrão foram 0,17 e 0,19, com percentual de  $r^2 > 0,3$  igual a  
 2340 19,57%. Estes valores, bem como os reportados para distâncias maiores, também estão  
 2341 de acordo com os resultados aqui obtidos.

2342 Para avaliar o nível de diferenciação genética entre as populações simuladas, foi  
 2343 realizada uma análise de componentes principais. O percentual de variância explicado por  
 2344 cada um dos dois principais componentes foi de 14,4% e 3,89%, respectivamente. As  
 2345 figuras do decaimento do desequilíbrio de ligação e da análise de componentes principais  
 2346 podem ser visualizadas no material publicado por Selli e colaboradores (2024).

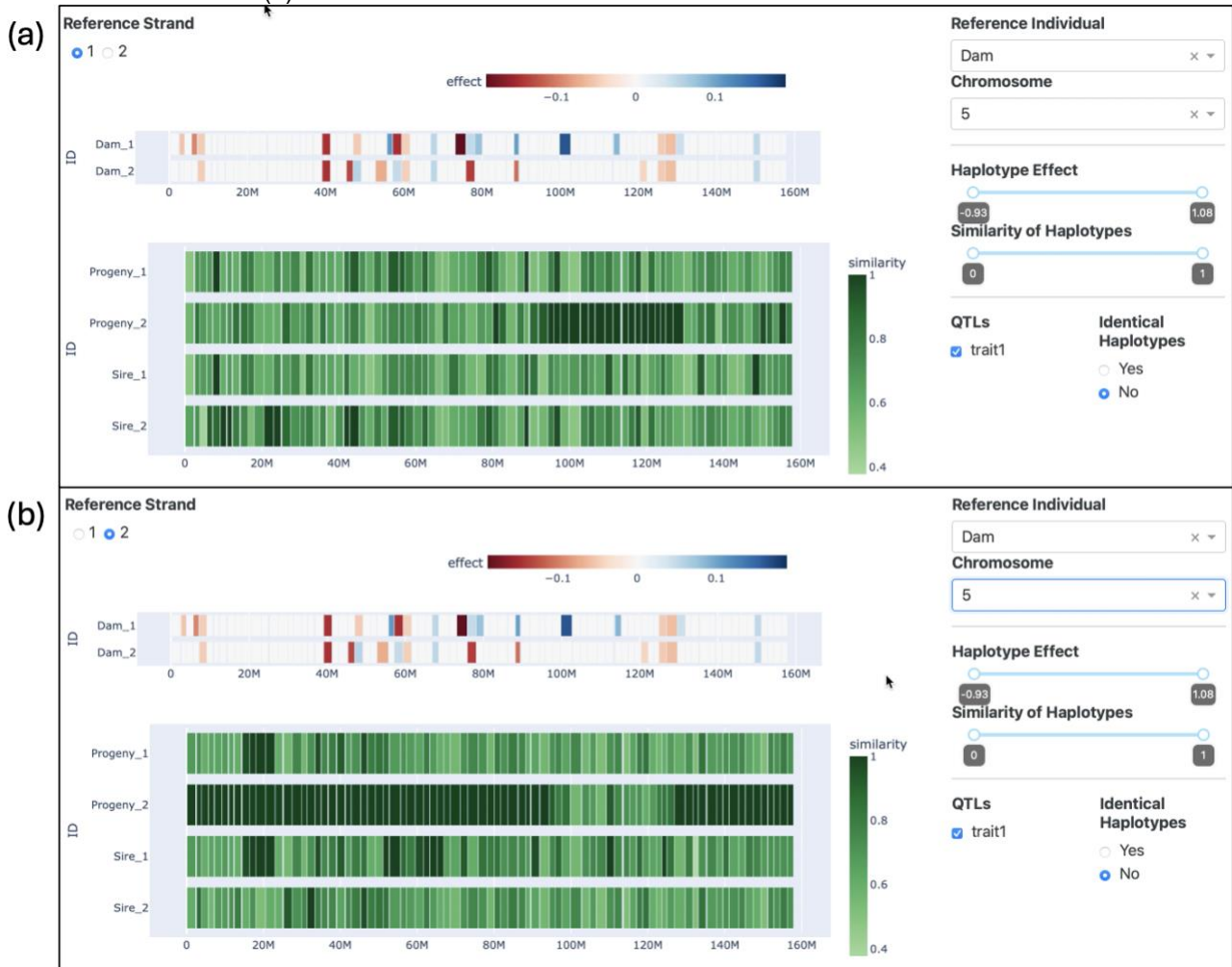
#### 2347 **4.4.2 Visualização de Trios**

2348 Neste cenário inicial, apresentamos as funcionalidades fundamentais da ferramenta  
 2349 de visualização. Neste contexto, comparamos um macho e uma fêmea da população D2,  
 2350 juntamente com seu descendente. A Figura 4 ilustra a interface tendo a vaca escolhida  
 2351 como indivíduo de referência e a opção “Haplótipos Idênticos” definida como “Não”. Os  
 2352 haplótipos dos demais animais são coloridos conforme o grau de similaridade em relação  
 2353 aos respectivos haplótipos da progênie.

2354

2355  
2356

Figura 4-4 – Visualização de trios – vaca como indivíduo de referência com a fita cromossômica paterna (a) ou materna (b) como referência.

2357  
2358

Fonte: Autoria própria.

2359

2360

2361

2362

2363

2364

Nas Figuras 4a e 4b, os haplótipos da progênie e do touro são comparados com a primeira e a segunda fitas da vaca, respectivamente. É possível visualizar trechos de maior similaridade entre a progênie e a vaca, evidenciando a transmissão de um pequeno trecho da primeira fita e trechos maiores da segunda fita, devido a um evento de crossover. Também existem trechos semelhantes entre a vaca e o touro, embora estes indivíduos não sejam aparentados.

2365

2366

2367

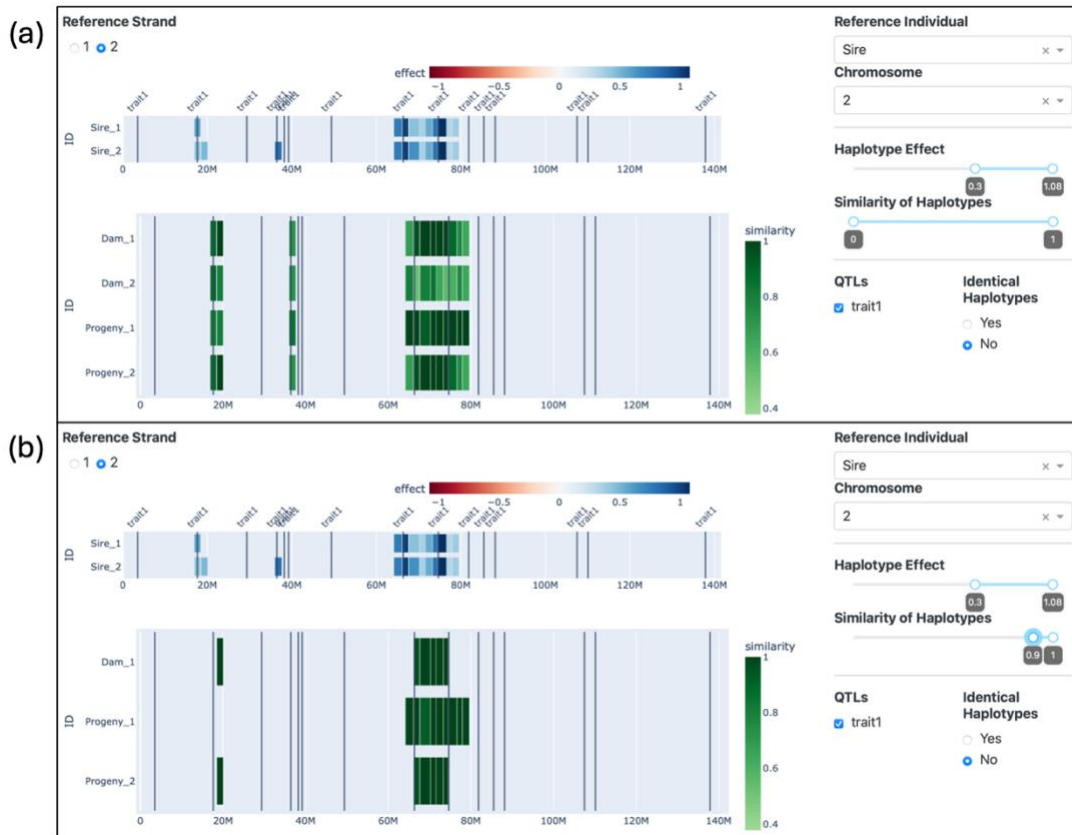
2368

Definir a vaca como indivíduo de referência permitiu visualizar que a segunda fita da progênie foi resultado de um evento de cruzamento entre as duas fitas da mãe. Além disso, alguns haplótipos apresentaram nível máximo de similaridade entre a mãe e o pai, revelando regiões idênticas entre indivíduos que não são imediatamente relacionados.

2369 Estes haplótipos idênticos representam regiões potenciais de menor diversidade genética  
2370 nessa população. Essa baixa diversidade pode ser atribuída a uma miríade de fatores,  
2371 nomeadamente desequilíbrio de ligação, seleção artificial, estrutura familiar e endogamia.

2372 A Figura 5 ilustra a comparação do Touro com a Vaca e a Progênie, em que a  
2373 segunda fita do segundo cromossomo do Touro é comparada com ambas as fitas do  
2374 segundo cromossomo dos demais indivíduos. O filtro de efeito foi ajustado em 0,3 para  
2375 selecionar apenas os haplótipos do pai com efeito para a característica simulada igual ou  
2376 superior a 0,3 e os haplótipos correspondentes dos demais indivíduos. Tais haplótipos do  
2377 Touro localizaram-se em torno das marcas de 20 Mb (efeito médio de 0,46) e 40 Mb (efeito  
2378 de 0,81) e entre 60 e 80 Mb (efeito médio de 0,65), embora os QTL estejam distribuídos ao  
2379 longo do cromossomo. Notavelmente, os efeitos mais pronunciados foram observados  
2380 entre 60 e 80 Mb, variando de 0,3 a 1,08. Na Figura 8a, a similaridade dos haplótipos da  
2381 mãe e da progênie em relação à segunda linhagem do touro variou de 52% a 100%.

2382 Figura 4-5 – Visualização de haplótipos com grandes efeitos positivos, onde a segunda fita do segundo  
2383 cromossomo do touro foi selecionada como referência.



2384

2385 Fonte: A autoria própria.

2386 As imagens mostram (a) a aplicação do filtro de efeito > 0,3 e (b) aplicação do filtro de efeito > 0,3 e do filtro  
2387 de similaridade > 0,9.

2388 Posteriormente, na Figura 5b, o filtro de similaridade foi definido para 0,9, levando  
2389 à exclusão de diversos haplótipos presentes na Dam\_1, Progeny\_1 e Progeny\_2.  
2390 Notavelmente, a segunda fita da Vaca não exibe quaisquer haplótipos com uma pontuação  
2391 de similaridade com os haplótipos selecionados do Touro igual ou superior a 0,9 e, portanto,  
2392 esta foi totalmente removida da visualização.

#### 2393 **4.4.3 Visualização dos avós**

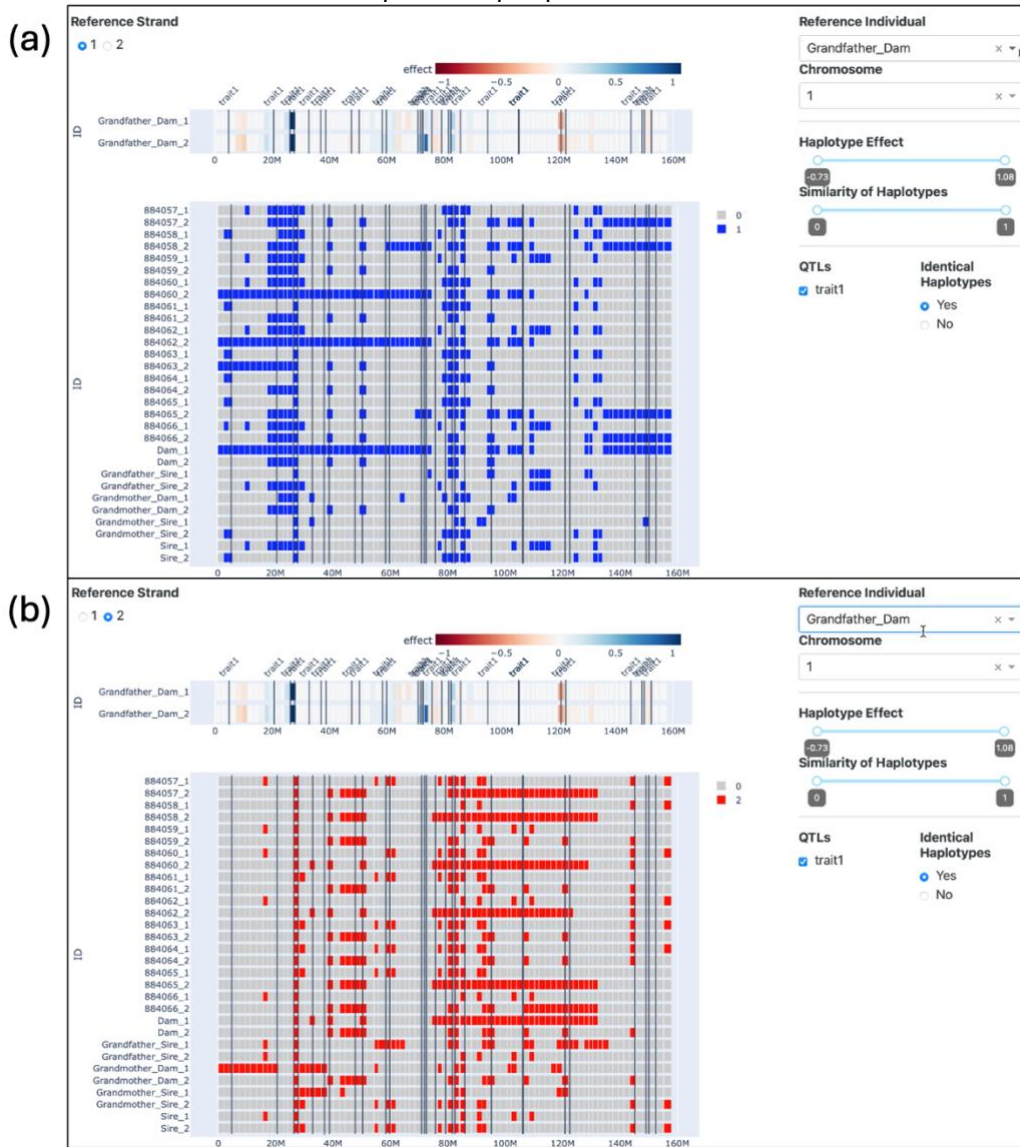
2394 Com o intuito de visualizar a herança genômica ao longo das gerações, nos  
2395 aprofundamos na construção de um cenário para o monitoramento de três gerações  
2396 consecutivas usando nossa ferramenta de visualização. Nosso objetivo foi rastrear a  
2397 transmissão de segmentos genômicos através dessas gerações. Para ilustrar esse  
2398 processo, quatro indivíduos foram escolhidos aleatoriamente da população D2 para  
2399 representar os avós – compreendendo tanto os avôs quanto as avós das linhagens do pai  
2400 e da mãe – bem como seus descendentes (o próprio pai e a mãe) e seus netos  
2401 (compreendendo 10 irmãos completos, cada um identificado por um ID numérico).

2402 A Figura 6 apresenta a nossa ferramenta com o avô materno (ID:  
2403 Grandfather\_Dam) como indivíduo de referência e a opção de haplótipos idênticos definida  
2404 como “sim”. É possível observar numerosos fragmentos genômicos idênticos à primeira  
2405 (Figura 6a) e segunda (Figura 6b) fitas do indivíduo de referência. Também se pode  
2406 visualizar blocos maiores que foram passados do avô materno para a vaca e para as  
2407 progênies: na fita 1 estes blocos estão localizados entre a posição 0 e 75 Mb  
2408 (aproximadamente), e da posição 130 Mb até o final do cromossomo (Figura 6a), enquanto  
2409 na fita 2, o trecho herdado pela vaca e por algumas das progênies é o complemento da fita  
2410 1, entre as posições 75 e 130 Mb, aproximadamente (Figura 6b).

2411

2412  
2413

Figura 4-6 – Visualização de três gerações. O avô materno com (a) a primeira fita ou (b) a segunda fita selecionada como referência. Apenas Haplótipos Idênticos são destacados.



2414

2415 Fonte: Autoria própria.

2416 Neste cenário, podemos visualizar como um segmento cromossômico é transmitido  
2417 de uma geração para a seguinte, e como dois irmãos completos podem apresentar  
2418 porcentagens de material genético muito diferentes de seus avós, o que por sua vez pode  
2419 afetar características econômicas ou de saúde relevantes. Roach et al. (ROACH et al.,  
2420 2010) analisaram as sequências do genoma completo (WGS) de quatro membros de uma  
2421 família (dois irmãos e seus pais) e foram capazes de identificar locais de recombinação  
2422 com grande precisão, identificar erros de sequenciamento, detectar SNPs muito raros e  
2423 reduzir significativamente o número de genes candidatos associados a duas doenças  
2424 genéticas. Os autores enfatizam a eficácia do emprego do WGS em um conjunto de dados

2425 que abrange quatro membros de uma família. No entanto, os autores alertam que a eficácia  
2426 de tais análises diminuiria substancialmente em estudos com menos marcadores ou  
2427 envolvendo um número menor de familiares. No presente estudo, nós simulamos um array  
2428 SNP de média densidade de ~ 50.000 marcadores (quase 2.000 marcadores por  
2429 cromossomo em bovinos). Isto provavelmente não serviria ao propósito de identificar SNPs  
2430 raros, como realizado em Roach et al. (2010); Contudo, ainda poderia ser informativo sobre  
2431 regiões mais amplas do genoma onde uma mutação importante poderia estar localizada.

2432           Conforme apresentado na Figura 6, a ferramenta desenvolvida no presente estudo  
2433 nos permitiu rastrear segmentos genômicos longos e curtos que são Idênticos por  
2434 Descendência (IBD), via comparação de fêmeas e suas progênes caracterizadas por  
2435 serem irmãos completos, ou Idênticos por Estado (IBS), situação na qual temos indivíduos  
2436 não relacionados ao avô materno. Os segmentos IBD oferecem informações claras sobre  
2437 quais descendentes herdaram segmentos genéticos específicos do avô, facilitando a busca  
2438 por regiões genômicas comuns entre descendentes que exibem uma característica ou  
2439 doença específica. Esta funcionalidade pode fornecer grande valor para a seleção  
2440 genômica em programas ou fazendas menores. Evidentemente, a eficácia desta estratégia  
2441 depende da natureza da herança e do número de loci que controlam a característica.

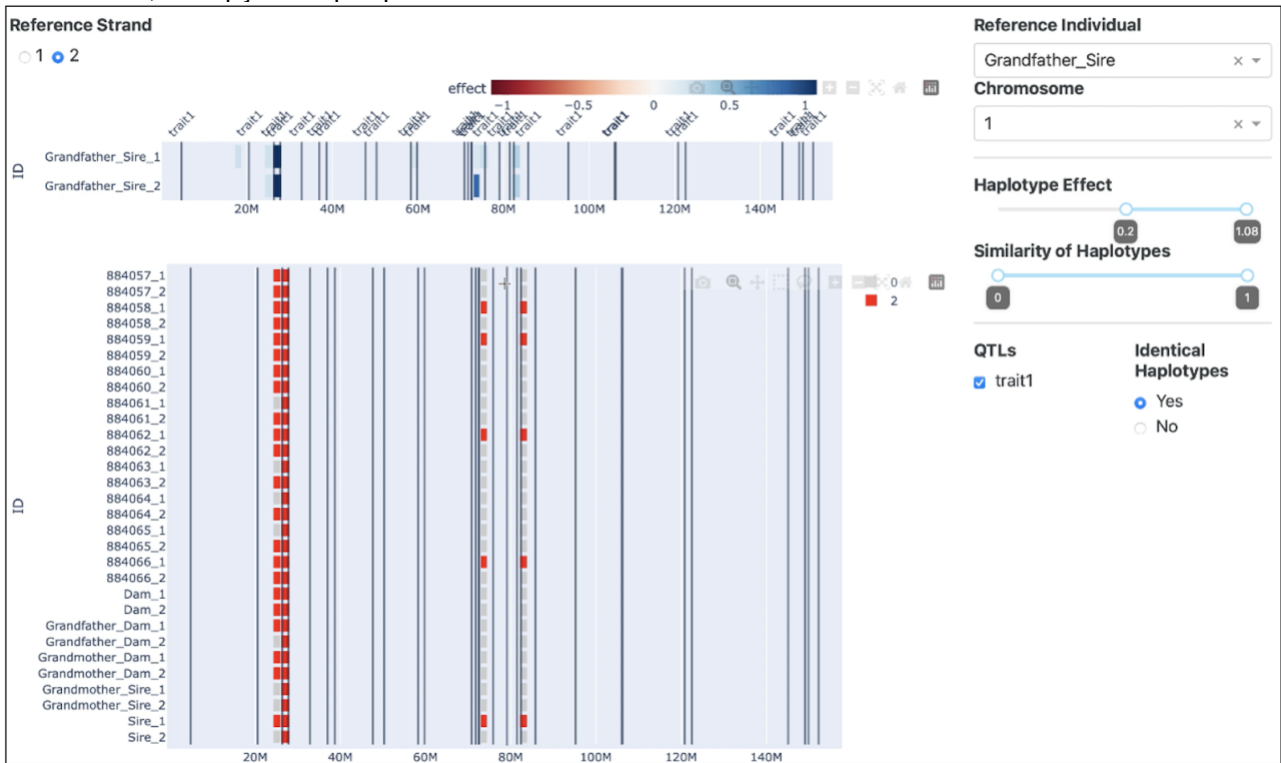
2442           Ao aplicar o filtro de efeitos é possível visualizar os haplótipos do Grandfather\_Sire  
2443 com maiores efeitos (Figura 7). Ao comparar tais haplótipos da segunda fita do indivíduo  
2444 de referência com o restante dos animais, é possível observar que alguns indivíduos  
2445 carregam exatamente os mesmos haplótipos. Além do haplótipo que é idêntico para todos  
2446 os indivíduos analisados, apenas um outro haplótipo (localizado na janela anterior) é  
2447 idêntico em indivíduos que não sejam os descendentes do Grandfather\_Sire (Figura 7).  
2448 Também podemos observar quais progênes herdaram os melhores haplótipos do avô  
2449 paterno. Quatro progênes herdaram haplótipos com efeitos positivos localizados em torno  
2450 de 80 Mb.

2451



2452  
2453

Figura 4-7 – Visualização de haplótipos da segunda fita do avô paterno com efeitos iguais ou maiores que 0,2. A opção "Haplótipos idênticos" é definida como "sim".



2454  
2455

Fonte: Autoria própria.

2456 Ao selecionar apenas regiões de efeitos acima de 0,2 (Figura 7), é possível rastrear  
2457 quais indivíduos herdaram estes haplótipos específicos. Isto permite-nos escolher  
2458 seletivamente descendentes que herdaram os segmentos genéticos desejados. Vale  
2459 ressaltar que nesta simulação foram simulados números variados de QTL em cada  
2460 cromossomo. O método de regressão SNP único falhou em capturar todos os efeitos de  
2461 QTL e atribuiu erroneamente alguns efeitos a regiões sem QTL reais, introduzindo certas  
2462 limitações à eficácia do filtro de efeitos. No entanto, conseguimos fazer comparações  
2463 visuais entre o número de efeitos e a presença de QTL em cada cromossomo.

2464 É importante ressaltar que o foco principal deste estudo não foi a estimativa dos  
2465 efeitos dos marcadores, servindo o GWAS como análise complementar à ferramenta. Em  
2466 vez disso, a nossa investigação teve como foco a identificação de segmentos genéticos  
2467 específicos e seus padrões de herança dentro da população simulada. Isso nos permitiu  
2468 explorar o potencial de seleção e filtragem direcionada com base na presença de haplótipos  
2469 desejados ou indesejados no contexto de análise de características complexas.

2470 Este artigo teve como objetivo mostrar os usos potenciais de visualizações  
2471 interativas, focando especificamente na questão: “Por que alguns irmãos completos são tão  
2472 diferentes?”. Em vez de fornecer uma ferramenta para todos os fins, incentivamos  
2473 acadêmicos, criadores e professores a adotarem esse conceito de visualização (seja  
2474 usando ferramentas de código aberto ou proprietárias) para desenvolver seus próprios  
2475 pipelines e painéis, a fim de responder às suas perguntas específicas. Nos esquemas de  
2476 criação de gado leiteiro, por exemplo, existe a preocupação com genes importantes, como  
2477 a beta-caseína A1/A2 (OLENSKI et al., 2010), doenças recessivas, nomeadamente  
2478 Malformação Vertebral Complexa (MALHER; BEAUDEAU; PHILIPOT, 2006) e Brachyspina  
2479 (CHARLIER et al., 2012), e diminuição da fertilidade causada por alguns haplótipos em  
2480 particular (VANRADEN et al., 2011). Uma vez que os genes subjacentes a estas condições  
2481 foram identificados e podem ser mapeados por marcadores, esta informação poderia ser  
2482 facilmente incluída no painel e assim proporcionar uma melhor compreensão do processo  
2483 de herança de tais genes e como eles se relacionam com regiões de interesse para os  
2484 produtores. As representações visuais oferecem uma forma intuitiva de observar resultados  
2485 obtidos por análises estatísticas e genômicas, facilitando a comunicação entre  
2486 pesquisadores, educadores, criadores e agricultores, conforme ilustrado por Deeb et al.  
2487 (2024).

2488 Além disso, enfatizamos as vantagens do emprego de ferramentas de código aberto  
2489 no desenvolvimento de tipos específicos de gráficos. Em outro estudo de Selli et al. (2021),  
2490 o uso de ferramenta proprietária permitiu a visualização rápida e eficiente das análises  
2491 genômicas; entretanto, a dificuldade de replicação dos gráficos em novos conjuntos de  
2492 dados foi um dos obstáculos que levaram à adoção de ferramentas mais flexíveis no  
2493 presente estudo. A facilidade de incorporar cálculos back-end complexos para preparar os  
2494 dados também foi uma grande vantagem observada no uso de software de código aberto,  
2495 além da flexibilidade para criar visualizações personalizadas para responder a questões  
2496 específicas colocadas pelos pesquisadores.

#### 2497 4.5 CONCLUSÃO

2498 Neste estudo, utilizamos ferramentas de código aberto para criar uma estratégia  
2499 interativa e eficaz para visualização de dados de SNPs compartilhados entre indivíduos,

2500 além de rastrear a transmissão de haplótipos em uma população de bovinos de corte. Os  
 2501 nossos objetivos foram demonstrar as vantagens das visualizações interativas e fornecer  
 2502 informações sobre a herança genética em populações de animais de produção. Usamos  
 2503 uma ferramenta de visualização para explorar dois cenários. No primeiro cenário,  
 2504 visualizamos padrões de herança genética dentro de trios, destacando a transmissão de  
 2505 segmentos genômicos entre pais e filhos. Isso nos permitiu identificar regiões de haplótipos  
 2506 idênticos, esclarecendo fatores como endogamia e seleção artificial. No segundo cenário,  
 2507 rastreamos a transmissão de segmentos genômicos ao longo de três gerações. Assim,  
 2508 pudemos obter informações sobre quais descendentes herdaram segmentos genéticos  
 2509 específicos de seus avós, oferecendo aplicações potenciais para seleção genômica e  
 2510 programas de melhoramento. No geral, nosso estudo ressalta o poder do emprego de  
 2511 ferramentas de código aberto, em oposição a programas proprietários, a fim de esclarecer  
 2512 questões específicas, mostrando aplicações para a genética pecuária e a comunidade de  
 2513 criação, bem como para educadores e pesquisadores que buscam comunicar conceitos  
 2514 complexos e explorar estratégias de seleção.

#### 2515 4.6 REFERÊNCIAS

- 2516 ABRAHAM, G.; QIU, Y.; INOUE, M. FlashPCA2: principal component analysis of Biobank-  
 2517 scale genotype datasets. **Bioinformatics**, v. 33, n. 17, p. 2776–2778, 1 set. 2017.  
 2518
- 2519 ALI, S. M.; GUPTA, N.; NAYAK, G. K.; LENKA, R. K. Big data visualization: Tools and  
 2520 challenges. Em: Proceedings of the 2016 2nd International Conference on Contemporary  
 2521 Computing and Informatics, IC3I 2016, 2016, [...]. Institute of Electrical and Electronics  
 2522 Engineers Inc., 2016. p. 656–660.  
 2523
- 2524 BATEMAN, S.; LEPROU, J.; JULIAN, A.; WEIR, A. Atypical cases of Zinc poisoning in  
 2525 Taranaki dairy herds during the spring of 2012. n. January 2013, p. 1–19, 2018.  
 2526
- 2527 BOHMANOVA, J.; SARGOLZAEI, M.; SCHENKEL, F. S. Characteristics of linkage  
 2528 disequilibrium in North American Holsteins. **BMC Genomics**, v. 11, n. 1, 2010.  
 2529
- 2530 BRITO, F. V.; NETO, J. B.; SARGOLZAEI, M.; COBUCI, J. A.; SCHENKEL, F. S.  
 2531 Accuracy of genomic selection in simulated populations mimicking the extent of linkage  
 2532 disequilibrium in beef cattle. **BMC Genetics**, v. 12, 2011.  
 2533
- 2534 BUELS, R.; YAO, E.; DIESH, C. M.; HAYES, R. D.; MUNOZ-TORRES, M.; HELT, G.;  
 2535 GOODSTEIN, D. M.; ELSIK, C. G.; LEWIS, S. E.; STEIN, L.; HOLMES, I. H. JBrowse: A  
 2536 dynamic web platform for genome visualization and analysis. **Genome Biology**, v. 17, n.  
 2537 1, p. 1–12, 2016.

- 2538  
2539 CHARLIER, C.; AGERHOLM, J. S.; COPPIETERS, W.; KARLSKOV-MORTENSEN, P.; LI,  
2540 W.; DE JONG, G.; FASQUELLE, C.; KARIM, L.; CIRERA, S.; CAMBISANO, N.; AHARIZ,  
2541 N.; MULLAART, E.; GEORGES, M.; FREDHOLM, M. A Deletion in the Bovine FANCI  
2542 Gene Compromises Fertility by Causing Fetal Death and Brachyspina. **PLoS ONE**, v. 7, n.  
2543 8, p. e43085, 29 ago. 2012.  
2544  
2545 COOK, M. P. Visual Representations in Science Education: The Influence of Prior  
2546 Knowledge and Cognitive Load Theory on Instructional Design Principles. **Science**  
2547 **Education**, v. 90, n. 6, p. 1073–1091, 2006.  
2548  
2549 CURTI, P. D. F.; SELLI, A.; PINTO, D. L.; MERLOS-RUIZ, A.; BALIEIRO, J. C. D. C.;  
2550 VENTURA, R. V. Applications of Livestock Monitoring Devices and Machine Learning  
2551 Algorithms in Animal Production and Reproduction: An Overview. **Animal Reproduction**,  
2552 v. 20, n. 2, p. e20230077, 2023.  
2553  
2554 DOSHER, B.; LU, Z.-L. Visual Perceptual Learning and Models. **Annual Review of Vision**  
2555 **Science**, v. 3, n. 1, p. 343–363, 15 set. 2017.  
2556  
2557 GAYNOR, R. C.; GORJANC, G.; HICKEY, J. M. AlphaSimR: an R package for breeding  
2558 program simulations. **G3 Genes|Genomes|Genetics**, v. 11, n. 2, 12 abr. 2021. Disponível  
2559 em: <<https://academic.oup.com/g3journal/article/doi/10.1093/g3journal/jkaa017/6025179>>.  
2560 Acesso em: 31 mar. 2022.  
2561  
2562 GONDRO, C. Primer to Analysis of Genomic Data Using R.—Use R! **Springer**  
2563 **International Publishing**: Cham, Switzerland, 2015; ISBN 978-3-319-14474-0.  
2564  
2565 HULLMAN, J.; DIAKOPOULOS, N. Visualization Rhetoric: Framing Effects in Narrative  
2566 Visualization. **IEEE Transactions on Visualization and Computer Graphics**, v. 17, n.  
2567 12, p. 2231–2240, dez. 2011.  
2568  
2569 LARKIN, D. M.; DAETWYLER, H. D.; HERNANDEZ, A. G.; WRIGHT, C. L.; HETRICK, L.  
2570 A.; BOUCEK, L.; BACHMAN, S. L.; BAND, M. R.; AKRAIKO, T. V.; COHEN-ZINDER, M.;  
2571 THIMMAPURAM, J.; MACLEOD, I. M.; HARKINS, T. T.; MCCAGUE, J. E.; GODDARD, M.  
2572 E.; HAYES, B. J.; LEWIN, H. A. Whole-genome resequencing of two elite sires for the  
2573 detection of haplotypes under selection in dairy cattle. **Proceedings of the National**  
2574 **Academy of Sciences**, v. 109, n. 20, p. 7693–7698, 15 maio 2012.  
2575  
2576 LU, D.; SARGOLZAEI, M.; KELLY, M.; LI, C.; VOORT, G. V.; WANG, Z.; PLASTOW, G.;  
2577 MOORE, S.; MILLER, S. P. Linkage disequilibrium in Angus, Charolais, and Crossbred  
2578 beef cattle. **Frontiers in Genetics**, v. 3, n. AUG, p. 1–10, 2012.  
2579  
2580 L'YI, S.; GEHLENBORG, N. Multi-View Design Patterns and Responsive Visualization for  
2581 Genomics Data. **IEEE Transactions on Visualization and Computer Graphics**, v. 29, n.  
2582 1, p. 559–569, jan. 2023.  
2583

- 2584 MALHER, X.; BEAUDEAU, F.; PHILIPOT, J. M. Effects of sire and dam genotype for  
2585 complex vertebral malformation (CVM) on risk of return-to-service in Holstein dairy cows  
2586 and heifers. **Theriogenology**, v. 65, n. 6, p. 1215–1225, 1 abr. 2006.  
2587
- 2588 MCKINNEY, W. *Data Structures for Statistical Computing in Python*. Em: 2010, Austin,  
2589 Texas. [...]. Austin, Texas: 2010. p. 56–61.  
2590
- 2591 MOROTA, G.; CHENG, H.; COOK, D.; TANAKA, E. ASAS-NANP SYMPOSIUM:  
2592 Prospects for Interactive and Dynamic Graphics in the Era of Data-Rich Animal Science1.  
2593 **Journal of Animal Science**, v. 99, n. 2, p. skaa402, 1 fev. 2021.  
2594
- 2595 OLDENBROEK, K.; VAN DER WAAIJ, L. **Textbook Animal Breeding: Animal Breeding  
2596 and Genetics for BSc Students**. [s.l.] Centre for Genetic Resources and Animal Breeding  
2597 and Genomics Group, Wageningen University and Research Centre, 2014.  
2598
- 2599 OLENSKI, K.; KAMIŃSKI, S.; SZYDA, J.; CIESLINSKA, A. Polymorphism of the Beta-  
2600 Casein Gene and Its Associations with Breeding Value for Production Traits of Holstein–  
2601 Friesian Bulls. **Livestock Science**, v. 131, n. 1, p. 137–140, jun. 2010.  
2602
- 2603 R CORE TEAM. **R: A Language and Environment for Statistical Computing** Vienna,  
2604 Austria. R Foundation for Statistical Computing, 2020. Disponível em: <[https://www.r-  
2605 project.org/](https://www.r-project.org/)>.  
2606
- 2607 ROACH, J. C.; GLUSMAN, G.; SMIT, A. F. A.; HUFF, C. D.; HUBLEY, R.; SHANNON, P.  
2608 T.; ROWEN, L.; PANT, K. P.; GOODMAN, N.; BAMSHAD, M.; SHENDURE, J.;  
2609 DRMANAC, R.; JORDE, L. B.; HOOD, L.; GALAS, D. J. Analysis of Genetic Inheritance in  
2610 a Family Quartet by Whole-Genome Sequencing. **Science**, v. 328, n. 5978, p. 636–639,  
2611 30 abr. 2010.  
2612
- 2613 ROUGIER, N. P.; DROETTBOOM, M.; BOURNE, P. E. Ten Simple Rules for Better  
2614 Figures. **PLoS Computational Biology**, v. 10, n. 9, p. e1003833, 11 set. 2014.  
2615
- 2616 SARGOLZAEI, M.; SCHENKEL, F. S. QMSim: A large-scale genome simulator for  
2617 livestock. **Bioinformatics**, v. 25, n. 5, p. 680–681, 2009.  
2618
- 2619 SELLI, A.; MILLER, S. P.; VENTURA, R. V. The Use of Interactive Visualizations for  
2620 Tracking Haplotypic Inheritance in Livestock. **Ruminants**, v. 4, n. 1, p. 90–111, 21 fev.  
2621 2024.  
2622
- 2623 SELLI, A.; VENTURA, R. V.; FONSECA, P. A. S.; BUZANSKAS, M. E.; ANDRIETTA, L.  
2624 T.; BALIEIRO, J. C. C.; BRITO, L. F. Detection and Visualization of Heterozygosity-Rich  
2625 Regions and Runs of Homozygosity in Worldwide Sheep Populations. **Animals**, v. 11, n.  
2626 9, p. 2696, 15 set. 2021.  
2627
- 2628 VAN ESSEN, D. C.; ANDERSON, C. H.; FELLEMAN, D. J. Information Processing in the  
2629 Primate Visual System: An Integrated Systems Perspective. **Science**, v. 255, n. 5043, p.  
2630 419–423, 24 jan. 1992.  
2631

- 2632 VAN ROSSUM, G.; DRAKE JR, F. L. **Python reference manual**. [s.l.] Centrum voor  
2633 Wiskunde en Informatica Amsterdam, 1995.
- 2634
- 2635 VANRADEN, P. M.; NULL, D. J.; OLSON, K. M.; HUTCHISON, J. L. Reporting of  
2636 Haplotypes with Recessive Effects on Fertility. n. 44, 2011.
- 2637
- 2638 ZACKS, J.; TVERSKY, B. Bars and Lines: A Study of Graphic Communication. **Memory &**  
2639 **Cognition**, v. 27, n. 6, p. 1073–1079, nov. 1999.
- 2640
- 2641 ZHOU, L.; HANSEN, C. D. A Survey of Colormaps in Visualization. **IEEE Transactions on**  
2642 **Visualization and Computer Graphics**, v. 22, n. 8, p. 2051–2069, 1 ago. 2016.
- 2643

## 2644 5 CONSIDERAÇÕES FINAIS

2645 Este trabalho reuniu diferentes aspectos do melhoramento genético, análises  
2646 genômicas, simulação e visualização de dados (fazendo o uso de imagens tanto estáticas  
2647 quanto dinâmicas). Apesar de não ter como objetivo a proposta de uma nova metodologia,  
2648 este estudo apresenta uma base para a construção de pipelines complexos que envolvem  
2649 o uso de diferentes ferramentas computacionais. Foram verificadas vantagens na inclusão  
2650 dos coeficientes de endogamia estimados através das corridas de homozigosidade, em um  
2651 processo de acasalamentos dirigidos, em relação a metodologias mais tradicionalmente  
2652 utilizadas, como os coeficientes de endogamia de pedigree e da matriz genômica. Também  
2653 foi demonstrada, por meio de dois estudos distintos, a importância do uso de visualizações  
2654 para a comunicação dos resultados e a compreensão de fenômenos biológicos. Algumas  
2655 limitações foram pontuadas, tais como a necessidade de ajustes em uma das simulações  
2656 e a adequação da metodologia para estimação de valores genéticos, e foram propostas  
2657 alternativas para superar ambas as limitações. Assim, espera-se que o presente trabalho  
2658 contribua para o desenvolvimento de pesquisas futuras na área do melhoramento animal,  
2659 simulação de dados e compartilhamento eficiente de resultados científicos.