



**UNIVERSIDADE FEDERAL DA BAHIA - UFBA
PROGRAMA DE DOUTORADO EM ZOOTECNIA**

**CARACTERIZAÇÃO DE CORRIDAS DE HOMOZIGOSE,
HETEROZIGOSE E ANÁLISE DA DIVERSIDADE GENÔMICA
EM POPULAÇÕES DE BOVINOS**

HENRIQUE ALBERTO MULIM

**SALVADOR – BA
FEVEREIRO-2023**

HENRIQUE ALBERTO MULIM

**CARACTERIZAÇÃO DE CORRIDAS DE HOMOZIGOSE,
HETEROZIGOSE E ANÁLISE DA DIVERSIDADE GENÔMICA
EM POPULAÇÕES DE BOVINOS**

Tese apresentada ao Programa de
Doutorado em Zootecnia, da Universidade
Federal da Bahia como requisito parcial para
obtenção do título de Doutor em Zootecnia.

Área de Concentração: Produção Animal

Orientador: Dr. Victor Breno Pedrosa
Coorientador: Dr. Luiz Fernando Brito

**SALVADOR - BA
FEVEREIRO - 2023**

Dados internacionais de catalogação-na-publicação
(SIBI/UFBA/Biblioteca Universitária Reitor Macedo Costa)

Mulim, Henrique Alberto.

Caracterização de corridas de homozigose, heterozigose e análise da diversidade genômica em populações de bovinos / Henrique Alberto Mulim. - 2023.
156 f.: il.

Orientador: Prof. Dr. Victor Breno Pedrosa.

Coorientador: Prof. Dr. Luiz Fernando Brito.

Tese (doutorado) - Universidade Federal da Bahia, Escola de Medicina Veterinária e Zootecnia, Salvador, 2023.

1. Zootecnia. 2. Genética de populações. 3. Bovinos - Genética. 4. Bovinos - Mapeamento genômico. 5. Genômica. I. Pedrosa, Victor Breno. II. Universidade Federal da Bahia. Escola de Medicina Veterinária e Zootecnia. III. Título.

CDD - 636.20821
CDU - 636.2

HENRIQUE ALBERTO MULIM

CARACTERIZAÇÃO DE CORRIDAS DE HOMOZIGOSE, HETEROZIGOSE E ANÁLISE DA DIVERSIDADE GENÔMICA EM POPULAÇÕES DE BOVINOS

Tese defendida e aprovada pela Comissão Examinadora em 06 de fevereiro de 2023.

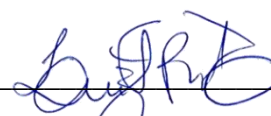
Comissão examinadora:



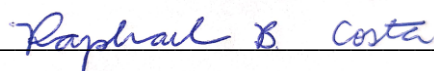
Dr. Victor Breno Pedrosa
UFBA
Orientador / Presidente



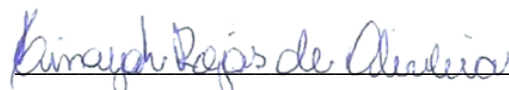
Dra Luís Fernando Batista Pinto
UFBA



Dr. Luiz Fernando Brito
Purdue University



Dr. Raphael Bermal Costa
UFBA



Dra. Hinayah Rojas de Oliveira
Purdue University

SALVADOR-BA
FEVEREIRO – 2023

“O poder da ciência não está em suas descobertas, mas em sua estrutura: trata-se da única atividade humana projetada para reconhecer, revisar e aprender com os próprios erros”

Pasternak & Orsi

AGRADECIMENTOS

Bem, chegou ao fim mais uma etapa. E não foi fácil não ... quem por vezes me viu sorrir, não imagina as vezes em que tive que me fazer forte. Não é fácil fazer uma pós-graduação, é um sonho, uma oportunidade, mesmo assim vem com uma demanda muita alta, principalmente a auto demanda. “Você tem que ser forte, você tem que mostrar que você pode. Fraquejar? Jamais!”

Bom, era isso que eu me falava no início desse sonho ... e só demonstra o quanto eu ainda era imaturo. E quando falo sobre maturidade, falo sobre a maturidade emocional mesmo. A gente cresce em um mundo em que nos cria para sermos fortes, para sermos “frios” e não demonstrarmos o que sentimos. Certamente, isso pode nos levar a alcançar voos, mas mais certo ainda irá fazer com que percamos uma experiência ainda maior: a de viver! E como eu vivi nesses últimos 4 anos. Encerro esse ciclo com uma paixão pela minha história, pelo que vivi e estou vivendo e amando as pessoas que encontrei por todo esse caminho. Sendo grato a tudo e todos...

Essa não foi uma jornada que trilhei sozinho, e quero muito agradecer a todos que contribuíram para que eu chegasse até aqui.

Primeiramente, agradeço muito à Deus ... o tanto que já Lhe pedi orientação não é brincadeira e certamente meu anjo da guarda anda com uma bússola (brinks) ... Sempre soube que o Senhor esteve perto e diversas vezes senti Sua presença, ela foi primordial para me dar paz nos momentos de aflição e achego quando me sentia só. Que meus passos sejam guiados conforme a Sua vontade e que o Senhor me use para que, assim como usou outras pessoas para me mostrar o Seu amor, eu possa ser usado.

Agradeço também a minha família que ainda que por muitas vezes não entendia o que eu estava fazendo me apoiou e comemorou todas as minhas conquistas. Em especial, quero destacar aqui a minha mãe, Tereza Cristina Gomides, a pessoa com quem compartilhei a maioria das situações, frustrações, medos e principalmente alegrias. Mãe, eu te amo! E obrigado ... obrigado por acreditar em mim e nos meus sonhos, e por me mostrar os preceitos mais importantes da vida: generosidade, perseverança, amor e humildade. Meu maior orgulho é ser seu filho!

Todo esse sonho, só foi possível porque junto comigo estava um cara fera e que me mostrou como é ser um pesquisador. Eu tive a sorte de ter um orientador que fez muito mais do que orientar, ele me inspirou. Dr. Victor Breno Pedrosa, obrigado por todo os

seus ensinamentos, seja na nossa área ou mesmo na vida. Obrigado pelos puxões de orelha, incentivos, conversas e principalmente todas as oportunidades que você me proporcionou. Você mudou minha vida e serei eternamente grato por tudo o quanto o senhor fez. Espero poder inspirar vidas e ajudar a realizar sonhos assim como o senhor um dia me inspirou e me ajudou.

Como já dizia C.S. Lewis (segundo o Google), “A amizade nasce no momento em que uma pessoa diz para a outra: — “O quê? Você também! Pensei que eu era o único.” E foi assim que eu encontrei vocês, de início era somente um desabafo de convivência, com o passar do tempo começamos a ver que tinha muito mais em comum. Começamos a compartilhar sonhos, aliás temos o mesmo (sermos MILIS), nossas angústias, reclamações e visões de vida. Trabalho em grupo? Perfeito! Pode ser nós 4??? É minhas meninas, minha experiência na Bahia não teria sido tão maravilhosa se não tivesse vocês ali comigo. Serei eternamente grato à Deus por ter colocado vocês na minha vida e como, de um grupo de trabalho, uma Ohana se formou. Marina Ribeiro, Taiana Cortez e Tatiana Cortez, vocês sempre serão as meninas do meu coração!

Durante meu tempo em Salvador, conheci pessoas maravilhosas com quem aprendi muito e me diverti ainda mais. Aqui deixo o espaço de muito obrigado a duas pessoas em especial: a Cláudia Loianny, que com seu jeito ímpar me conquistou desde o primeiro dia com seu belo sorriso. Mulher forte e guerreira, exemplo de dedicação no trabalho e amor a vida. E a Fernando Correia Cairo, um profissional incrível e com uma dedicação exemplar! Obrigado por todas as conversas, risadas e descontrações dentro e fora da Universidade, vocês fizeram meu tempo em Salvador ser ainda mais leve.

Agradeço também a toda Universidade Federal da Bahia, em especial ao Programa de Pós-Graduação em Zootecnia, toda a equipe técnica, professores e alunos com que tive o prazer de conviver e trabalhar. Em especial, 5 profissionais que compartilharam seus conhecimentos na área do melhoramento animal comigo. Dra. Iara Del Pilar Solar Diaz, Dra. Valdecy Aparecida da Rocha Cruz, Dr. Raphael Bernal da Costa, Dr. Gregório Miguel Ferreira de Camargo e Dr. Luís Fernando Batista Pinto, obrigado por toda a dedicação dentro de sala de aula. Agradeço também a FAPESB (Fundação de Amparo à Pesquisa do Estado da Bahia), pela concessão da bolsa de estudos durante todo o período de doutoramento.

Nesse último ano, muitas coisas aconteceram e uma mudança radical fez com que

novas pessoas entrassem em minha vida e conquistassem um espaço grande nela.

Primeiramente, à Gustavo Rodrigues e João Paulo Grizenti, dois seres iluminados com quem tenho o prazer de ter uma amizade tão linda e verdadeira. Obrigado por todas as conversas, momentos de reflexão e apoio. A amizade de vocês me trouxe vigor em momentos em que por vezes eu me senti cabisbaixo ou desiludido.

A Hui Wen, Lirong Hu, Felipe Eguti e Juliana Ferreira ... juntos formamos a “Connection China – Brazil”. Obrigado pessoal por todos nossos momentos de viagem, festas e comemorações, mas principalmente pelo momento de apoio que tive de cada um de vocês. Em especial, a Juliana ou Juuh, que conquistou meu ser e em tão pouco tempo construímos uma irmandade linda e forte. Obrigado por me fazer perceber coisas sobre mim que nem eu mesmo sabia, me entender sem me julgar, me alegrar e me divertir, mas principalmente por estar sempre ao meu lado. A sim também a Felipe Eguti, que não é mais somente meu colega de trabalho, mas sim meu irmão. Obrigado por proporcionar momentos incríveis de descontrações e risadas, mas também me ouvir e me fazer entender que a vida também pode ser leve.

A Dr. Katiene Regia Silva Sousa, Lorena Benfica, Leticia Fernanda de Oliveira, Artur Rocha, Dr. Leonardo Glória, Dr. André Araújo, Jacob Maskal, Sharlene Hartman, Rick Hernandez, Jason Graham, Dr. Pedro Freitas, Dra. Amanda Alvarenga, Caitlin Watts e Ali Saleem, agradeço de coração todas as risadas conversas e papos durante esse ano, mas principalmente pela força durante a execução a todo o momento desse trabalho. Obrigado pelas ideias, ajudas e motivações. Em especial agradeço a Katiene, Lorena e Leticia por nos momentos em que estive desanimado, pude ter consolo e palavras de incentivo.

A Universidade de Purdue pela oportunidade de desenvolver parte do meu doutorado como *Visiting Scholar*. Em especial ao Dr. Luiz Fernando Brito, por toda coorientação, colaboração e dedicação no desenvolvimento desse trabalho. Como já lhe disse pessoalmente, tenho uma grande admiração pelo profissional que és. E essa admiração só cresceu com a oportunidade de trabalhar contigo. Obrigado por me propor desafios e me fazer, por vezes, sair da zona de conforto.

Longe de casa a saudade bateu, mas quando encontrei vocês, pude ter um gostinho de estar no meio do meu povo novamente. Mayara Rocha, Fernando Aguilera, Cíntia Sant’ana, Vitor Haetinger, Mirian Campos, Amanda Miranda e Vinicius Lima, obrigado

por me fazer sentir em casa, bem e acolhido. Nossos momentos de comemoração, aflição, risadas e histórias certamente marcaram minha trajetória aqui em Purdue.

Bem, durante todo esse tempo de doutoramento, muitas pessoas passaram pela minha vida e estiveram comigo. E isso só foi possível porque me permitiu viver enquanto sonhava. O certo é que as demandas ainda continuaram, mas entendi que viver o presente é tão importante quanto batalhar pelo futuro. Chorei, sorri, fiquei bravo e feliz ... vivi e demonstrei tudo aquilo que senti! E isso só fez com que eu me apaixonasse por todo o processo e por mim.

Aos que chegaram na leitura até aqui a grande mensagem que aprendi durante esse tempo foi: Seja feliz! e “que a felicidade não dependa do tempo, nem da paisagem, nem da sorte, nem do dinheiro. Que ela possa vir com toda simplicidade, de dentro para fora, de cada um para todos. Que as pessoas saibam falar, calar, e acima de tudo ouvir. Que tenham amor ou então sintam falta de não o ter. Que tenham ideais e medo de perdê-lo. Que amem ao próximo e respeitem sua dor. Para que tenhamos certeza de que, como já dizia Drummond: Ser feliz sem motivo é a mais autêntica forma de felicidade!” (Autor Desconhecido).

Obrigado!

LISTA DE TABELAS

CAPÍTULO 1

Tabela 1 - Rebanhos utilizados no estudo com as respectivas amostragens (N), sigla referente ao rebanho, densidade do chip de genotipagem e o banco de dados ao qual se encontram os genótipos.....	35
Tabela 2 - Estatísticas descritivas dos conjuntos de dados genômicos após o controle de qualidade do genótipo	36
Tabela 3 - Média dos coeficientes de endogamia estimados para as cinco metodologias de cálculo de endogamia.....	43
Tabela 4 - Ilhas de homozigose encontradas em diferentes cromossomos e grupos de indivíduos	45
Tabela 5 - Regiões ricas em heterozigose (HER), nas diferentes populações, que aparecem em pelo menos 10% dos indivíduos da população	47
Tabela 6 - Parâmetros para identificação de corridas de homozigose em diferentes estudos com diferentes densidades de painéis de genotipagem	51

CAPÍTULO 2

Tabela 1 - Estimativas do coeficiente de endogamia com diferentes metodologias para animais da raça Purunã.....	89
Tabela 2 - Consistência da fase gamética baseada na correlação Person, entre a raça Purunã e suas raças fundadoras: Angus, Canchim, Charolês e Nelore	91
Tabela 3 - Proporção de SNPs polimórficos classificados pela frequência do menor alelo (MAF)	92
Tabela 4 - Termos significativos ($p < 0,05$) da ontologia gênica (GO) para os genes localizados em corridas de regiões de homozigose na raça Purunã.....	94
Tabela 5 - Regiões enriquecidas em heterozigotos (HER) que aparecem em pelo menos 10% dos indivíduos Purunã	97
Tabela 6 - Termos significativos ($p < 0,05$) da Gene Ontology (GO) para regiões enriquecidas com heterozigotos encontrados na raça Purunã	98

CAPÍTULO 3

Tabela 1 - Número de regiões ricas em heterozigose detectadas e classificação	
--	--

com base no comprimento para os cenários com o número mínimo de marcadores iguais a 5 e 10.....	130
Tabela 2 - Ilhas de regiões rica em heterozigose, média do desequilíbrio de ligação (R2) dos marcadores dentro das ilhas e porcentagem em que que essas aparecem em cada cenário.....	135
Tabela 3 - Anotação dos genes para as Ilhas de regiões ricas em heterozigose	138
Tabela 4 - Termos significativos ($p < 0,05$) da Ontologia Gênica (GO) e vias para os genes localizados em ilhas de regiões ricas com heterozigose.....	140
Tabela 5 - QTLs (quantitative trait loci) significativo ($p < 0,05$) para os genes localizados em ilhas de regiões ricas em heterozigose.	142

LISTA DE FIGURAS

CAPÍTULO 1

Figura 1 - Classificação das corridas de homozigose (ROH) de acordo com o comprimento e as raças bovinas.....	41
Figura 2 - Correlação entre os métodos de estimativa de endogamia.....	42
Figura 3 - Comportamento do tamanho efetivo da população (N_e) ao longo das últimas gerações	44
Figura 4 - Número de regiões ricas em heterozigose (HER) por raça estudada.	46
Figura 5 - Número de regiões ricas em heterozigose (HER) e classificação pelo comprimento, com base em diferentes valores de parâmetros nas raças Montana (MON) e Nelore (NEL50).....	48
Figura 6 - Comparação entre corridas de painel SNP de homozigose e análises WGS para o cromossomo 1.....	49
Figura 7 - Comparação entre corridas de painel SNP de homozigose e análises WGS para o cromossomo 20.....	50

CAPÍTULO 2

Figura 1 - Animais Purunã do Instituto Agronômico do Paraná (IAPAR, Ponta Grossa, Paraná, Brasil).....	79
Figura 2 - Estratificação Populacional da raça Purunã. A Análise de Componentes Principais (PCA) incluindo animais das raças Purunã, Angus, Canchim, Charolês e Nelore. B Análise de Admixture das Raças Purunã, Angus, Canchim, Charolês e Nelore. C Árvore de filogenética usando a distância de Reynolds para as populações Purunã (PUR), Angus (ANG), Canchim (CAN), Charolês (CHL) e Nelore (NEL).	88
Figura 3 - Correlação entre métodos de estimativa de endogamia	90
Figura 4 - Classificação das corridas de homozigose (ROH - A) e regiões ricas em heterozigose (HER - B), por cromossomo, de acordo com o tamanho na raça Purunã e a porcentagem média dos cromossomos cobertos por ROHs/HERs. ..	93
Figura 5 - Árvores de filogenéticas. A Árvore filogenética para região em homozigose comparando as populações Purunã (PUR), Angus (ANG), Canchim	

(CAN), Charolês (CHL) e Nelore (NEL). **B** Árvore de filogenética da via de crescimento em regiões ricas em heterozigose96

CAPÍTULO 3

Figura 1 - Roteiro das análises realizadas. Min SNP: número mínimo de SNPs em um HER; Max Gap: distância máxima entre dois SNPs consecutivos (Kb); Min Length: comprimento mínimo de um HER (Kb); Max n hom: Número máximo de homozigotos permitido em um HER.....	124
Figura 2 - Efeito do parâmetro na detecção de regiões ricas em heterozigose	128
Figura 3 - Média e desvio padrão da detecção de regiões ricas em heterozigose em sequência do genoma completo (WGS) e painel comercial.....	129
Figura 4 - Classificação das regiões ricas em heterozigose (HER), por cromossomo, de acordo com o comprimento, e a porcentagem média de cobertura do cromossomo por HERs em cada um dos cenários com o número mínimo de marcadores iguais a 5 e 10	131
Figura 5 - Correlação de Spearman entre os cenários com o número mínimo de marcadores iguais a 5 e 10, diferencial baseado na heterozigosidade observada e esperada (diff), diversidade de nucleotídeos (PI) e teste D de Tajima.....	132
Figura 6 - Gráfico de dispersão da frequência do alelo menor (MAF) para cada marcador e estatística D de Tajima	133
Figura 7 - Distribuição das ilhas de regiões ricas em heterozigose através do genoma da raça Holandesa.....	134
Figura 8 - Desequilíbrio de ligação do mapa de calor para a ilha localizada no cromossomo 23	137

SUMÁRIO

Resumo.....	17
Abstract.....	18
INTRODUÇÃO.....	19
REFERENCIAS.....	24

CAPITULO 1

CARACTERIZAÇÃO DE CORRIDAS DE HOMOZIGOSE, REGIÕES RICAS EM HETEROZIGOSE E ESTRUTURA POPULACIONAL DE POPULAÇÕES DE BOVINOS SELECIONADAS PARA DIFERENTES OBJETIVOS DE SELEÇÃO.....	28
Resumo.....	29
Abstract.....	31
Introdução.....	32
Material e Métodos.....	34
Genótipos.....	34
Controle de qualidade.....	35
Corridas de homozigose.....	36
Regiões Ricas em Heterozigose.....	37
Estrutura genômica da população.....	38
Coeficiente de endogamia genômico.....	38
Tamanho efetivo populacional.....	39
Análises Funcionais.....	40
Comparação entre painéis de SNP e WGS.....	40
Resultados.....	41
Corridas de homozigose.....	41
Coeficientes de endogamia genômica e tamanho efetivo da população.....	41
Ilhas de homozigose.....	44
Regiões ricas em heterozigose.....	45
Comparação entre os resultados do painel SNP e WGS.....	49
Análises genômicas funcionais.....	50

Discussão	51
Coeficientes de endogamia genômica e tamanho efetivo da população.....	53
Ilhas de homozigose	55
Regiões ricas em heterozigose.....	56
Painéis SNP e comparação WGS	57
Conclusões	59
Referências.....	61

CAPITULO 2

CARACTERIZAÇÃO GENÉTICA E GENÔMICA DE UMA NOVA RAÇA COMPOSTA DE GADO DE CORTE (PURUNÃ) DESENVOLVIDA PARA PRODUÇÃO EM SISTEMA A PASTO	73
Resumo.....	74
Abstract	76
Introdução	78
Material e métodos.....	80
Estratificação da população	81
Análise de componentes principais.....	81
Análise de Admixture	81
Árvore de filogenética.....	82
Estrutura da população	83
Medidas de endogamia	83
Desequilíbrio de ligação	85
Tamanho efetivo da população	85
Consistência da fase gamética	85
Proporção de SNPs polimórficos e distribuições de SNPs por intervalo de MAF	86
Corridas de homozigose	86
Regiões ricas em Heterozigose.....	87
Análises funcionais.....	87
Resultados	87
Estratificação da população	87
Análise do componente principal.....	87

Árvore de filogenética.....	88
Estrutura da População	89
Endogamia	89
Desequilíbrio de ligação	90
Tamanho Efetivo da População	91
Consistência da fase gamética	91
Proporção de SNPs polimórficos e distribuição de SNPs por faixa de MAF	92
Corridas de homozigose	92
Regiões ricas em heterozigose	93
Ilhas ROH e HER e Análises Funcionais.....	94
Corridas de homozigose	94
Regiões ricas em heterozigose	96
Discussão	100
Medidas de endogamia.....	101
Desequilíbrio de ligação, tamanho efetivo da população e consistência da fase gamética	102
Corridas de homozigose	104
Região rica em heterozigose.....	105
Conclusões	106
Referências.....	107

CAPITULO 3

AVALIAÇÃO DE PARÂMETROS QUE INFLUENCIAM A IDENTIFICAÇÃO DE REGIÕES RICAS EM HETEROZIGOSE EM REBANHO HOLANDÊS COM BASE EM PAINEL DE SNP OU DADOS DE SEQUÊNCIA DO GENOMA COMPLETO	117
Resumo.....	118
Abstract	120
Introdução	121
Material e Métodos	123
Dados e controle de qualidade.....	123
Identificação de Regiões Ricas em Heterozigotos.....	125

Estimativa de heterozigosidade, diversidade de nucleotídeos e teste estatístico D de Tajima	126
Identificação de ilhas HER, desequilíbrio de ligação e anotação genômica	126
Comparação das ilhas HER da raça Holandesa com outras raças taurina	127
Resultados	127
O impacto dos parâmetros usados na identificação da HER	127
Classificação e distribuição de HERs em todo o genoma	129
Correlações	132
Tajima D estatística	133
Ilhas de regiões enriquecidas em heterozigotos, desequilíbrio de ligação e análises funcionais.....	133
Comparação entre raças	143
Discussão	145
Impacto dos parâmetros utilizados na identificação de HERs.....	145
Ilhas da região ricas em heterozigose	147
Conclusões	149
Referências	150

Resumo

O genoma animal é caracterizado por extensas regiões homozigóticas, em que haplótipos idênticos foram herdados ao longo das gerações. A extensão de cada corrida de homozigose é determinada, em parte, pelo número de gerações analisadas e pela conexão genética existente entre os animais avaliados. Por outro lado, extensas regiões heterozigóticas também podem ser observadas, as quais podem estar relacionadas ao cruzamento entre raças ou associadas ao desequilíbrio de ligação. Com isso, entende-se que um dos principais aspectos estudado no âmbito da genética populacional é a verificação da possível diversidade existente entre espécies, raças e rebanhos. A diversidade torna-se assim, um importante ponto de observação, visto que seu declínio pode resultar em menores respostas a seleção. Com isso, nosso objetivo foi investigar a estrutura populacional de diferentes raças bovinas, verificando a existência de regiões em homozigose e heterozigose no DNA bovino e alterações nas frequências dos genótipos devido a pressões seletivas, bem como compará-las com a finalidade de observar as divergências e/ou semelhanças genéticas existentes. Para tal, bovinos de diferentes raças e aptidões foram genotipados com painéis de diferentes densidades, juntamente com informações de sequências do genoma completo, e as informações adquiridas utilizadas para verificar as regiões do DNA autossomais dos bovinos. Essas regiões também foram utilizadas para verificação da existência de assinaturas de seleção, endogamia, desequilíbrio de ligação, tamanho efetivo populacional, além da identificação dos genes candidatos presentes nas regiões com alta homozigose e heterozigose. A caracterização uma nova raça bovina desenvolvida no sul do Brasil para sistemas de produção a pasto (Purunã), o qual apresentou proximidade genômica com raças utilizadas na formação, porém com particularidades que não permitem uma avaliação multirracial. Os resultados desse projeto permitiram caracterizar geneticamente algumas das principais raças de bovinos, possibilitando aprofundar o conhecimento genômico nas raças estudadas, bem como resultados que poderão servir de guia em estudos futuros na área de diversidade genética animal.

Palavras chaves: endogamia, frequência gênica, haplótipos, similaridade genética

Abstract

The animal genome is characterized by long homozygous regions, in which identical haplotypes had been inherited over generations. The extension of each run of homozygosity is determined, in part, by the generations' number analyzed and the genetic relationship among animals evaluated. Still, long heterozygous regions can be observed, and it can be related to crossbreeding or associated with linkage disequilibrium. Therefore, it is understandable that one of the main aspects of genetic population studies is the verification of possible diversity among species, breeds, and cattle. Diversity becomes an important point of observation since the decrease in diversity can result in lesser selection results. With this, our main goal was to investigate the population structure of different bovine breeds, checking the existence of homozygous and heterozygous regions at bovine DNA and possible changes in genotype frequencies due to selection pressure, as well as comparing them to observe the genetic differences and/or resemblances. Thus, bovines of different breeds and purposes were be genotyped with different densities SNP panels, together with whole genome sequence, and the information acquired used to check the bovine autosomal DNA regions. These regions also were checked for selection signature, inbreeding, linkage disequilibrium, effective population size, besides genes identification that appears in regions with high homozygosity and heterozygosity. The characterization of a new bovine breed, developed in Southern Brazil for production in pasture-based systems (Purunã), which showed genomic similarity with the breeds used in the formation of the breed, but with particularities that do not allow a multi-breed evaluation. Besides, the characterization of the impact of different parameters on the evaluation of the heterozygous-enriched region, proves that different parameters regulated the detection of heterozygous-enriched regions. The results of this project make it possible to genetically characterize some of the main breeds of cattle, making it possible to deepen the genomic knowledge of the studied breeds, as well as results that may serve as a guide for future studies in the area of animal genetic diversity.

Keywords: genetic frequency, genetic similarity, haplotypes, inbreeding

INTRODUÇÃO

A intensificação do uso de seleção genômica, aliado a avanços de técnicas reprodutivas tem facilitado a rápida dispersão de animais com alto valor genético em diversas populações de bovinos (ZANELLA *et al.*, 2018). Com isso, grandes avanços na produção animal têm sido alcançados principalmente para características de difícil mensuração, mensurações somente em um dos sexos, ou ainda, que não pudessem ser mensuradas em vida. Avanços produtivos mais rápidos também são uma das vantagens da utilização de ferramentas genômicas, uma vez que seleção pode ser realizada em animais ainda jovens, conforme o mérito genético estimado precocemente. Porém, verifica-se que um aumento nos níveis de endogamia ainda está presente nas populações, mesmo com o acesso ao conhecimento de segregações mendelianas com maior precisão (BAES *et al.*, 2019), ocasionando algumas perdas na diversidade genética populacional pelo aumento de homozigoses em determinadas regiões do genoma.

Com o aumento da homogeneidade genética, além do não acompanhamento de tal, resultados como baixas variações genéticas, baixas respostas a seleção, ou ainda, aparecimento de homozigoses deletérias podem surgir na população (FORUTAN *et al.*, 2018) prejudicando a produção animal. Com isso, observar o estado da diversidade genética da população, bem como modificações do genoma bovino, podem auxiliar no direcionamento genético em seleções de características de interesse econômico e potencializar ganhos das seleções genéticas realizadas.

Dada a importância da manutenção da variabilidade das populações na diversidade genética, é imprescindível o acompanhamento em conjunto das taxas de homozigose e heterozigose como ponto importante da diversidade em programas de melhoramento animal e o efeito das práticas de domesticação e seleção sobre estas (FERREIRA DE CAMARGO, 2019). A mensuração dos níveis de heterozigose podem fornecer fortes indícios de como os indivíduos estão relacionados em grandes populações, além de favorecer a diminuição de riscos de aparecimento de variantes deletéria, sendo possíveis fontes de proteção contra doenças ligadas a essas (SAMUELS *et al.*, 2016). Porém, poucos estudos ainda visam a utilização dessa na verificação das estruturas populacional, e no auxílio a mecanismos que venham a favorecer a variabilidade dentro de programas de melhoramento animal.

A estimação dos níveis de homozigose e heterozigose revela importante

informação sobre a diversidade animal, sendo a manutenção dos baixos níveis de endogamia bem como variação genética existente na população um desafio no setor pecuário, principalmente com a intensificação dos sistemas de produção (ZANELLA *et al.*, 2018). Sendo assim, o acesso a estrutura populacional bem como a análise da diversidade genética é uma importante parte da avaliação dos programas de melhoramento genético (LI *et al.*, 2020). Esse acesso as informações poderão auxiliar nos sistemas de acasalamento norteando seleções mais eficazes, uma vez que haverá a identificação da diversidade existente entre e dentro de populações.

Sabe-se que a diversidade genética não é estática, sendo um processo contínuo de criação e perdas de acordo com os processos utilizados sob a população. A manutenção de suficiente diversidade genética é importante para a sustentabilidade dos diferentes tipos de população selecionados (FORUTAN *et al.*, 2018), mesmos os processos aplicados sobre as populações afetando essa. Muitos dos trabalhos realizados até então incorporam as informações dos painéis de genotipagem de SNPs para identificar a atual situação da diversidade existente, os quais são comumente usados em programas de seleção genômica (SZMATOŁA *et al.*, 2016). Uma das diferentes abordagens ainda pouco utilizada nos estudos de genética de população em animais é a utilização da sequência do genoma completo para tais estudos. As abordagens baseadas no sequenciamento do genoma completo fornecem possibilidades adicionais para a investigação da diversidade genética de raças bovinas (WELDENEGODGUAD *et al.*, 2019), essas informações podem melhor caracterizar a população adicionando robustez e melhor acurácia aos dados encontrados.

REVISÃO DE LITERATURA

Os diferentes objetivos de seleção levaram a diferentes caracterizações entre as raças bovinas. Ao longo do tempo, as raças distintas estabeleceram características próprias que definiriam uma especificidade às populações, como produção, robustez, padrões raciais, aspectos qualitativos e quantitativos, entre outras. Para que fossem atingidos tais objetivos, genótipos superiores foram selecionados para as características desejadas, perpetuando-se ao longo das gerações.

Com a possibilidade de se avaliar o genoma animal, verificou-se que tal prática poderia propiciar a avaliação de regiões que determinassem a homozigidade de haplótipos, distribuídos por todo o genoma (KIM *et al.*, 2013), gerando longos trechos

ininterruptos conhecidos como corridas de homozigose ou ROH. Esses longos trechos homozigóticos são em decorrência da transferência de haplótipos pelos progenitores idênticos aos seus descendentes (PURFIELD *et al.*, 2012), normalmente relacionados a aproximação genética desses devido a algum ancestral em comum.

As ROH's tem sido estudadas como uma forma de verificar e distinguir a história das populações e suas estruturas em diversas espécies (PURFIELD *et al.*, 2012). Através do comprimento desses segmentos é possível estabelecer se a homozigose é oriunda de um processo endogâmico recente, onde os comprimentos desses segmentos são maiores, ou proveniente de um antigo ancestral em comum, onde curtos segmentos são distribuídos no genoma (FORUTAN *et al.*, 2018). Outra propriedade da aplicação de estudos de ROH's, é a verificação da composição genética populacional, na avaliação de possíveis assinaturas de seleção presentes nos rebanhos. Essas são tidas como alterações genotípicas em populações submetidas a alguma forma de pressão seletiva (natural ou artificial), as quais podem ser caracterizadas pelo aumento da frequência de alelos em um ou mais genes ou grupo de genes (BERTOLINI *et al.*, 2018). Com isso, é possível verificar se houve algum tipo de redução, eliminação ou alteração na variação genética populacional, ou ainda a identificação de genes envolvidos em processos relacionados à produtividade dos animais.

O estudo das ROH's também possibilitam avaliações secundárias do genoma, permitindo análises genéticas mais profundas, como a verificação de taxas de consanguinidades, mapeamento e identificação de alelos deletérios recessivos, caracterização da demografia, da estrutura e do histórico das populações e ainda estimar genomicamente o parentesco e a autozigosidade individual e populacional (REBELATO; CAETANO, 2018a). Através dessas, avaliações mais acuradas da caracterização e diversidade da população são possíveis de serem realizadas uma vez que o que se leva em consideração os próprios genótipos dos animais como base das informações (FORUTAN *et al.*, 2018).

A distribuição e o padrão com que as ROHs aparecem sob o genoma podem indicar padrões de eventos de seleção (seja natural relacionados a adaptabilidade da população ou artificial por seleção de características importantes por longo período) sendo um ponto chave nos programas de melhoramento animal (NANDOLO *et al.*, 2018). Com isso, torna-se possível visualizar e identificar padrões de haplótipos característicos de

raças ou espécies, fixados em consequência de pressões seletivas (REBELATO; CAETANO, 2018).

Dada a natureza aleatória dos eventos de recombinação, a ocorrência das ROH's é altamente heterogênea em todo o genoma, e os pontos de alta incidência de ROH's em um grande número de amostras podem ser indicativos de pressão seletiva (ZAVAREZ *et al.*, 2015), sendo essas pressões estabelecidas conforme a especificidade da raça. Essas regiões com altas quantidades de homozigose podem ser definidas como ilhas de ROH's e são regiões genômicas com redução da diversidade genética e, conseqüentemente, alta homozigose ao redor do locus selecionado que podem ter sido alvo de seleção positiva e forte pressão seletiva (MAIORANO *et al.*, 2018; PERIPOLLI *et al.*, 2018). Semelhante as análises de ROH's, as corridas de heterozigose (*runs of heterozygosity – ROHet*) ou regiões ricas em heterozigose (*heterozygous-enriched regions – HER*) são tidas como regiões do genoma com altos níveis de heterozigosidade alélica. Essas regiões estão ligadas a diversidade genética da população e podem contribuir com características importantes como fertilidade e taxas de sobrevivência (MC PARLAND; KEARNEY; BERRY, 2009), ou ainda como saúde e resistência a doenças (TSAIRIDOU *et al.*, 2018).

Apesar do conhecimento dos benefícios e acompanhamento da heterozigose dentro da produção animal, pouco se tem feito para identificar essas regiões com alta diversidade genética. Como benefício da aplicação dos estudos com HERs, esses podem fornecer indícios sobre alguns parâmetros como seleção negativa, introgressão e regiões com grande variação no genoma que podem estar atuando nas populações em estudo (MARRAS *et al.*, 2018), assim uma completa descrição da estrutura populacional pode estar sendo realizada ao se unir as metodologias de identificação de homozigose e heterozigose nos estudos de diversidade genética.

A diversidade genética é um dos principais pilares da biodiversidade dentro e entre populações, sendo importante a sua análise em programas de melhoramento genético devido a essa ser um indicativo da variação herdável existente dentro e entre populações (LI *et al.*, 2020). As avaliações por genotipagem permitem que haja a avaliação da diversidade populacional bem como a distância genéticas entre populações (PORTONETO *et al.*, 2013) identificando a estrutura genética existentes. Essa estrutura refere-se as diferenças genéticas que existem entre grupos e/ou populações de diferentes regiões geográficas e produtivas. Como resultado dessas diferenças genéticas podemos encontrar

diferentes efeitos nas frequências de alelos de variantes genéticas entre populações e quando essas não são levadas em consideração potenciais erros e viés nas estimativas podem acontecer nas respostas a seleção (TEO, 2008).

Ao se mensurar a diversidade genética entre população é possível estabelecer a situação genética populacionais e histórico de seleção estabelecido entre tais (MARTÍN-BURRIEL; GARCÍA-MURO; ZARAGOZA, 1999). O conhecimento da distância genômica entre rebanhos é útil para se explorar a heterozigosidade entre populações distintas, onde a homozigosidade pode estar estabelecendo algum efeito sobre determinada população (KELLEHER *et al.*, 2017), influenciando os potenciais ganhos da seleção. O objetivo desse trabalho foi analisar a estrutura populacional de diferentes populações de bovinos, contrapô-las e assim, melhor caracterizá-los para fins de conservação de recursos genéticos e melhoramento genético. Mais especificamente:

- Identificar e comparar as ROHs e HERs dentro e entre populações, caracterizando as diferenças no processo de seleção a longo prazo;
- Analisar a assinatura de seleção no genoma bovino das diferentes raças e as alterações em populações submetidas a pressão seletiva;
- Analisar a diversidade genômica existente entre as populações de bovinos;
- Contrapor os resultados obtidos com painéis de menor densidade com as sequencias completas do genoma bovino;
- Caracterizar genética e genomicamente a população de bovinos da raça Purunã, identificando os níveis de endogamia, tamanho efetivo populacional e estrutura e histórico da população.
- Caracterizar o efeito de diferentes parâmetros na identificação de HER em sequências completas do genoma e painéis de SNP na população de bovinos Holandês.

REFERENCIAS

BAES, C. F.; MAKANJUOLA, B. O.; MIGLIOR, F.; MARRAS, G.; HOWARD, J. T.; FLEMING, A.; MALTECCA, C. Symposium review: The genomic architecture of inbreeding: How homozygosity affects health and performance. **Journal of Dairy Science**, v. 102, n. 3, p. 2807–2817, 1 mar. 2019. . Acesso em: 30 out. 2019.

BERTOLINI, F.; SERVIN, B.; TALENTI, A.; ROCHAT, E.; KIM, E. S.; OGET, C.; PALHIÈRE, I.; CRISÀ, A.; CATILLO, G.; STERI, R.; AMILLS, M.; COLLI, L.; MARRAS, G.; MILANESI, M.; NICOLAZZI, E.; ROSEN, B. D.; VAN TASSELL, C. P.; GULDBRANDTSEN, B.; SONSTEGARD, T. S.; TOSSER-KLOPP, G.; STELLA, A.; ROTHSCHILD, M. F.; JOOST, S.; CREPALDI, P. Signatures of selection and environmental adaptation across the goat genome post-domestication. **Genetics Selection Evolution**, v. 50, n. 1, p. 57, 19 dez. 2018. Disponível em: <<https://doi.org/10.1186/s12711-018-0421-y>>. Acesso em: 17 fev. 2019.

FERREIRA DE CAMARGO, G. M. The role of molecular genetics in livestock production. **Animal Production Science**, v. 59, n. 2, p. 201, 21 jan. 2019. Disponível em: <<http://www.publish.csiro.au/?paper=AN18013>>. Acesso em: 1 dez. 2020.

FORUTAN, M.; ANSARI MAHYARI, S.; BAES, C.; MELZER, N.; SCHENKEL, F. S.; SARGOLZAEI, M. Inbreeding and runs of homozygosity before and after genomic selection in North American Holstein cattle. **BMC Genomics**, v. 19, n. 1, p. 98, 27 dez. 2018. . Acesso em: 14 fev. 2019.

KELLEHER, M. M.; BERRY, D. P.; KEARNEY, J. F.; MCPARLAND, S.; BUCKLEY, F.; PURFIELD, D. C. Inference of population structure of purebred dairy and beef cattle using high-density genotype data. **Animal**, v. 11, n. 1, p. 15–23, 22 jan. 2017. Disponível em: <http://www.journals.cambridge.org/abstract_S1751731116001099>. Acesso em: 28 fev. 2019.

KIM, E.-S.; COLE, J. B.; HUSON, H.; WIGGANS, G. R.; VAN TASSELL, C. P.; CROOKER, B. A.; LIU, G.; DA, Y.; SONSTEGARD, T. S. Effect of Artificial Selection on Runs of Homozygosity in U.S. Holstein Cattle. **PLoS ONE**, v. 8, n. 11, p. e80813, 14 nov. 2013. Disponível em: <<https://dx.plos.org/10.1371/journal.pone.0080813>>. Acesso em: 14 fev. 2019.

LI, J.; GUO, H.; ZONG, J.; CHEN, J.; LI, D.; LIU, J. Genetic diversity in centipede grass [*Eremochloa ophiuroides* (Munro) Hack.]. **Horticulture Research**, v. 7, n. 1, p. 4, 1 dez. 2020. Disponível em: <www.nature.com/hortres>. Acesso em: 7 jan. 2020.

MAIORANO, A. M.; LOURENCO, D. L.; TSURUTA, S.; TORO OSPINA, A. M.; STAFUZZA, N. B.; MASUDA, Y.; FILHO, A. E. V.; DOS SANTOS GONCALVES CYRILLO, J. N.; CURI, R. A.; DE VASCONCELOS SILVA, J. A. Assessing genetic architecture and signatures of selection of dual purpose Gir cattle populations using genomic information. **PLoS ONE**, v. 13, n. 8, 1 ago. 2018. . Acesso em: 30 out. 2019.

MARRAS, G.; WOOD, B. J.; MAKANJUOLA, B.; MALCHIODI, F.; PEETERS, K.; AS, P. van; BAES, C. F.; BISCARINI, F.; TURKEYS, H. Characterization of runs of homozygosity and heterozygosity-rich regions in a commercial turkey (*Meleagris gallopavo*) population. Em: Proceedings of the 11th World Congress of Genetics Applied to Livestock Production, Szwachkowski 2017., 2018, Auckland. [...]. Auckland: 2018. v. 11, p. 763–767.

MARTÍN-BURRIEL, I.; GARCÍA-MURO, E.; ZARAGOZA, P. Genetic diversity analysis of six Spanish native cattle breeds using microsatellites. **Animal Genetics**, v. 30, n. 3, p. 177–182, jun. 1999. Disponível em: <<http://doi.wiley.com/10.1046/j.1365-2052.1999.00437.x>>.

MC PARLAND, S.; KEARNEY, F.; BERRY, D. P. Purging of inbreeding depression within the Irish Holstein-Friesian population. **Genetics Selection Evolution**, v. 41, n. 1, p. 16, 21 dez. 2009. Disponível em: <<https://gsejournal.biomedcentral.com/articles/10.1186/1297-9686-41-16>>.

NANDOLO, W.; UTSUNOMIYA, Y. T.; MÉSZÁROS, G.; WURZINGER, M.; KHAYADZADEH, N.; TORRECILHA, R. B. P.; MULINDWA, H. A.; GONDWE, T. N.; WALDMANN, P.; FERENČAKOVIĆ, M.; GARCIA, J. F.; ROSEN, B. D.; BICKHART, D.; VAN TASSELL, C. P.; CURIK, I.; SÖLKNER, J. Misidentification of runs of homozygosity islands in cattle caused by interference with copy number variation or large intermarker distances. **Genetics Selection Evolution**, v. 50, n. 1, p. 43, 22 dez. 2018. . Acesso em: 29 out. 2019.

PERIPOLLI, E.; STAFUZZA, N. B.; MUNARI, D. P.; LIMA, A. L. F.; IRGANG, R.;

MACHADO, M. A.; PANETTO, J. C. do C.; VENTURA, R. V.; BALDI, F.; DA SILVA, M. V. G. B. Assessment of runs of homozygosity islands and estimates of genomic inbreeding in Gyr (*Bos indicus*) dairy cattle. **BMC Genomics**, v. 19, n. 1, p. 34, 9 dez. 2018. Disponível em:

<<https://bmcbgenomics.biomedcentral.com/articles/10.1186/s12864-017-4365-3>>.

Acesso em: 19 ago. 2020.

PORTO-NETO, L. R.; SONSTEGARD, T. S.; LIU, G. E.; BICKHART, D. M.; DA SILVA, M. V.; MACHADO, M. A.; UTSUNOMIYA, Y. T.; GARCIA, J. F.; GONDRO, C.; VAN TASSELL, C. P. Genomic divergence of zebu and taurine cattle identified through high-density SNP genotyping. **BMC Genomics**, v. 14, n. 1, p. 876, 2013. Disponível em: <<http://bmcbgenomics.biomedcentral.com/articles/10.1186/1471-2164-14-876>>. Acesso em: 28 fev. 2019.

PURFIELD, D. C.; BERRY, D. P.; MCPARLAND, S.; BRADLEY, D. G. Runs of homozygosity and population history in cattle. **BMC Genetics**, v. 13, n. 1, p. 70, 14 dez. 2012. . Acesso em: 17 fev. 2019.

REBELATO, A. B.; CAETANO, A. R. Runs of homozygosity for autozygosity estimation and genomic analysis in production animals. **Pesquisa Agropecuária Brasileira**, v. 53, n. 9, p. 975–984, set. 2018. Disponível em:

<[http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-204X2018000900975&lng=en&tlng=en)

[204X2018000900975&lng=en&tlng=en](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-204X2018000900975&lng=en&tlng=en)>. Acesso em: 29 out. 2019.

SAMUELS, D. C.; WANG, J.; YE, F.; HE, J.; LEVINSON, R. T.; SHENG, Q.; ZHAO, S.; CAPRA, J. A.; SHYR, Y.; ZHENG, W.; GUO, Y. Heterozygosity Ratio, a Robust Global Genomic Measure of Autozygosity and Its Association with Height and Disease Risk. **Genetics**, v. 204, n. 3, p. 893–904, 1 nov. 2016. . Acesso em: 1 dez. 2020.

SZMATOŁA, T.; GURGUL, A.; ROPKA-MOLIK, K.; JASIELCZUK, I.; ZABEK, T.; BUGNO-PONIEWIERSKA, M. Characteristics of runs of homozygosity in selected cattle breeds maintained in Poland. **Livestock Science**, v. 188, p. 72–80, 1 jun. 2016. Disponível em:

<<https://www.sciencedirect.com/science/article/pii/S1871141316300671>>. Acesso em: 24 jul. 2019.

TEO, Y. Y. **Common statistical issues in genome-wide association studies: A review**

on power, data quality control, genotype calling and population structure*Current Opinion in Lipidology*abr. 2008. Disponível em:

<<https://insights.ovid.com/crossref?an=00041433-200804000-00007>>. Acesso em: 28 fev. 2019.

TSAIRIDOU, S.; ALLEN, A. R.; PONG-WONG, R.; MCBRIDE, S. H.; WRIGHT, D. M.; MATIKA, O.; POOLEY, C. M.; MCDOWELL, S. W. J.; GLASS, E. J.; SKUCE, R. A.; BISHOP, S. C.; WOOLLIAMS, J. A. An analysis of effects of heterozygosity in dairy cattle for bovine tuberculosis resistance. **Animal Genetics**, v. 49, n. 2, p. 103–109, abr. 2018. Disponível em: <<http://doi.wiley.com/10.1111/age.12637>>.

WELDENEGODGUAD, M.; POPOV, R.; POKHAREL, K.; AMMOSOV, I.; MING, Y.; IVANOVA, Z.; KANTANEN, J. Whole-Genome Sequencing of Three Native Cattle Breeds Originating From the Northernmost Cattle Farming Regions. **Frontiers in Genetics**, v. 9, n. JAN, p. 728, 11 jan. 2019. Disponível em: <<https://www.frontiersin.org/article/10.3389/fgene.2018.00728/full>>. Acesso em: 1 dez. 2020.

ZANELLA, R.; LAGO, L.; DA SILVA, A.; PÉRTILLE, F.; DE CARVALHO, N.; DO CARMO PANETTO, J.; ZANELLA, G.; FACIOLI, F.; DA SILVA, M. Genetic Characterization of Indubrasil Cattle Breed Population. **Veterinary Sciences**, v. 5, n. 4, p. 98, 3 dez. 2018. Disponível em: <<http://www.mdpi.com/2306-7381/5/4/98>>.

ZAVAREZ, L. B.; UTSUNOMIYA, Y. T.; CARMO, A. S.; NEVES, H. H. R.; CARVALHEIRO, R.; FERENCAKOVIC, M.; PÉREZ O'BRIEN, A. M.; CURIK, I.; COLE, J. B.; VAN TASSELL, C. P.; DA SILVA, M. V. G. B.; SONSTEGARD, T. S.; SÖLKNER, J.; GARCIA, J. F. Assessment of autozygosity in Nellore cows (*Bos indicus*) through high-density SNP genotypes. **Frontiers in Genetics**, v. 5, n. JAN, p. 5, 29 jan. 2015. Disponível em: <<http://journal.frontiersin.org/article/10.3389/fgene.2015.00005/abstract>>. Acesso em: 19 ago. 2020.

CAPÍTULO 1

CARACTERIZAÇÃO DE CORRIDAS DE HOMOZIGOSE, REGIÕES RICAS EM HETEROZIGOSE E ESTRUTURA POPULACIONAL DE POPULAÇÕES DE BOVINOS SELECIONADAS PARA DIFERENTES OBJETIVOS DE SELEÇÃO

Artigo publicado na revista BMC Genomics – doi: 10.1186/s12864-022-08384-0

Resumo

O declínio nos níveis da diversidade genética em rebanhos pode resultar em redução a resposta à seleção, maior incidência de defeitos genéticos e depressão endogâmica. Nesse contexto, várias medidas têm sido propostas para avaliar o nível da diversidade genética em populações selecionadas. Portanto, os objetivos deste estudo foram: 1) investigar a estrutura populacional de 16 populações de bovinos de 15 diferentes raças puras ou compostas, selecionadas para diferentes objetivos de seleção; e, 2) identificar e comparar corridas de homozigose (*runs of homozygosity* - ROH) e regiões ricas em heterozigose (*heterozygous-enriched regions* - HER) com base em diferentes painéis de polimorfismo de nucleotídeo único (*single nucleotide polymorphism* - SNP) e dados de sequência do genoma completo (*whole-genome sequence* - WGS), seguidos por análises genômicas funcionais. Um total de 24.187 ROH foram encontradas em todas as populações de bovinos, com 55% classificados com o tamanho de 2-4 Mb. Quatorze ilhas de homozigose foram encontradas em cinco populações, onde quatro ilhas de ROH localizadas nos cromossomos BTA1, BTA5, BTA16 e BTA19 se sobrepuseram entre as raças Brahman (BRM) e Gir (GIR). A análise funcional dos genes encontrados nessas ilhas revelou genes candidatos conhecidos por desempenhar papéis nas vias da melanogênese, sinalização de prolactina e vias de sinalização de cálcio. As correlações entre as medidas de endogamia variaram de 0,02 a 0,95, onde os métodos com base em genótipos homozigotos (F_{HOM}), união de gametas (F_{UNI}) e variância aditiva genotípica (F_{GRM}) apresentaram fortes correlações entre si. Todos os métodos apresentaram correlações baixas a moderadas com o coeficiente de endogamia com base na corridas de homozigose (F_{ROH}). Para a HER, 3.576 corridas e 26 ilhas, distribuídas sobre todo o genoma autossomal, foram encontradas em regiões contendo genes relacionados principalmente ao sistema imunológico, indicando possível balanço de seleção. Embora as análises com WGS não tenham permitido a detecção dos mesmos padrões das ilhas, elas revelaram novas regiões não capturadas ao usar dados de painel de SNP. As populações bovinas que apresentaram maior quantidade de ROH e HER foram Senepol (SEN) e Montana (MON), respectivamente. Ilhas ROH sobrepostas foram identificadas entre as raças GIR e BRM, indicando uma possível conexão histórica entre as populações. A distribuição e o padrão de ROH e HER foram específicos de cada população, indicando que diferentes raças experimentaram processos de seleção divergentes ou diferentes processos genéticos.

Palavras-chave: autozigosidade; assinatura de seleção; coeficiente de endogamia; corridas de heterozigose

Abstract

A decline in the level of genetic diversity in livestock can result in reduced response to selection, greater incidence of genetic defects, and inbreeding depression. In this context, various metrics have been proposed to assess the level of genetic diversity in selected populations. Therefore, the main goals of this study were to: 1) investigate the population structure of 16 cattle populations from 15 different pure breeds or composite populations, which have been selected for different breeds goals; and 2) identify and compare runs of homozygosity (ROH) and heterozygosity-enriched regions (HER) based on different single nucleotide polymorphism (SNP) panels and whole-genome sequence data (WGS), followed by functional genomic analyses. A total of 24,187 ROH were found across all cattle populations, with 55% classified in the 2-4 Mb size group. Fourteen homozygosity islands were found in five populations, where four ROH islands located on BTA1, BTA5, BTA16, and BTA19 overlapped between the Brahman (BRM) and Gyr (GIR) breeds. A functional analysis of the genes found in these islands revealed candidate genes known to play a role in the melanogenesis, prolactin signaling, and calcium signaling pathways. The correlations between inbreeding metrics ranged from 0.02 to 0.95, where the methods based on homozygous genotypes (FHOM), uniting of gametes (FUNI), and genotype additive variance (FGRM) showed strong correlations among them. All methods yielded low to moderate correlations with the inbreeding coefficients based on runs of homozygosity (FROH). For the HER, 3,576 runs and 26 islands, distributed across all autosomal chromosomes, were found in regions containing genes mainly related to the immune system, indicating potential balancing selection. Although the analyses with WGS did not enable detection of the same island patterns, it unraveled novel regions not captured when using SNP panel data. The cattle populations that showed the largest amount of ROH and HER were Senepol (SEN) and Montana (MON), respectively. Overlapping ROH islands were identified between GIR and BRM breeds, indicating a possible historical connection between the populations. The distribution and pattern of ROH and HER are population specific, indicating that different breeds have experienced divergent selection processes or different genetic processes.

Keywords: autozygosity; runs of heterozygosity; inbreeding coefficient; signature of selection

Introdução

Desde a domesticação dos bovinos, que ocorreu há cerca de 10.000 anos (PITT *et al.*, 2019), mais de 1.000 raças (FELIUS *et al.*, 2011; FELIUS, 1995) foram desenvolvidas por meio da seleção para diferentes características e objetivos de criação. Portanto, bovinos são um valioso modelo animal para estudar mudanças genômicas em resposta a processos como seleção, cruzamento e domesticação (MASTRANGELO *et al.*, 2018b). A seleção genética para características específicas pode resultar em assinaturas de seleção, que se caracterizam por regiões genômicas com reduzida variabilidade genética, ou seja, maior concentração de alelos homozigotos (MARCHESI *et al.*, 2018). Nesse contexto, vários estudos têm revelado que uma das consequências da seleção intensiva é o aumento da homozigose (BAES *et al.*, 2019), resultando na perda da diversidade genética dentro das populações (FORUTAN *et al.*, 2018; PURFIELD *et al.*, 2012). Além disso, altos níveis de endogamia estão diretamente relacionados a uma maior incidência de ROH, que, se não for controlada, pode resultar em outros problemas, como anomalias congênitas (FORUTAN *et al.*, 2018) e depressão endogâmica (CURIK; FERENČAKOVIĆ; SÖLKNER, 2014).

Um efeito dessas concentrações de alelos homozigotos é o surgimento de ROH. Uma ROH é definida como o comprimento contínuo de genótipos homozigotos que estão presentes no genoma de um indivíduo devido aos progenitores transmitirem haplótipos idênticos aos seus descendentes (PURFIELD *et al.*, 2012). A identificação de ROH permite estimar parâmetros relativos à estrutura genética e histórico da população, incluindo autozigosidade e coeficientes de endogamia (REBELATO; CAETANO, 2018b). Dada a natureza aleatória da recombinação, a ocorrência de ROH é altamente heterogênea em todo o genoma, onde regiões com alta incidência de ROH em um grande número de amostras é indicativo de uma pressão de seleção sofrida por essa população (ZAVAREZ *et al.*, 2015). Estudos anteriores foram realizados para avaliar a incidência de ilhas ROH em populações bovinas (e.g., DIXIT *et al.* (2020) , PERIPOLLI *et al.* (2018)), bem como em muitas outras espécies (e.g., ONZIMA *et al.* (2018)).

Regiões ricas em heterozigose, também conhecidas como corridas de heterozigose, consistem na identificação de regiões genômicas com alta variabilidade, podendo fornecer informações sobre a diversidade populacional e histórico evolutivo (SAMUELS *et al.*, 2016). Por exemplo, Biscarini *et al.* (2020), Marras *et al.* (2018), e

Santos et al. (2021) identificaram HER em bovinos, aves e equinos, respectivamente. Este método visa identificar regiões genômicas com alta variabilidade genética, fornecendo informações sobre o nível de diversidade genética da população e histórico evolutivo (SAMUELS *et al.*, 2016), bem como identificar segmentos específicos no genoma onde a manutenção de maior diversidade genética pode ser benéfica (BISCARINI *et al.*, 2020).

A diversidade genética não é estática, sendo um processo contínuo de criação e perda de acordo com as mudanças evolutivas e de seleção de uma população. A manutenção de suficiente diversidade genética é importante para a sustentabilidade a longo prazo das populações (FORUTAN *et al.*, 2018; PERIPOLLI *et al.*, 2017; PURFIELD *et al.*, 2012). Grande parte dos estudos de diversidade genética foram realizada com base em dados de painéis de SNP, que são comumente usados em programas de melhoramento genômico (SZMATOŁA *et al.*, 2016). Alternativamente, WGS oferece uma oportunidade para avaliar com mais precisão a diversidade genética em raças bovinas (WELDENEGODGUAD *et al.*, 2019). Isso é possível devido à maior cobertura do genoma ao usar dados WGS. Além disso, o uso de dados de WGS pode revelar variantes raras e comuns em ROH/HER que não foram identificadas ao usar apenas dados de painéis de SNP. Portanto, novas possibilidades para entender melhor a estrutura genômica envolvida nas regiões em ROH/HER podem ser reveladas para novas populações ou características (CEBALLOS; HAZELHURST; RAMSAY, 2018).

Os principais objetivos deste estudo foram: 1) caracterizar a estrutura populacional de 16 populações de bovinos de 15 raças puras ou cruzadas selecionadas para diferentes objetivos de seleção; 2) quantificar e classificar ROH e HER em cada população com base em seu comprimento e parâmetros alternativos de detecção; 3) realizar análises de anotação funcional para identificar genes candidatos e vias envolvidas nas regiões genômicas com maior concentração de ROH e HER; 4) estimar coeficientes de endogamia e tamanho efetivo da população com base em informações genômicas; e, 5) comparar os resultados obtidos nas análises de SNP e WGS.

Material e Métodos

Genótipos

Um conjunto de dados genômicos de 16 populações de bovinos ($n = 2.415$) espalhados pelo mundo todo selecionados para diferentes propósitos foram usados neste estudo. O tamanho amostral de cada raça e a densidade dos painéis SNP utilizados são apresentados na Tabela 1. Esses conjuntos de dados foram fornecidos por: Universidade de Purdue (West Lafayette, IN, EUA) – dados do cruzamento Angus x Simental (ANGSIM - população F1); Universidade de São Paulo (Pirassununga, SP, Brasil) – dados da população composta Montana Tropical (MON); Pecuária Katayama Ltda (Guararapes, SP, Brasil) – forneceram os dados da raça Nelore com dois painéis SNP: 35K (NEL35) e 50K (NEL50); o banco de dados WIDDE (SEMPÉRÉ *et al.*, 2015)- <http://widde.toulouse.inra.fr/widde/>) dados de Angus (ANG), Borgou (BOR), Brahman (BRM), Creolo de Guadalupe (CGU), Charolês (CHL), Gir (GIR), Hereford (HFD), Holândes (HOL), Jersey (JER), Limousin (LMS), Senepol (SEN) e Santa Gertrudis (SGT). Para aumento do tamanho amostral em algumas raças, conjuntos de dados de painéis SNP de alta densidade e 50K foram mesclados, como o caso das raças ANG, BRM, GIR, HFD, JER e LIM. Isso feito através da plataforma WIDDE antes do download dos dados. Todas as coordenadas dos SNP foram atualizadas para o genoma de referência ARS-UCD 1.2 (ROSEN *et al.*, 2020) antes da realização das análises.

Tabela 1 - Rebanhos utilizados no estudo com as respectivas amostragens (N), sigla referente ao rebanho, densidade do chip de genotipagem e o banco de dados ao qual se encontram os genótipos

População	N	Abreviação	Densidade	BASE DE DADOS
ANGUS	99	ANG	46.989	WIDDE
ANGUS x SIMENTAL	487	ANGSIM	52.597	Purdue
BORGOU	158	BOR	52.497	WIDDE
BRAHMAN	70	BRM	46.989	WIDDE
CRIOLO DE GUADALUPE	140	CGU	52.497	WIDDE
CHAROLES	62	CHL	46.989	WIDDE
GIR	50	GIR	46.989	WIDDE
HEREFORD	61	HFD	46.989	WIDDE
HOLÂNDES	137	HOL	46.989	WIDDE
JERSEY	84	JER	46.989	WIDDE
LIMOUSINA	87	LMS	46.989	WIDDE
MONTANA	271	SEG	51.084	USP
NELORE 35K	209	NEL35	35.237	Katayama
NELORE 50K	192	NEL50	54.791	Katayama
SENEPOL	153	SEN	52.497	WIDDE
SANTA GERTRUDIS	55	SGT	46.989	WIDDE

Controle de qualidade

O controle de qualidade (CQ) dos genótipos foi feito separadamente para cada análise. Para ROH e HER, SNPs com call-rate inferior a 0,90, posições duplicadas, presente em cromossomos não autossomais ou sem posição definida foram removidos. Para as demais análises, a frequência do alelo menor ($MAF < 0,05$) e os parâmetros de equilíbrio de Hardy-Weinberg ($HWE < 10^{-6}$) também foram usados para filtrar os SNPs. A densidade dos painéis de genótipos e a quantidade de SNP descartados em cada QC são apresentados na Tabela 2.

Tabela 2 - Estatísticas descritivas dos conjuntos de dados genômicos após o controle de qualidade do genótipo

Raças	N	SNP	Call-rate	dup	NA/SPD	Total 1	MAF	HWE	Total 2
ANG	99	46.989	260	0	0	46.729	10.399	19	36.311
ANGSIM	487	52.597	0	0	4.014	48.583	9.633	31	38.919
BOR	158	52.497	1.315	0	0	51.182	16.525	161	34.496
BRM	70	46.989	1.047	0	0	45.942	19.456	44	26.442
CGU	140	52.497	3.435	0	0	49.062	10.428	130	38.504
CHL	62	46.989	401	0	0	46.588	9.352	18	37.218
GIR	50	46.989	1.260	0	0	45.729	24.289	33	21.407
HFD	61	46.989	252	0	0	46.737	8.816	105	37.816
HOL	137	46.989	1.350	0	0	45.639	9.416	16	36.207
JER	84	46.989	1.377	0	0	45.612	13.163	56	32.393
LMS	87	46.989	243	0	0	46.746	9.925	37	36.784
SEG	271	51.084	0	0	0	51.084	142	127	50.815
NEL 35	209	35.237	780	16	1.624	32.817	2.529	1.385	28.903
NEL 50	192	54.791	942	9	3.647	50.193	9.895	1.804	38.494
SEN	153	52.497	1.633	0	0	50.864	12.635	91	38.138
SGT	55	46.989	619	0	0	46.370	9.768	28	36.574

N: número de animais utilizados nas análises; dup: SNPs duplicados; NA / SPD: SNPs não autossômicos ou SNPs sem posições definidas; Total 1: número de SNPs usados para análises de ROH e HER; Total 2: número de SNPs usados na endogamia e análises de tamanho efetivo da população realizadas; MAF: frequência do alelo menor; HWE: Equilíbrio de Hardy-Weinberg; Todas as abreviaturas de raça são definidas na Tabela 1

Corridas de homozigose

O software PLINK v1.9 (PURCELL *et al.*, 2007) foi utilizado para a identificação de ROH com base nos seguintes critérios:

- a) 1 SNP heterozigoto e 1 ausente foram permitidos;
- b) A sobreposição de janelas utilizada foi de 0,05;
- c) A distância entre dois SNPs consecutivos não poderia ser superior a 1.000 kb;
- d) O comprimento mínimo de uma ROH de 500 kb;

- e) O número mínimo de SNPs consecutivos que criam uma ROH deve ser igual ou superior a 30;
- f) A densidade de 1 SNP a cada 50 kb;
- g) Uma janela genômica de 50 SNPs.

As ROHs foram classificadas nas seguintes classes: <2 Mb, 2-4 Mb, 4-8 Mb, 4-16 Mb e >16 Mb. As regiões de ROHs presentes em pelo menos 50% mais um dos indivíduos foram consideradas como ilhas de ROH, e foram utilizadas para posteriores análises funcionais.

Regiões Ricas em Heterozigose

O pacote *detectRUNS* (BISCARINI *et al.*, 2019) foi utilizado para a detecção de HER seguindo dois métodos – a metodologia baseada em janelas de SNP e a metodologia baseada em SNPs consecutivos. O método com janelas é usado para escanear o genoma e os parâmetros da janela selecionados para determinar se o SNP está incluído ou não em uma HER. A metodologia baseada em SNPs consecutivos verifica SNP por SNP no genoma. Para as análises das janelas de SNP, os seguintes parâmetros foram considerados:

- a) Uma janela de 50 SNPs;
- b) Um mínimo de 20 SNPs consecutivos que constituem uma ROH;
- c) Comprimento mínimo de 500 kb;
- d) A densidade de 1 SNP a 100 kb;
- e) Permitindo o número mínimo de dois SNP homocigotos e um ausente; e,
- f) O intervalo máximo entre SNPs consecutivos não pode ser maior que 1.000 kb.

Para a análise de SNPs consecutivos, foram considerados os seguintes parâmetros:

- a) Um número mínimo de 20 SNPs consecutivos constitui um HER;
- b) Comprimento mínimo de 500 kb;
- c) permitido um mínimo de dois SNP homocigotos e um ausente; e,
- d) O intervalo máximo entre SNPs consecutivos não pode ser superior a 1.000 kb.

As regiões genômicas que apresentaram pelo menos 10% da população com HER foram incluídas nas análises funcionais subsequentes.

Para entender melhor o impacto das modificações de parâmetros na identificação de HER, análises adicionais usando diferentes parâmetros foram realizadas para os

principais parâmetros utilizados na determinação de HER, incluindo: a distância entre SNPs consecutivos (500 a 5.000 kb), e número de SNPs homozigotos permitidos em uma HER (0 a 5) e número de SNPs ausentes permitidos em uma HER (0 a 5). Neste caso, as duas populações com maior quantidade de HER foram utilizadas para testar o impacto de alterações de parâmetros nos resultados de HER (MON e NEL50).

Estrutura genômica da população

Coeficiente de endogamia genômico

Cinco modelos de estimativas de coeficientes de endogamia foram analisados. O primeiro método foi baseado nos genótipos homozigotos observados e esperados (F_{HOM1}), calculados como (PURCELL *et al.*, 2007):

$$F_{HOM1} = \frac{H_{exp} - H_{obs}}{H_{exp}}$$

onde, H_{exp} é o valor esperado para genótipos homozigotos e H_{obs} é o valor observado para genótipos homozigotos.

O segundo método foi baseado na variância aditiva do genótipo (F_{GRM}), usando o seguinte modelo (VANRADEN, 2008):

$$F_{GRM} = \frac{[x_i - 2p_i]^2}{h_i - 1} \text{ no qual } h_i = 2p_i(1 - p_i)$$

onde, x_i é o número de cópias do alelo de referência do i^{th} SNP, p_i é a frequência do alelo de referência na população. Semelhante ao primeiro método, a metodologia F_{HOM2} foi baseada nos genótipos homozigotos seguindo o modelo:

$$F_{HOM2} = 1 - \frac{x_i * (2 - x_i)}{h_i}$$

Os modelos acima são todos dependentes da frequência alélica dos genótipos, por esta razão, um quarto modelo foi aplicado com base na correlação entre união de gametas (F_{UNI}) usando o seguinte modelo (YANG *et al.*, 2010):

$$F_{UNI} = \frac{[x_i^2 - (1 + 2p_i) * x_i + 2p_i^2]}{h_i}$$

O último método foi baseado na soma do comprimento individual de ROH dividido pelo comprimento total do genoma autossômico (F_{ROH}) usando a seguinte equação (MCQUILLAN *et al.*, 2008):

$$F_{ROH} = \frac{\sum_{i=1}^n f(ROH_i)}{\sum_{j=1}^A h(j)}$$

onde $f(ROH_i)$ é o comprimento ROH do indivíduo i^{th} , n é o total de regiões genômicas homozigóticas de cada indivíduo, $h(j)$ é o comprimento do cromossomo j^{th} , e A é o número de cromossomos autossômicos ($A = 29$). Ainda, para cada classe de ROH (<2 Mb, 2-4 Mb, 4-8 Mb, 4-16 Mb, >16 Mb, <8 Mb e >8 Mb), as estimativas de endogamia foram feitas dividindo a soma total dos segmentos de ROH pelo comprimento total do genoma autossômico bovino. Todos os coeficientes de endogamia genômicos foram calculados usando o software PLINK v1.9 (PURCELL *et al.*, 2007). A opção PROC CORR do software estatístico SAS (SAS INSTITUTE INC., 2013) foi usada para correlacionar as estimativas dos coeficientes de endogamia. Um mapa de calor foi criado para melhor visualização dos resultados através do pacote *ploty* (SIEVERT, 2020).

Tamanho efetivo populacional

O tamanho efetivo populacional (N_e) foi investigado com o método de relação entre as variâncias do desequilíbrio de ligação e o tamanho efetivo da população usando o software SNeP (BARBATO *et al.*, 2015) e a seguindo a seguinte fórmula (CORBIN *et al.*, 2012):

$$N_{e(T)} = \left(4f(c_t)^{-1} \left(E[r_{adj}^{|ct}]^{-1} - \alpha \right) \right)$$

onde N_e é o tamanho efetivo populacional na geração T , c_t é a taxa de recombinação para a distância física entre os marcadores, α é a probabilidade de ocorrência de mutação e r_{adj}

é o valor do desequilíbrio de ligação calculado pela correlação entre dois alelos em loci separados, assumindo o seguinte modelo:

$$r_{adj}(p_a, p_b, p_{ab}) = \frac{(p_{ab} - p_a p_b)^2}{p_a(1 - p_a)p_b(1 - p_b)}$$

onde p_a é a frequência do haplótipo-a, p_b é a frequência do haplótipo-b e p_{ab} é a frequência do haplótipo com o alelo a no primeiro locus e o alelo b no segundo locus.

Análises Funcionais

As regiões genômicas consideradas como ilhas ROH e HER foram usadas para anotações genômicas. O pacote GALLO (FONSECA *et al.*, 2020) foi utilizado para a anotação de genes nessas regiões, com os dados anotados para *Bos taurus* do banco de dados Ensembl (https://www.ensembl.org/Bos_taurus/Info/Index), versão ARS-UCD1.2 (ROSEN *et al.*, 2020). Posteriormente, o pacote WebGestaltR (WANG *et al.*, 2020) foi usado para as análises de ontologia gênica (*Gene Ontology* - (GO)) para identificar processos biológicos, funções moleculares e componentes celulares nos quais os genes candidatos estão envolvidos.

Comparação entre painéis de SNP e WGS

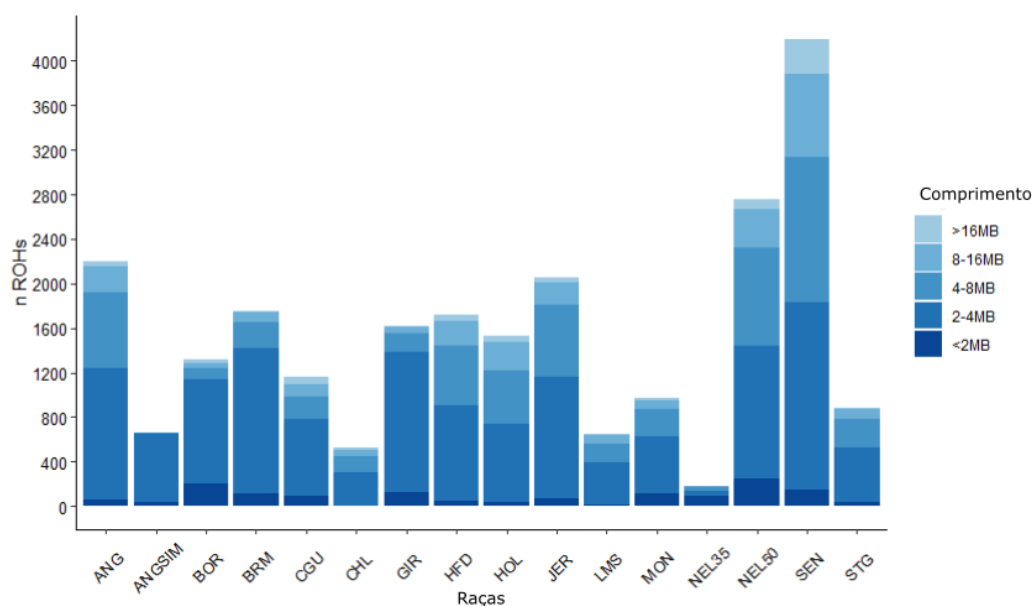
Os resultados da análise do painel de SNP, tanto para análises ROH quanto HER, foram comparados com os resultados usando dados WGS. Os dados do WGS foram obtidos através do banco de dados do “*The 1000 Bull Genomes Project – Run 8*” (HAYES; DAETWYLER, 2019). Um total de 1.842 animais de 138 raças entre *Bos taurus taurus* e *Bos taurus indicus* e seus cruzamentos. Nas análises, foram consideradas apenas raças em comum com os painéis de SNP, o que configurou um total de 914 animais. Marcadores com call-rate menor que 0,90, posições duplicadas, não autossomais ou sem posição definida para foram removidos. As análises foram realizadas separadamente por cromossomo. Os parâmetros usados para identificar ROHs e HER nos dados WGS foram os mesmos usados nas análises de painel SNP, exceto o número de SNPs heterozigotos/homozigotos (3) e SNPs ausentes (5), onde tais parâmetros foram adaptados do trabalho de Ceballos, Hazelhurst e Ramsay (2018).

Resultados

Corridas de homozigose

O grupo que apresentou maior número de ROH foi a raça SEN (Figura 1) com 4.198 ROHs distribuídos por todo o cromossomo autossomal. A população com menor quantidade de ROH foi o grupo Nelore genotipado com o painel 35K (NEL35), que também apresentou o menor percentual de indivíduos com pelo menos uma ROH por animal (63,16%). A maioria os animais de todas as populações apresentaram pelo menos uma ROH, exceto alguns casos em animais das raças Angus x Simental (ANGSIM), MON, NEL35 e SEN.

Figura 1 - Classificação das corridas de homozigose (ROH) de acordo com o comprimento e as raças bovinas



No total, foram identificados 24.187 ROHs, distribuídos por todo o cromossomo autossomal. A maioria das ROHs foram classificadas como 2-4 Mb, representando 55% de todas as ROHs encontradas. Em resumo, apenas 14% das ROHs eram maiores que 8 Mb; destes, 24% eram maiores que 16 Mb.

Coefficientes de endogamia genômica e tamanho efetivo da população

A população ANGSIM apresentou a menor taxa de endogamia (-0,026), exceto

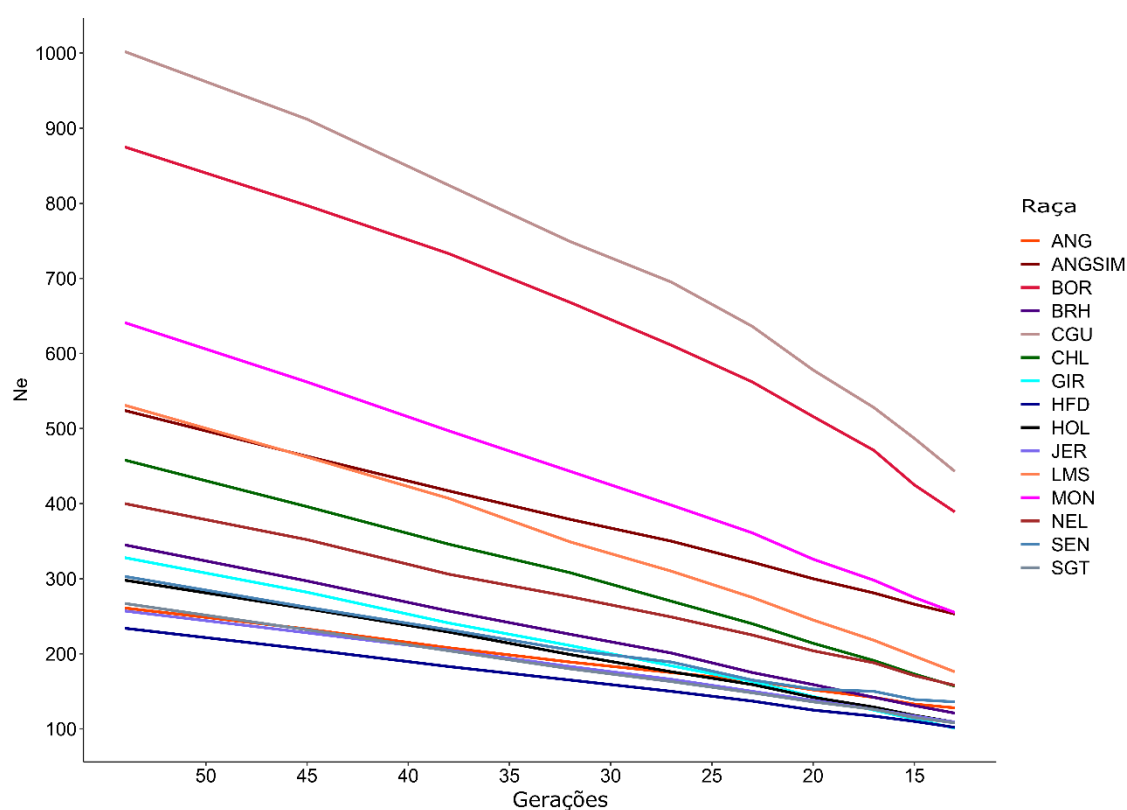
Tabela 3 - Média dos coeficientes de endogamia estimados para as cinco metodologias de cálculo de endogamia

	F_{HOM1}	F_{GRM}	F_{HOM2}	F_{UNI}	F_{ROH}	<2 MB	2-4 MB	4-8 MB	8-16 MB	>16 Mb	<8 MB	>8 Mb
ANG	-0,003	-0,003	-0,003	-0,003	0,043	0,000	0,014	0,015	0,010	0,004	0,029	0,014
ANGSIM	-0,026	-0,024	-0,024	-0,024	0,001	0,000	0,001	0,000	0,000	0,000	0,001	0,000
BOR	0,002	0,002	0,002	0,002	0,012	0,001	0,006	0,001	0,001	0,002	0,008	0,004
BRM	0,014	0,011	0,011	0,011	0,036	0,001	0,020	0,007	0,006	0,002	0,028	0,008
CGU	0,024	0,024	0,024	0,024	0,018	0,000	0,005	0,003	0,004	0,006	0,009	0,009
CHL	-0,004	-0,004	-0,004	-0,004	0,018	0,000	0,005	0,005	0,004	0,004	0,011	0,008
GIR	-0,011	-0,011	-0,011	-0,011	0,043	0,002	0,027	0,007	0,005	0,003	0,036	0,008
HFD	0,086	0,087	0,087	0,087	0,061	0,001	0,016	0,019	0,016	0,009	0,037	0,024
HOL	-0,012	-0,012	-0,012	-0,012	0,026	0,000	0,006	0,008	0,008	0,004	0,014	0,012
JER	0,023	0,030	0,030	0,030	0,047	0,001	0,015	0,017	0,010	0,004	0,033	0,015
LMS	-0,007	-0,006	-0,006	-0,006	0,015	0,000	0,005	0,004	0,004	0,002	0,009	0,006
SEG	-0,017	-0,018	-0,018	-0,018	0,006	0,000	0,002	0,002	0,001	0,001	0,004	0,002
NEL35	-0,019	-0,019	-0,019	-0,019	0,002	0,000	0,000	0,000	0,000	0,001	0,000	0,000
NEL50	-0,014	-0,013	-0,013	-0,013	0,030	0,001	0,007	0,010	0,008	0,004	0,018	0,012
SEN	0,003	0,003	0,003	0,003	0,075	0,001	0,013	0,019	0,021	0,021	0,033	0,043
SGT	-0,010	-0,010	-0,010	-0,010	0,030	0,001	0,010	0,010	0,007	0,003	0,021	0,009

F_{HOM1} : Coeficiente de endogamia baseado no número de genótipos homozigotos observados e esperados; F_{GRM} : Coeficiente de endogamia baseado na variância genotípica aditiva; F_{HOM2} : Coeficiente de endogamia baseado na homozigose dos genótipos (semelhante a F1); F_{UNI} : Coeficiente de endogamia baseado na correlação entre gametas unidos; F_{ROH} : Coeficiente de endogamia baseado no comprimento dos ROH e no comprimento total do genoma autossômico

O tamanho efetivo da população (N_e) para cada população da 54^a à 13^a geração mais recente é apresentado na Figura 3 e, como esperado, o N_e de cada população teve uma redução ao longo dos anos. A população com maior N_e , na geração mais recente, foi a Crioula de Guadalupe (CGU - 443), e o menor N_e foi encontrado em HFD (101). Em média, as estimativas de N_e diminuíram cerca de 59,6% nas últimas gerações, sendo as populações GIR, Limousin (LMS), Charolês (CHL), BRM, Holândes (HOL) e NEL a apresentarem as maiores reduções de N_e .

Figura 3 - Comportamento do tamanho efetivo da população (N_e) ao longo das últimas gerações



Ilhas de homozigose

A mais longa ilha de ROH foi encontrada no cromossomo BTA16 na raça BRM (Tabela 4), com tamanho aproximado de 14 Mb, incluindo 146 SNPs e 229 genes (Tabela S1). A menor ilha de ROH (3 Mb e 63 SNPs) foi identificada na população ANGSIM presente em mais de 75% dos animais ANGSIM incluídos nas análises. Essa região é conhecida por codificar 35 genes.

Tabela 4 - Ilhas de homozigose encontradas em diferentes cromossomos e grupos de indivíduos

n animal: número de animais avaliados; n ROH: número de ROHs encontrados na posição; %: percentagem da população que apresentou esta ilha; CHR: cromossomo, início: início da ROH; fim: fim da ROH;

Raça	n animal	n ROH	%	CHR	começar	fim	comprimento	n SNP
ANGSIM	487	369	75,77	16	62.578.656	66.253.552	3.674.896	62
BRM	70	36	51,43	1	78.237.770	84.586.062	6.348.292	58
BRM	70	36	51,43	5	105.576.062	117.735.828	12.159.766	171
BRM	70	39	55,71	8	56.051.150	63.444.254	7.393.104	58
BRM	70	36	51,43	16	44.071.454	58.289.347	14.217.893	146
BRM	70	39	55,71	19	42.110.400	46.627.006	4.516.606	74
GIR	50	31	62,00	1	80.333.027	84.911.107	4.578.080	80
GIR	50	46	92,00	5	110.192.579	116.240.339	6.047.760	110
GIR	50	30	60,00	14	21.914.329	26.212.648	4.298.319	56
GIR	50	26	52,00	16	45.727.235	52.149.496	6.422.261	111
GIR	50	31	62,00	19	42.054.880	46.678.246	4.623.366	82
JER	84	43	51,19	6	100.066.570	110.600.517	10.533.947	180
SEN	153	99	64,71	1	776.231	10.605.227	9.828.996	160
SEN	153	82	53,59	20	36.135.896	42.174.483	6.038.587	78

comprimento: comprimento ROH; n SNP: número de SNPs que o ROH cobre; Todas as abreviaturas de raça são definidas na Tabela 1.

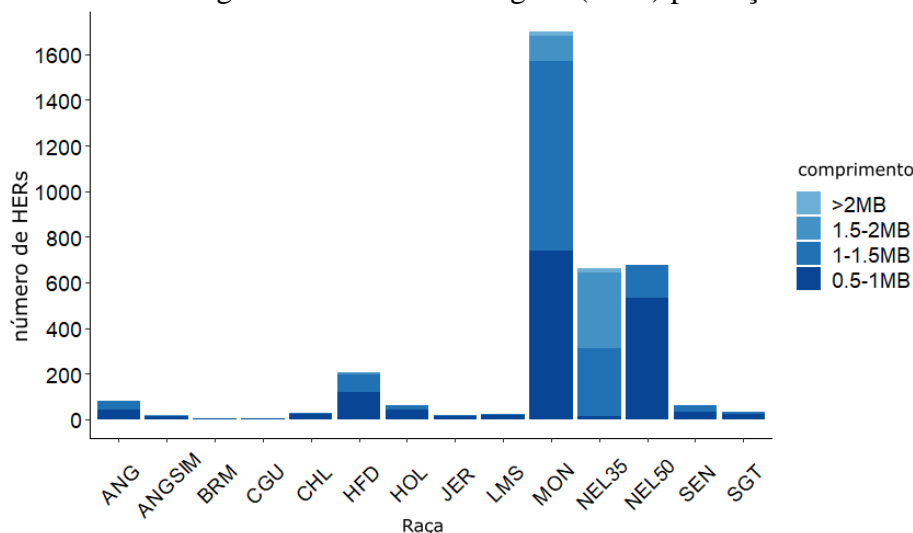
A ilha de ROH presente em maior proporção da população (92%) foi encontrada no cromossomo BTA5 para a raça GIR. Essa ilha também está presente na raça BRM, em menor proporção dos animais (51,43%). Uma sobreposição de ilhas de ROH foi observada nos cromossomos BTA1, BTA16 e BTA19 em ambas as raças GIR e BRM. Essas raças também apresentaram o maior número de ilhas de ROH (5 ilhas em cada raça). No total, as ilhas de ROH nessas raças contêm 556 e 863 genes nas raças GIR e BRM, respectivamente, que estão envolvidas nos principais processos biológicos, componentes celulares e funções moleculares.

Regiões ricas em heterozigose

Dois métodos foram usados para detectar HERs: o método de janelas e o método

de SNP consecutivos (BISCARINI *et al.*, 2019). Apesar de ser utilizado, o método de janela não detectou nenhuma corrida como o método de SNPs consecutivos. Portanto, serão apresentados apenas os resultados obtidos pelo método consecutivo (Figura 4).

Figura 4 - Número de regiões ricas em heterozigose (HER) por raça estudada.



No total, foram encontrados 3.576 HERs para todas as populações, sendo o maior número de HERs encontrado na população MON (1.702 corridas). O menor número de HERs foi encontrado na população CGU (2 HERs). Em geral, o maior percentual de HER foram classificados como comprimento de 0,5-1 Mb (45,13%) e 1-1,5 Mb (41,11%). Os percentuais de HER maiores que 2 Mb foram detectados apenas em 1% dos HER. A distribuição de HER por cromossomo, em cada população, mostram que o cromossomo BTA5 apresentou o maior número de HER, com 12% de todas as HER (Figura S2).

As regiões de ilhas de HER presentes em pelo menos 10% dos animais dentro de cada população podem ser observadas na Tabela 5. A maior HER foi encontrada em 20,3% da população MON com comprimento de 4,25 Mb (Tabela 5). Esta HER foi encontrada no cromossomo BTA5, mesma região considerada como uma ilha ROH nas raças BRM e GIR. Esta região abriga 114 genes, onde 71 deles são genes codificadores de proteínas. A menor região com concentração de HER foi encontrada no cromossomo BTA23 da população Nelore genotipada com painel 50K (NEL50), com comprimento aproximado de 0,7 Mb. Esta região abriga apenas um gene: *KHDRBS2*.

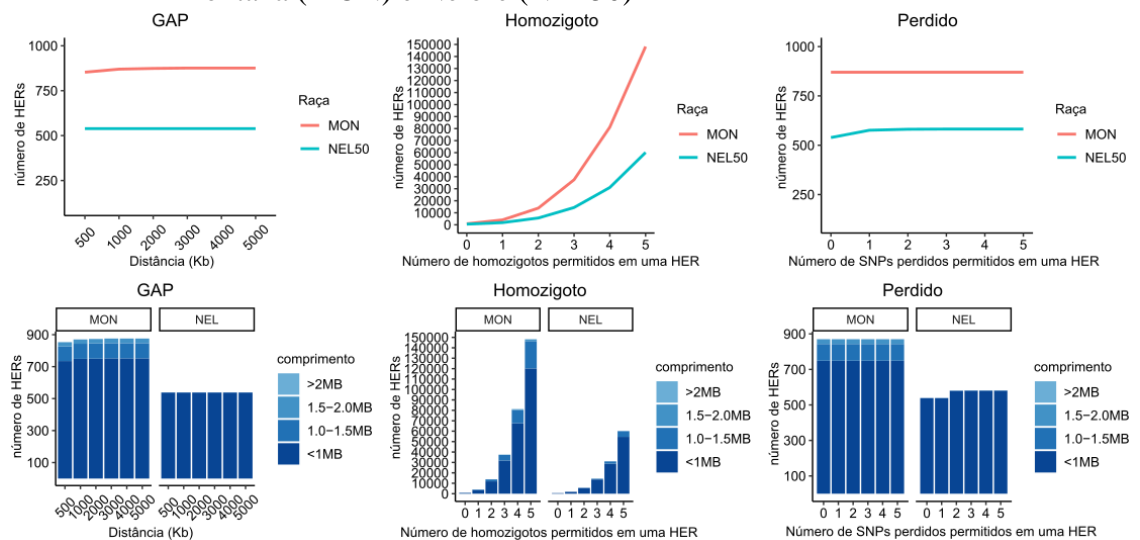
Tabela 5 - Regiões ricas em heterozigose (HER), nas diferentes populações, que aparecem em pelo menos 10% dos indivíduos da população

Raça	n animal	nHER	%	CRM	inicio	fim	comprimento	nSNP
ANG	99	20	20,20	13	40.318.645	41.433.750	1.115.105	21
HFD	61	8	13,11	1	32.091.115	33.034.361	943.246	23
HFD	61	9	14,75	1	101.195.925	104.830.589	3.634.664	47
HFD	61	9	14,75	3	75.829.200	76.986.902	1.157.702	27
HFD	61	9	14,75	3	89.652.649	90.628.822	976.173	25
HFD	61	13	21,31	6	78.147.926	79.620.230	1.472.304	25
HFD	61	12	19,67	14	50.316.790	52.240.902	1.924.112	25
HFD	61	7	11,48	20	46.124.910	46.912.182	787.272	20
HFD	61	7	11,48	24	28.353.262	29.204.162	850.900	20
SEG	271	67	24,72	5	48.569.574	50.434.637	1.865.063	37
SEG	271	62	22,88	5	95.248.428	96.123.569	875.141	22
SEG	271	55	20,30	5	110.220.384	114.471.702	4.251.318	36
SEG	271	52	19,19	12	44.975.037	46.232.946	1.257.909	21
SEG	271	50	18,45	16	58.037.089	59.263.750	1.226.661	24
SEG	271	35	12,92	19	60.384.670	61.177.248	792.578	23
SEG	271	38	14,02	27	42.506.836	43.575.359	1.068.523	30
NEL35	209	38	18,18	23	103.505	1.712.734	1.609.229	22
NEL50	192	83	43,23	1	63.642.322	65.584.086	1.941.764	37
NEL50	192	24	12,50	2	36.202.336	37.386.478	1.184.142	23
NEL50	192	30	15,63	5	75.447.962	76.805.229	1.357.267	29
NEL50	192	26	13,54	6	77.309.507	78.206.076	896.569	24
NEL50	192	33	17,19	11	16.880.546	18.643.939	1.763.393	24
NEL50	192	23	11,98	18	23.656.888	24.920.756	1.263.868	24
NEL50	192	47	24,48	23	588.741	1.284.183	695.442	22
SEN	153	26	16,99	8	49.323.954	50.336.594	1.012.640	27
SGT	55	14	25,45	5	76.209.127	76.942.872	733.745	22

n animal: número de animais avaliados; n HER: número de HER encontrado na posição; %: percentagem da população que apresentou esta ilha; CRM: cromossomo; começo: começo da HER; fim: fim da HER; comprimento: comprimento HER; n SNP: número de SNPs que o HER cobre; Todas as abreviaturas de raça são definidas na Tabela 1.

O HER com maior proporção (43,23%) foi encontrado no cromossomo BTA1 em NEL50. Esta região abriga 28 genes. Ilhas de HER sobrepostas entre os grupos foram observadas no cromossomo BTA5 para as populações NEL50 e Santa Gertrudis (SGT) (próximo à região de 76 Mb) e para HFD e NEL50 (região de 78 Mb). Também foram avaliados o impacto de diferentes parâmetros na detecção de HER, conforme mostrado na Figura 5. O principal parâmetro que afetou a quantidade de HER foi o número de SNPs homozigotos permitidos dentro das HERs.

Figura 5 - Número de regiões ricas em heterozigose (HER) e classificação pelo comprimento, com base em diferentes valores de parâmetros nas raças Montana (MON) e Nelore (NEL50)



Quando zero SNPs homozigotos foram permitidos nas análises de HER, 870 e 539 HERs foram encontrados para as populações MON e NEL50, respectivamente. Quando cinco SNPs homozigotos foram permitidos, 148.233 HERs foram encontrados para MON e 60.155 para NEL50. Assim, a modificação de zero a cinco SNPs homozigotos permitidos nas análises representaram um aumento no número de HERs de 17.039% e 11.160%, para MON e NEL50, respectivamente. No entanto, nenhum efeito relevante foi observado quando a distância máxima entre dois SNPs consecutivos foi aumentada, demonstrando que o aumento no parâmetro não afetou o número de HERs obtidos. Por fim, outro parâmetro testado foi o número de SNPs ausentes permitidos nas análises de HER, que para MON não foi observado nenhum efeito relevante, mas para NEL50 foi detectado um aumento de 8% do HER encontrado quando o parâmetro foi alterado de

zero para cinco SNPs ausentes permitidos.

Comparação entre os resultados do painel SNP e WGS

As ilhas apresentadas na análise com painel SNP para as raças GIR e BRM não foram encontradas na análise com WGS (Figura 6). Esse resultado também se repetiu nos demais grupos avaliados. Um ponto interessante foi observado no cromossomo BTA20, onde os animais HOL, com base na análise com WGS, tiveram ROHs próximos à ilha encontrada no SEN nas análises usando o painel de SNP (Figura 7). Para as análises HER, o padrão foi semelhante ao que ocorre na análise ROH onde não houve repetição das ilhas na análise com WGS.

Figura 6 - Comparação entre corridas de painel SNP de homozigose e análises WGS para o cromossomo 1

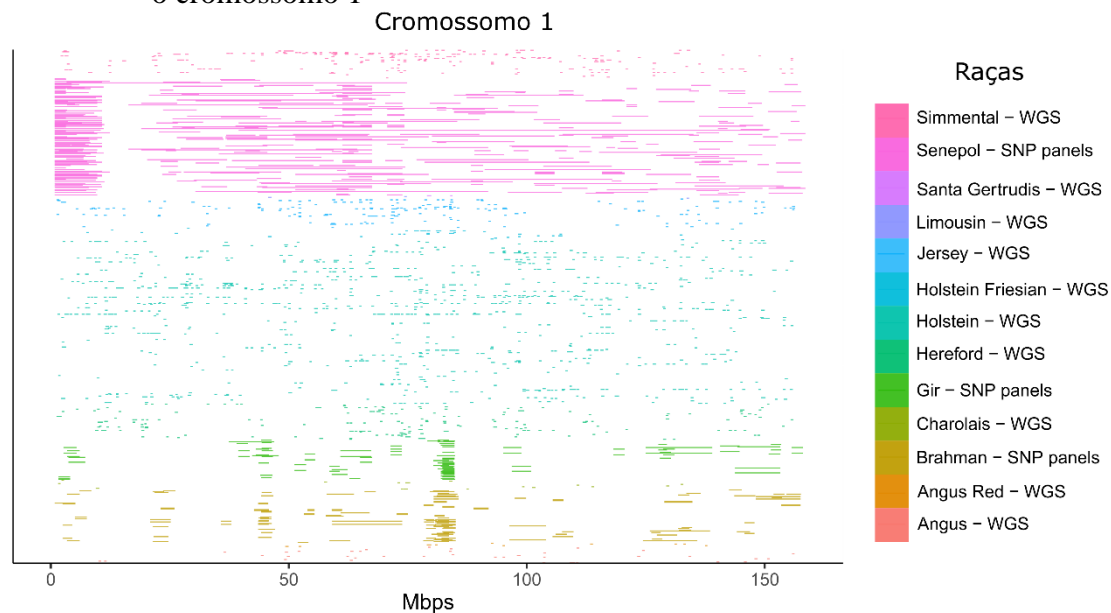
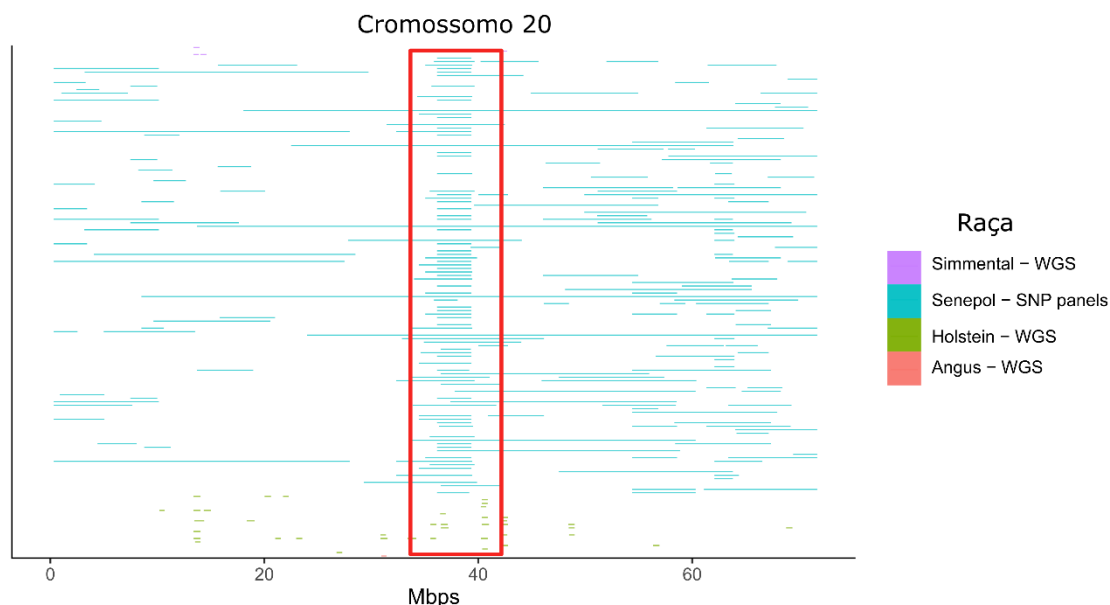


Figura 7 - Comparação entre corridas de painel SNP de homozigose e análises WGS para o cromossomo 20



Análises genômicas funcionais

Em resumo, 1.662 genes foram encontrados nas ilhas de ROH e 850 genes nas de HER. Os genes encontrados nas ilhas de ROH estão relacionados a 379 vias, 389 componentes celulares, 400 processos biológicos e 319 funções moleculares. Os genes encontrados nas ilhas de HER estão relacionados a 217 vias, 180 componentes celulares, 373 processos biológicos e 224 funções moleculares. As ilhas de ROH das populações GIR e BRM contêm genes relacionados à via da melanogênese como *DVL3* no BTA1, *EP300* no BTA5, *CREB3* no BTA8 (somente para os animais BRM), *DVLI* em BTA16 e *FZD2* e *WNT3* no BTA19.

A via de sinalização da prolactina foi outra importante via observada neste estudo. Quatro genes identificados na raça BRM estão envolvidos nesta via (*CCND2*, *STAT5B*, *STAT5A* e *STAT3*); na raça GIR, três genes estão envolvidos (*STAT5B*, *STAT5A* e *STAT3*), enquanto que para Jersey (JER) e SEN, apenas os genes *MPK10* e *PRLR* foram envolvidos, respectivamente. Outra via em comum entre os grupos foi a via de sinalização de cálcio em ANGSIM e JER, onde foram encontrados os genes *CACNA1E* e *DRD5* relacionados a esta via.

Em relação ao HER, a maioria dos genes encontrados estão envolvidos com o sistema imune como *ANXA1*, *INFG*, *KLRD1*, *RAC2* e *NKG2C*. Para SEN, o gene identificado *ANXA1* é conhecido por influenciar processos biológicos como resposta

imune do tipo 2 e resposta a antibióticos, proliferação, migração de leucócitos, respostas de imunidade adaptativa e desenvolvimento do sistema imunológico. Na população MON, foram identificados seis genes conhecidos pela via de processamento e apresentação de antígenos: *IFNG*, *ENSBTAG00000046268*, *NKG2C*, *KLRD1*, *ENSBTAG00000000966* e *CD4*. Ainda, em NEL50, a via de migração transendotelial leucocitária é influenciada por três genes: *NCF4*, *RAC2* e *MMP2*, ou por genes que estão envolvidos na proliferação leucocitária como *CD80*, *MARCHF7* e *RAC2*.

Discussão

Há uma grande variabilidade na literatura sobre os parâmetros utilizados para identificar as ROH, o que dificulta a comparação dos resultados de diferentes estudos, conforme reforçado por Peripolli, et al. (2017) e mostrado na Tabela 6. Além dos parâmetros ROH, também avaliamos o impacto da densidade do painel SNP na detecção de ROH.

Tabela 6 - Parâmetros para identificação de corridas de homozigose em diferentes estudos com diferentes densidades de painéis de genotipagem

Estudos	Painel SNP	het	hom	trhs	Gap (kb)	min (kb)	nSNP	Densidade		janela
								snp	kb	
Presente estudo	35K, 50K	1	1	0,05	1000	500	30	1	50	50
PERIPOLLI et al. (2018a)	20K, 30K, 50K, 70K E HD	1	5	-	500	1000	100	1	50	50
MASTRANGELO et al. (2018))	50K	1	1	-	1000	4000	30	1	100	-
ZAVAREZ et al. (2015))	HD	2	5	-	500	100	30	1	100	-
PERIPOLLI et al. (2020)	30K	0	2	-	1000	500	15	1	120	40
PERIPOLLI et al. (2018b)	HD	1	5	0,05	500	1000	100	1	50	50
DE SOUZA FONSECA et al. (2016))	50K	0	5	-	1000	-	15	-	-	-
ZANELLA et al. (2018)	HD	1	1	-	-	1000	-	-	-	50
VENTURA et al. (2020)	HD	1	5	0,05	500	1000	100	1	50	50

het: número de SNPs heterozigotos permitidos em uma ROH; miss: número de SNPs ausentes permitidos em uma ROH; trhs : sobreposição de janela; min: tamanho mínimo de uma ROH; nSNP : número mínimo de SNP que compõem uma ROH.

Conforme demonstrado na Tabela 6, a diversidade de parâmetros escolhidos entre os estudos leva a uma dificuldade na comparação de estudos e/ou raças (PERIPOLLI *et al.*, 2017). Conforme mencionado por Biscarini *et al.* (2020), poucos estudos objetivaram avaliar o impacto de diferentes parâmetros na detecção de ROH, sendo que tais parâmetros ainda não estão bem estabelecidos. Esta definição é de fundamental importância, pois tem efeito direto nos resultados obtidos (MACCIOTTA *et al.*, 2021); não apenas na detecção de ROH, mas também nas análises secundárias (por exemplo, coeficiente de endogamia baseado em ROH).

Foram avaliadas 16 populações de bovinos selecionados para diferentes objetivos de seleção, comparando ROHs e HERs. A maioria das ROHs foram definidas como sendo de 2-4 Mb e levando em consideração que os eventos de recombinação que acontecem a cada geração quebram os segmentos homozigotos em haploblocos menores, essas corridas provavelmente foram formadas em gerações mais antigas. Howrigan, Simonson e Keller (2011) estimaram que os comprimentos de ROH de 10 Mb, 5 Mb e 2,5 Mb seriam correlacionados com 5, 10 e 20 gerações atrás, respectivamente. Além disso, Cardoso *et al.* (2020) estimaram que o comprimento de ROH maior que 16 Mb foram formados há menos de três gerações e a ROH menor que 8 Mb foi formado há mais de seis gerações atrás.

SEN apresentou o maior número de ROHs e o maior número de segmentos de ROHs maiores que 16Mb. Isso sugere que eventos recentes de endogamia ocorreram com mais frequência nessa população. Como esperado, as populações mestiças ou compostas, como ANGSIM e MON, apresentaram a menor quantidade de ROHs. A verificação da incidência de ROH em raças compostas e mestiças é um fator importante, pois uma alta incidência de ROH pode indicar uma diminuição da heterose (PERIPOLLI *et al.*, 2020).

As populações com alta pressão de seleção, como ANG, NEL50, HOL e JER, apresentaram maiores números de ROHs. É bem conhecido que a seleção aumenta a autozigosidade, embora ainda haja uma falta de informações sobre os efeitos da seleção em relação à distribuição de ROH ao longo do genoma (FORUTAN *et al.*, 2018). A prevalência de ROHs é mais comum em regiões com maior desequilíbrio de ligação e baixa recombinação, ou regiões de baixa diversidade genética (CEBALLOS *et al.*, 2018). No entanto, sabe-se também que a seleção pode causar uma pressão substancial em regiões genômicas específicas, resultando em um aumento dos números de ROHs e

determinando um padrão de aparecimento na população (FLEMING *et al.*, 2018; NANDOLO *et al.*, 2018).

A diferença na densidade do painel de SNP afetou o número de ROHs detectados. Para NEL35 e NEL50 (mesma raça, mas genotipada usando diferentes painéis SNP), houve uma discrepância considerável entre a quantidade de ROH detectada (Figura 1). Conforme discutido por Hillestad *et al.* (2017), painéis de SNP mais densos tendem a resultar na identificação de um maior número de segmentos ROH. Isso acontece porque uma densidade crescente de marcadores permite a detecção de marcadores heterozigotos que não são apresentados nos painéis SNP de menor densidade. No entanto, esta pode não ser a única razão neste contexto, uma vez que o painel SNP utilizado para NEL35 foi criado especificamente para as raças *Bos taurus indicus* (GGP indicus 35K), com a seleção de SNPs com maior MAF em populações de bovinos Zebu e projetado para otimizar espaçamento equidistante de marcadores (ALILOO *et al.*, 2018). Esses fatores podem ter afetado a detecção de ROHs com base nos parâmetros usados.

Alguns estudos relataram que a densidade do painel de SNP afeta a detecção de ROH (ALILOO *et al.*, 2018; CEBALLOS; HAZELHURST; RAMSAY, 2018; PURFIELD *et al.*, 2012). Segundo Ceballos, Hazelhurst e Ramsay (2018), a precisão da detecção de ROH é afetada pela densidade de marcadores e sua distribuição ao longo do genoma. Rebelato e Caetano (2018) relataram que ROHs são mais bem identificadas em painéis SNP com mais de 50.000 SNPs. No presente estudo, o painel SNP de 35K foi menos apropriado para detectar ROHs em comparação com os painéis SNP de maior densidade. Uma possível alternativa seria a imputação a painéis de SNP de maior densidade, uma vez que a acurácia desse procedimento pode ser alta em animais Nelore (ALILOO *et al.*, 2018).

Coefficientes de endogamia genômica e tamanho efetivo da população

Em geral, as estimativas de endogamia das quatro primeiras metodologias mostraram resultados semelhantes. Cada uma das metodologias utilizadas no presente estudo tem suas particularidades quanto à estimativa de endogamia e dependem de alguns fatores. Por exemplo, os três primeiros métodos (F_{HOM1} , F_{HOM2} , F_{GRM}) são dependentes da frequência alélica dos genótipos diferente da metodologia por união de gametas (YANG *et al.*, 2011). Essas particularidades devem ser levadas em conta na definição dos

conceitos de endogamia, por exemplo, o método UNI que leva em consideração os métodos de Wright (1922) e Malécot (1948) ou os métodos HOM que levam em consideração a redução da heterozigosidade. As metodologias HOM e ROH pesam todos os alelos igualmente, enquanto os métodos UNI e GRM dão mais peso aos alelos raros (ALEMU *et al.*, 2021).

A maioria das informações moleculares medidas na estimativa dos coeficientes de endogamia é baseada na identidade do alelo marcador e não separa diretamente as regiões que são idênticas por descendentes e aquelas que são idênticas por estado (MEUWISSEN *et al.*, 2020). A vantagem da estimativa de endogamia por ROH, além de não ser dependente da frequência alélica, é a capacidade de diferenciar endogamia recente e antiga (GHOREISHIFAR *et al.*, 2020). No entanto, o método de estimativa de ROH é altamente dependente dos parâmetros de detecção utilizados nas análises.

Correlações baixas a moderadas foram observadas entre os quatro primeiros métodos com base nas análises de ROHs (Figura 2). A correlação entre as quatro metodologias e os grupos de comprimento de ROH diminui para segmentos de ROH mais curtos, reforçando que alguns métodos de estimativa de endogamia têm menor poder de detecção da endogamia mais antiga (DOEKES *et al.*, 2019). Esses resultados corroboram com achados anteriores em estudos com bovinos, como Gurgul *et al.* (2016) que avaliaram a correlação das metodologias GRM e ROH, ou Zhang *et al.* (2015) que avaliaram a correlação das metodologias HOM e UNI com o método ROH.

Em relação ao N_e , houve redução, ao longo das gerações, para todas as populações. No entanto, a maior redução de N_e foi detectada para as populações GIR e LMS. O provável efeito de uma diminuição de N_e é a redução da diversidade genética, afetando as taxas de homozigotos e heterozigotos nessas populações. Essa redução de N_e pode ser consequência de intensas práticas de seleção e uso de tecnologias reprodutivas (PERIPOLLI *et al.*, 2017; PURFIELD *et al.*, 2012; REBELATO; CAETANO, 2018). Atenção é necessária no manejo da diversidade genética de populações com N_e menor que 100 (MEUWISSEN, 2009). A redução nos valores de N_e pode resultar em aumento do desequilíbrio de ligação e taxas de fixação de alelos, causando uma redução na variabilidade de certas regiões genômicas. Todas as populações do presente estudo apresentaram valores recentes de N_e superiores a 100, o que demonstra razoável controle da endogamia. No entanto, recomenda-se, especialmente para GIR, HFD, SGT, HOL e

JER, que as taxas de Ne sejam constantemente monitoradas para evitar perda de diversidade nas próximas gerações. EDEA et al. (2018) trabalhando com algumas das mesmas raças (ANG, CHL, HFD, HOL e JER) encontraram um padrão Ne semelhante.

Ilhas de homozigose

As ilhas de homozigose foram definidas como as regiões com incidência de ROHs em pelo menos 50% mais um indivíduo da população. Isso ocorre devido ao aumento da frequência alélica em determinadas regiões como resposta a processos adaptativos ou intensa seleção de características de alto interesse econômico (REBELATO; CAETANO, 2018b). Muitas das ilhas encontradas, estavam presentes em animais GIR e BRM. Curiosamente, GIR contribuiu para a formação de BRM (KOUFARIOTIS *et al.*, 2018), sugerindo que essas ilhas podem ter sido mantidas ao longo de gerações e foram passadas para BRM. Uma das vias metabólicas encontradas com os genes nessas ilhas é a via da melanogênese, que é responsável por determinar o padrão de coloração da pelagem em cada raça (RANDHAWA *et al.*, 2016), mantendo o equilíbrio entre a eumelanina marrom-preta e a feomelanina amarelo-avermelhada (KIM; HWANG; YOON, 2014), e está associado à termorregulação, resultando em melhor adaptação a determinadas condições ambientais (D'MELLO *et al.*, 2016; SENCZUK *et al.*, 2020).

Outra via encontrada em comum nas quatro populações que apresentaram as ilhas de homozigose foi a via de sinalização da prolactina, responsável por diversos processos biológicos. Vários genes nesta via foram previamente associados a características importantes em bovinos. Por exemplo, a família de genes *STAT* (transdutor de sinal e ativador da transcrição) é responsável pelo desenvolvimento e diferenciação das células da glândula mamária em diferentes fases da vida (OTTO *et al.*, 2020). O gene *MAPK10* (mitogen-activated protein kinase 10) está relacionado com a resposta inflamatória e imunidade das células epiteliais da glândula mamária. Segundo Silva et al. (2020), *MAPK10* é um gene candidato ao escore de células somáticas (SCS). O gene *PRLR* (receptor de prolactina), também conhecido como gene *SLICK*, foi associado à tolerância ao calor (HANSEN, 2020), pois resulta em pelos curtos e lisos que resultam em melhor adaptação a altas temperaturas (DIKMEN *et al.*, 2014).

A via de sinalização do cálcio encontrada nas ilhas em comum, foi observada nas populações ANGSIM e JER. Na raça ANGSIM, o gene relacionado a esta via é o

CACNA1E (*calcium voltage-gated channel subunit alpha1 E*). Hay e Roberts (2018) relataram *CACNA1E* como um gene candidato para crescimento e características relacionadas à carcaça. Na raça JER, o gene *DRD5* (receptor de dopamina D5) foi associado à via de sinalização do cálcio, que tem sido relacionado à regulação do consumo de ração e homeostase energética (MSALYA *et al.*, 2017).

Regiões ricas em heterozigose

Diferentemente das ROHs, espera-se que as HER ocorram em regiões sob balanço de seleção ou com alta taxa de recombinação, pois baixo desequilíbrio de ligação leva a alta diversidade de regiões. Normalmente, as HER estão concentradas em regiões genômicas associadas à resistência a doenças, onde níveis mais altos de diversidade podem ajudar a população a lidar com novos (e mutáveis) problemas potenciais de enfermidades (BISCARINI *et al.*, 2020). Apesar de estarem associados a regiões de interesse, as HER não são tão amplamente estudadas quanto os ROHs (SANTOS *et al.*, 2021).

A população que apresentou maior quantidade de HERs foi a MON (população composta), que representou mais de 47% do HER encontrado neste estudo. Este resultado era esperado, pois a população MON é uma raça composta que combina diferentes grupos biológicos, incluindo raças de *Bos taurus indicus* e *Bos taurus taurus* (PERIPOLLI *et al.*, 2020).

Também é importante ressaltar a diferença encontrada entre os grupos NEL35 e NEL50 para HER, principalmente pela diferença na densidade entre os painéis SNP. A classificação do comprimento do HER foi diferente entre os painéis SNP: o grupo NEL50 apresentou um comprimento menor do HER em comparação com o grupo NEL35. Como as mesmas regiões foram detectadas com base em ambos os painéis SNP, esse resultado sugere que muitas das longas regiões HER são compostas por pequenas HER, quebrados por regiões com SNPs homozigotos.

Apenas um gene foi encontrado na menor HER identificada nesse estudo – *KHDRBS2* (*KH RNA Binding Domain Containing, Signal Transduction Associated 2*). Este gene tem sido associado com fertilidade e desempenho reprodutivo em bovinos (DE LEÓN *et al.*, 2019). No entanto, a maioria dos genes encontrados nas ilhas de HER apresentaram alguma relação com a imunidade animal, demonstrando uma relação das

respostas biológicas aos desafios ambientais. Outro gene importante é o *IFNG* (*interferon-gamma*), que foi encontrado na ilha HER para a raça MON. *IFNG* foi previamente associada à resistência a carrapatos em bovinos taurinos e zebuínos (MARYAM *et al.*, 2012). O gene *ANXA1* (*Anexina A1*), proteína regulada por glicocorticóides, identificada para SEN, desempenha papel inibitório na síntese do ácido araquidônico, precursor de processos inflamatórios.

Essas altas concentrações de heterozigose em regiões de controle a resposta imune são um ponto interessante a ser investigado nessas populações, pois podem contribuir para a resiliência a doenças (SAMUELS *et al.*, 2016). Além disso, as regiões genômicas onde há maior variabilidade em características-chaves é um ponto importante a ser verificado em estudos que estimam a diversidade genética das populações (BISCARINI *et al.*, 2020), pois a manutenção da diversidade pode ser uma vantagem para características adaptativas e seleção (WILLIAMS *et al.*, 2016).

Embora poucos estudos tenham focado na identificação de ilhas HER, parece que o uso da abordagem consecutiva é o mais utilizado, como o caso de Biscarini *et al.* (2020), Marras *et al.* (2018) e Santos *et al.* (2021). Nesse trabalho, o uso da janela deslizante não capturou nenhuma corrida de HER. Considerando a abordagem consecutiva, as análises demonstraram um bom poder de detecção de HER. Nesse contexto, o número de homozigotos e ausentes permitidos e a máxima distância entre dois SNPs consecutivos estão diretamente relacionados à flexibilidade nos critérios para identificação de HER. Quando testados diferentes parâmetros de identificação do HER, a modificação na máxima distância e SNPs ausentes não pareceram interferir na detecção das HERs. No entanto, a quantidade de homozigotos permitida dentro de um HER é um parâmetro importante que afeta o número de HER detectados, como também observado por Biscarini *et al.* (2020). Mesmo com impacto no número de detecção de HER, é importante afirmar que a maioria das HER identificadas nesse estudo continuaram sendo classificadas como pequenas (<1MB), independentemente do número de homozigotos permitido nas análises (de 0 a 5). Portanto, se recomenda em estudos futuros que se permita um pequeno número de alelos homozigotos na detecção de HER.

Painéis SNP e comparação WGS

Atualmente, alguns estudos de WGS têm sido usados para detectar a estrutura

populacional e identificar as regiões que influenciam características economicamente importantes na pecuária (XIA *et al.*, 2021). O presente estudo não encontrou semelhanças entre painéis de SNP e resultados com WGS em nenhuma das populações avaliadas. A maioria das ROHs de longas sequências homozigóticas são, na verdade, muitas ROHs curtas distribuídas lado a lado sobre o genoma. Além disso, como nos painéis de SNP, assume-se que os loci entre dois SNPs homozigotos também estão em homozigose, a ROH tende a ser mais longa, superestimando os comprimentos de algumas ROHs. Por outro lado, o uso de painéis de SNPs de baixa densidade pode ser suficiente em alguns casos, considerando que estimar ROH usando WGS não é fácil devido aos obstáculos como os altos valores da técnica, erros ocasionais de sequenciamento, e novas mutações individuais. Os resultados de ROH com painéis SNP comerciais podem ser consistentes com análises WGS quando um haplótipo aparentemente longo está presente em certas regiões genômicas.

Uma observação interessante na comparação entre raças e painéis de SNP, que não foi identificada anteriormente com a análise apenas com painéis SNP, foi uma pequena concentração de ROH observada na análise WGS para a raça Holandesa (HOL) próxima a uma região encontrada na análise com painel de SNP na população SEN. Esta região contém o gene *SLICK*, já discutido anteriormente e originalmente encontrado em animais SEN. Estudos recentes já demonstraram que o gene *SLICK* foi introduzido em algumas populações HOL em um processo de seleção para controlar o estresse térmico (DIKMEN *et al.*, 2014).

Conforme encontrado nesse estudo, o número de regiões de ROH e HER difere entre as populações e fornece informações sobre suas diferenças no processo seletivo e na história evolutiva. Essas diferenças são esperadas uma vez que os eventos que atuaram em cada população são diferentes ou apresentam intensidades diferente (EDEA *et al.*, 2018). Algumas particularidades deste estudo devem ser levadas em consideração, como o desbalanceamento do número de animais entre as populações. Isso poderia afetar o número total de ROH e HER identificados para cada população e a comparação dos resultados obtidos. Além disso, os conjuntos de dados de WGS não foram obtidos nos mesmos animais genotipados para painéis de SNP e, portanto, a falta de regiões comuns não pode ser atribuída apenas à plataforma de genotipagem.

A comparação entre os painéis de SNP 35K e 50K na raça Nelore evidenciou

divergência na identificação e classificação de ROH e HER. Conforme relatado anteriormente (ALILLOO *et al.*, 2018; CEBALLOS; HAZELHURST; RAMSAY, 2018; PURFIELD *et al.*, 2012), o uso de pelo menos um painel de 50K na análise de ROH e HER é recomendado. A utilização de painéis de menor densidade pode subestimar o número de ROH/HER e a classificação, onde segmentos de menores comprimentos são menos identificados. Essa situação ocorre principalmente devido à maior distância entre os marcadores e menor número de marcadores distribuídos ao longo do genoma, resultando em menos informações disponíveis para as análises.

Vale ressaltar que esse trabalho traz informações substanciais sobre a diversidade genética em diferentes raças bovinas, apresentando novas regiões genômicas com ilhas de homozigose e regiões ricas em heterozigose, onde está localizado muitos genes. Um dos principais desafios do manejo da diversidade genética de populações pecuárias é saber quais regiões genômicas estão em homozigose/heterozigose, uma vez que são altamente heterogêneas em todo o genoma. Essa administração e caracterização da estrutura genética de uma população é essencial para acessar a diversidade e ajudar a entender a ação do tempo sobre raças específicas. Com o avanço das tecnologias moleculares, novos insights sobre o genoma animal podem ser acessados e as populações comparadas, para verificar quais regiões estão em alta ou baixa diversidade em cada população e, assim, gerenciar melhor as gerações futuras.

Conclusões

Os números de ROHs e HERs diferem para cada população sugerindo que diferentes eventos atuaram nas distintas populações ao longo do tempo. A população com maior número de ROH foi a SEN e para HER foi a população MON. Ilhas sobrepostas foram identificadas entre GIR e BRM sugerindo que essas regiões podem ter sido compartilhadas durante sua formação. As diferentes metodologias de estimativas de endogamia apresentaram correlação de baixa a moderada com o método ROH, principalmente com um menor comprimento de ROH, sugerindo que a endogamia antiga não foi bem capturada para essas populações. Ilhas de HER foram identificadas em regiões relacionadas à resposta imune dos animais. Por fim, na comparação entre os painéis SNP e WGS, observou-se que as longas ROH e HER identificadas nos painéis de SNP são trechos mais curtos lado a lado. Observou-se incidência de ROH em análises de

WGS para animais HOL em uma região contendo uma ilha de ROH, em análises de painéis de SNP, para animais SEN relacionados à tolerância ao calor, indicando que uma possível seleção para tal característica foi aplicada nesta população.

Referências

- ALEMU, S. W.; KADRI, N. K.; HARLAND, C.; FAUX, P.; CHARLIER, C.; CABALLERO, A.; DRUET, T. An evaluation of inbreeding measures using a whole-genome sequenced cattle pedigree. **Heredity**, v. 126, n. 3, p. 410–423, 6 mar. 2021. Disponível em: <<http://www.nature.com/articles/s41437-020-00383-9>>.
- ALILLOO, H.; MRODE, R.; OKEYO, A. M.; NI, G.; GODDARD, M. E.; GIBSON, J. P. The feasibility of using low-density marker panels for genotype imputation and genomic prediction of crossbred dairy cattle of East Africa. **Journal of Dairy Science**, v. 101, n. 10, p. 9108–9127, out. 2018.
- BAES, C. F.; MAKANJUOLA, B. O.; MIGLIOR, F.; MARRAS, G.; HOWARD, J. T.; FLEMING, A.; MALTECCA, C. Symposium review: The genomic architecture of inbreeding: How homozygosity affects health and performance. **Journal of Dairy Science**, v. 102, n. 3, p. 2807–2817, 1 mar. 2019. . Acesso em: 30 out. 2019.
- BARBATO, M.; OROZCO-TERWENGEL, P.; TAPIO, M.; BRUFORD, M. W. SNeP: a tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. **Frontiers in Genetics**, v. 6, 20 mar. 2015. Disponível em: <http://www.frontiersin.org/Livestock_Genomics/10.3389/fgene.2015.00109/abstract>. Acesso em: 2 mar. 2021.
- BISCARINI, F.; COZZI, P.; GASPA, G.; MARRAS, G. **detectRUNS: an R package to detect runs of homozygosity and heterozygosity in diploid genomes**.
- BISCARINI, F.; MASTRANGELO, S.; CATILLO, G.; SENCZUK, G.; CIAMPOLINI, R. Insights into Genetic Diversity, Runs of Homozygosity and Heterozygosity-Rich Regions in Maremmana Semi-Feral Cattle Using Pedigree and Genomic Data. **Animals**, v. 10, n. 12, p. 2285, 3 dez. 2020.
- CARDOSO, D. F.; FERNANDES JÚNIOR, G. A.; SCALEZ, D. C. B.; ALVES, A. A. C.; MAGALHÃES, A. F. B.; BRESOLIN, T.; VENTURA, R. V.; LI, C.; DE SENA OLIVEIRA, M. C.; PORTO-NETO, L. R.; CARVALHEIRO, R.; DE OLIVEIRA, H. N.; TONHATI, H.; ALBUQUERQUE, L. G. Uncovering Sub-Structure and Genomic Profiles in Across-Countries Subpopulations of Angus Cattle. **Scientific Reports**, v. 10, n. 1, p. 8770, 29 dez. 2020. Disponível em: <<https://doi.org/10.1038/s41598-020-65565-1>>. Acesso em: 12 abr. 2021.

- CEBALLOS, F. C.; HAZELHURST, S.; RAMSAY, M. Assessing runs of Homozygosity: a comparison of SNP Array and whole genome sequence low coverage data. **BMC Genomics**, v. 19, n. 1, p. 106, 30 dez. 2018. . Acesso em: 3 mar. 2021.
- CEBALLOS, F. C.; JOSHI, P. K.; CLARK, D. W.; RAMSAY, M.; WILSON, J. F. Runs of homozygosity: windows into population history and trait architecture. **Nature Reviews Genetics**, v. 19, n. 4, p. 220–234, 15 abr. 2018. . Acesso em: 29 out. 2019.
- CORBIN, L. J.; LIU, A. Y. H.; BISHOP, S. C.; WOOLLIAMS, J. A. Estimation of historical effective population size using linkage disequilibria with marker data. **Journal of Animal Breeding and Genetics**, v. 129, n. 4, p. 257–270, 2012.
- CURIK, I.; FERENČAKOVIĆ, M.; SÖLKNER, J. Inbreeding and runs of homozygosity: A possible solution to an old problem. **Livestock Science**, v. 166, n. 1, p. 26–34, ago. 2014. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1871141314003060>>. Acesso em: 26 maio. 2019.
- DE LEÓN, C.; MANRIQUE, C.; MARTÍNEZ, R.; ROCHA, J. F. Research Article Genomic association study for adaptability traits in four Colombian cattle breeds. **Genetics and Molecular Research**, v. 18, n. 3, 2019. Disponível em: <http://www.funpecrp.com.br/gmr/articles/year2019/vol18-3/pdf/gmr18373_-_genomic-association-study-adaptability-traits-four-colombian-cattle-breeds.pdf>. Acesso em: 26 abr. 2021.
- DE SOUZA FONSECA, P. A.; DOS SANTOS, F. C.; ROSSE, I. C.; VENTURA, R. V.; BRUNELLI, F. Â. T.; PENNA, V. M.; DA SILVA VERNEQUE, R.; MACHADO, M. A.; DA SILVA, M. V. G. B.; CARVALHO, M. R. S.; PEIXOTO, M. G. C. D. Retelling the recent evolution of genetic diversity for Guzerá: Inferences from LD decay, runs of homozygosity and Ne over the generations. **Livestock Science**, v. 193, p. 110–117, 1 nov. 2016.
- DIKMEN, S.; KHAN, F. A.; HUSON, H. J.; SONSTEGARD, T. S.; MOSS, J. I.; DAHL, G. E.; HANSEN, P. J. The SLICK hair locus derived from Senepol cattle confers thermotolerance to intensively managed lactating Holstein cows. **Journal of Dairy Science**, v. 97, n. 9, p. 5508–5520, set. 2014. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/24996281/>>. Acesso em: 26 abr. 2021.
- DIXIT, S. P.; SINGH, S.; GANGULY, I.; BHATIA, A. K.; SHARMA, A.; KUMAR,

- N. A.; DANG, A. K.; JAYAKUMAR, S. Genome-Wide Runs of Homozygosity Revealed Selection Signatures in *Bos indicus*. **Frontiers in Genetics**, v. 11, p. 1, 21 fev. 2020. Disponível em: <<https://www.frontiersin.org/article/10.3389/fgene.2020.00092/full>>. Acesso em: 28 abr. 2021.
- D'MELLO, S.; FINLAY, G.; BAGULEY, B.; ASKARIAN-AMIRI, M. Signaling Pathways in Melanogenesis. **International Journal of Molecular Sciences**, v. 17, n. 7, p. 1144, 15 jul. 2016. Disponível em: <[pmc/articles/PMC4964517/](https://pubmed.ncbi.nlm.nih.gov/3214517/)>. Acesso em: 25 abr. 2021.
- DOEKES, H. P.; VEERKAMP, R. F.; BIJMA, P.; DE JONG, G.; HIEMSTRA, S. J.; WINDIG, J. J. Inbreeding depression due to recent and ancient inbreeding in Dutch Holstein–Friesian dairy cattle. **Genetics Selection Evolution**, v. 51, n. 1, p. 54, 27 dez. 2019. Disponível em: <<https://doi.org/10.1186/s12711-019-0497-z>>. Acesso em: 1 nov. 2019.
- EDEA, Z.; DADI, H.; DESSIE, T.; UZZAMAN, M. R.; ROTHSCHILD, M. F.; KIM, E. S.; SONSTEGARD, T. S.; KIM, K. S. Genome-wide scan reveals divergent selection among taurine and zebu cattle populations from different regions. **Animal Genetics**, v. 49, n. 6, p. 550–563, 1 dez. 2018. . Acesso em: 7 jul. 2021.
- FELIUS, M.; KOOLMEES, P. A.; THEUNISSEN, B.; LENSTRA, J. A.; BAUMUNG, R.; MANATRINON, S.; MOMMENS, G.; HOLM, L. E.; WITHEIN, K. B.; PEDERSEN, B. v.; GRAVLUND, P.; VIINALASS, H.; KANTANEN, J.; TAPIO, I.; LI, M. H.; MOAZAMI-GOUDARZI, K.; GAUTIER, M.; LALOÃ«, D.; OULMOUDEN, A.; LEVÃ©ZIEL, H.; TABERLET, P.; HARLIZIUS, B.; SIMIANER, H.; TÃ©UBERT, H.; ERHARDT, G.; JANN, O.; WEIMANN, C.; PRINZENBERG, E. M.; MEDUGORAC, I.; MEDUGORAC, A.; FÃ¶RSTER, M.; MIX, H. M.; LOOFT, C.; KALM, E.; BRADLEY, D. G.; EDWARDS, C. J.; MACHUGH, D. E.; FREEMAN, A. R.; AJMONE MARSAN, P.; NEGRINI, R.; LONGERI, M.; CERIOTTI, G.; ZANOTTI, M.; MARLETTA, D.; CRISCIONE, A.; VALENTINI, A.; MARCHITELLI, C.; PARISET, L.; SAVARESE, M. C.; PILLA, F.; GRISLIS, Z.; MICEIKIENÄ—, I.; NIJMAN, I. J.; VAN HAERINGEN, W.; VAN DE GOOR, L.; OLSAKER, I.; GINJA, C.; GAMA, L. T.; MATEUS, J. C.; BEJA-PEREIRA, A.; FERRAND, N.; IVANOVA, Z.; POPOV, R.; AMMOSOV, I.; KISELYOVA, T.;

- MARZANOV, N.; STOJANOVIC, S.; SIMÄ• IÄ• , M.; DOVÄ• , P.; KOMPAN, D.; DUNNER, S.; RODELLAR, C.; MARTÄN-BURRIEL, I.; SÄ;NCHEZ, A.; PIEDRAFITA, J.; AZOR, P. J.; DELGADO, J. v.; MARTÄNEZ-MARTÄNEZ, A.; MOLINA, A.; RODERO, E.; DOLF, G.; WILLIAMS, J. L.; WIENER, P.; BRUFORD, M. W.; BRAY, T. C.; ALDERSON, L.; PENEDO, M. C. T. **On the breeds of cattle-
Historic and current classifications**DiversityMolecular Diversity Preservation International, 9 nov. 2011. Disponível em: <<https://www.mdpi.com/1424-2818/3/4/660/htm>>. Acesso em: 7 jul. 2021.
- FELIUS, Marleen. **Cattle breeds : an encyclopedia**. [s.l: s.n.]799 p.
- FLEMING, A.; ABDALLA, E. A.; MALTECCA, C.; BAES, C. F. Invited review: Reproductive and genomic technologies to optimize breeding strategies for genetic progress in dairy cattle. **Archives Animal Breeding**, v. 61, n. 1, p. 43–57, 23 jan. 2018. Disponível em: <<http://doisrpska.nub.rs/index.php/AGR/article/view/6227>>. Acesso em: 31 out. 2019.
- FONSECA, P. A. S.; SUÄREZ-VEGA, A.; MARRAS, G.; CÄNOVAS, Á. GALLO: An R package for genomic annotation and integration of multiple data sources in livestock for positional candidate loci. **GigaScience**, v. 9, n. 12, p. 1–9, 30 dez. 2020. . Acesso em: 2 mar. 2021.
- FORUTAN, M.; ANSARI MAHYARI, S.; BAES, C.; MELZER, N.; SCHENKEL, F. S.; SARGOLZAEI, M. Inbreeding and runs of homozygosity before and after genomic selection in North American Holstein cattle. **BMC Genomics**, v. 19, n. 1, p. 98, 27 dez. 2018. . Acesso em: 14 fev. 2019.
- GHOREISHIFAR, S. M.; MORADI-SHAHRBABA, H.; FALLAHI, M. H.; JALIL SARGHALE, A.; MORADI-SHAHRBABA, M.; ABDOLLAHI-ARPANAHI, R.; KHANSEFID, M. Genomic measures of inbreeding coefficients and genome-wide scan for runs of homozygosity islands in Iranian river buffalo, *Bubalus bubalis*. **BMC Genetics**, v. 21, n. 1, p. 16, 10 dez. 2020. Disponível em: <<https://bmcgenet.biomedcentral.com/articles/10.1186/s12863-020-0824-y>>. Acesso em: 19 abr. 2021.
- GURGUL, A.; SZMATOŁA, T.; TOPOLSKI, P.; JASIELCZUK, I.; ŻUKOWSKI, K.; BUGNO-PONIEWIERSKA, M. The use of runs of homozygosity for estimation of recent inbreeding in Holstein cattle. **Journal of Applied Genetics**, v. 57, n. 4, p. 527–

530, 23 nov. 2016. . Acesso em: 29 out. 2019.

HANSEN, P. J. Prospects for gene introgression or gene editing as a strategy for reduction of the impact of heat stress on production and reproduction in cattle.

Theriogenology, v. 154, p. 190–202, 15 set. 2020. Disponível em:

<<https://doi.org/10.1016/j.theriogenology.2020.05.010>>. Acesso em: 25 abr. 2021.

HAY, E. H.; ROBERTS, A. Genotype × prenatal and post-weaning nutritional environment interaction in a composite beef cattle breed using reaction norms and a multi-trait model. **Journal of Animal Science**, v. 96, n. 2, p. 444–453, 6 mar. 2018.

Disponível em: <[https://www.scopus.com/inward/record.uri?eid=2-s2.0-](https://www.scopus.com/inward/record.uri?eid=2-s2.0-85045472400&doi=10.1093%2Fjas%2Fskx057&partnerID=40&md5=b46806ea12be08ee258d186d19c1ac89)

85045472400&doi=10.1093%2Fjas%2Fskx057&partnerID=40&md5=b46806ea12be08ee258d186d19c1ac89

HAYES, B. J.; DAETWYLER, H. D. 1000 Bull Genomes Project to Map Simple and Complex Genetic Traits in Cattle: Applications and Outcomes. **Annual Review of Animal Biosciences**, v. 7, n. 1, p. 89–102, 15 fev. 2019. . Acesso em: 19 nov. 2020.

HILLESTAD, B.; WOOLLIAMS, J. A.; BOISON, S. A.; GROVE, H.; MEUWISSEN, T.; VÅGE, D. I.; KLEMETSDAL, G. Detection of runs of homozygosity in Norwegian Red: Density, criteria and genotyping quality control. **Acta Agriculturae Scandinavica, Section A — Animal Science**, v. 67, n. 3–4, p. 107–116, 2 out. 2017.

Disponível em:

<<https://www.tandfonline.com/doi/full/10.1080/09064702.2018.1501088>>. Acesso em: 29 out. 2019.

HOWRIGAN, D. P.; SIMONSON, M. A.; KELLER, M. C. Detecting autozygosity through runs of homozygosity: A comparison of three autozygosity detection algorithms. **BMC Genomics**, v. 12, n. 1, p. 460, 23 dez. 2011. Disponível em:

<<http://www.biomedcentral.com/1471-2164/12/460>>. Acesso em: 12 abr. 2021.

KIM, S. H.; HWANG, S. Y.; YOON, J. T. Microarray-Based Analysis of the Differential Expression of Melanin Synthesis Genes in Dark and Light-Muzzle Korean Cattle. **PLoS ONE**, v. 9, n. 5, p. e96453, 8 maio 2014. Disponível em:

<[/pmc/articles/PMC4014497/](https://doi.org/10.1371/journal.pone.0096453)>. Acesso em: 22 abr. 2021.

KOUFARIOTIS, L.; HAYES, B. J.; KELLY, M.; BURNS, B. M.; LYONS, R.; STOTHARD, P.; CHAMBERLAIN, A. J.; MOORE, S. Sequencing the mosaic genome of Brahman cattle identifies historic and recent introgression including polled.

Scientific Reports, v. 8, n. 1, p. 17761, 10 dez. 2018. Disponível em:

<www.nature.com/scientificreports>. Acesso em: 22 abr. 2021.

MACCIOTTA, N. P. P.; COLLI, L.; CESARANI, A.; AJMONE-MARSAN, P.; LOW, W. Y.; TEARLE, R.; WILLIAMS, J. L. The distribution of runs of homozygosity in the genome of river and swamp buffaloes reveals a history of adaptation, migration and crossbred events. **Genetics Selection Evolution**, v. 53, n. 1, p. 20, 27 dez. 2021.

Disponível em: <<https://doi.org/10.1186/s12711-021-00616-3>>. Acesso em: 12 abr. 2021.

MALÉCOT, G. **Les mathématiques de l'hérédité**. [s.l.] Masson, 1948. 63 p.

MARCHESE, J. A. P.; BUZANSKAS, M. E.; CANTÃO, M. E.; IBELLI, A. M. G.; PEIXOTO, J. O.; JOAQUIM, L. B.; MOREIRA, G. C. M.; GODOY, T. F.; SBARDELLA, A. P.; FIGUEIREDO, E. A. P.; COUTINHO, L. L.; MUNARI, D. P.; LEDUR, M. C. Relationship of runs of homozygosity with adaptive and production traits in a paternal broiler line. **Animal**, v. 12, n. 6, p. 1126–1134, 1 jun. 2018.

Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/29065939/>>. Acesso em: 28 abr. 2021.

MARRAS, G.; WOOD, B. J.; MAKANJUOLA, B.; MALCHIODI, F.; PEETERS, K.; AS, P. van; BAES, C. F.; BISCARINI, F.; TURKEYS, H. Characterization of runs of homozygosity and heterozygosity-rich regions in a commercial turkey (*Meleagris gallopavo*) population. Em: Proceedings of the 11th World Congress of Genetics Applied to Livestock Production, Szwaczkowski 2017., 2018, Auckland. [...]. Auckland: 2018. v. 11, p. 763–767.

MARYAM, J.; BABAR, M. E.; NADEEM, A.; HUSSAIN, T. Genetic variants in interferon gamma (IFN- γ) gene are associated with resistance against ticks in *Bos taurus* and *Bos indicus*. **Molecular Biology Reports**, v. 39, n. 4, p. 4565–4570, 30 abr. 2012.

Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/21960011/>>. Acesso em: 26 abr. 2021.

MASTRANGELO, S.; SARDINA, M. T.; TOLONE, M.; DI GERLANDO, R.; SUTERA, A. M.; FONTANESI, L.; PORTOLANO, B. Genome-wide identification of runs of homozygosity islands and associated genes in local dairy cattle breeds. **Animal**, v. 12, n. 12, p. 2480–2488, 1 dez. 2018. Disponível em:

<[/core/journals/animal/article/genomewide-identification-of-runs-of-homozygosity-](https://www.nature.com/core/journals/animal/article/genomewide-identification-of-runs-of-homozygosity-)

- islands-and-associated-genes-in-local-dairy-cattle-breeds/8DC5345A470219BAB2B35B768CC4BFCD>. Acesso em: 19 ago. 2020.
- MCQUILLAN, R.; LEUTENEGGER, A.-L.; ABDEL-RAHMAN, R.; FRANKLIN, C. S.; PERICIC, M.; BARAC-LAUC, L.; SMOLEJ-NARANCIC, N.; JANICIJEVIC, B.; POLASEK, O.; TENESA, A.; MACLEOD, A. K.; FARRINGTON, S. M.; RUDAN, P.; HAYWARD, C.; VITART, V.; RUDAN, I.; WILD, S. H.; DUNLOP, M. G.; WRIGHT, A. F.; CAMPBELL, H.; WILSON, J. F. Runs of Homozygosity in European Populations. **The American Journal of Human Genetics**, v. 83, n. 3, p. 359–372, 12 set. 2008. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S000292970800445X>>. Acesso em: 31 out. 2019.
- MEUWISSEN, T. H. Accuracy of breeding values of “unrelated” individuals predicted by dense SNP genotyping. **Genetics Selection Evolution**, v. 41, n. 1, p. 35, 11 dez. 2009. Disponível em: <www.ncbi.nlm.nih.gov/sites/entrez?db=genome.>. Acesso em: 21 abr. 2021.
- MEUWISSEN, T. H. E.; SONESSON, A. K.; GEBREGIWERGIS, G.; WOOLLIAMS, J. A. Management of Genetic Diversity in the Era of Genomics. **Frontiers in Genetics**, v. 11, p. 880, 13 ago. 2020. Disponível em: <<https://www.frontiersin.org/article/10.3389/fgene.2020.00880/full>>. Acesso em: 19 abr. 2021.
- MSALYA, G.; KIM, E. S.; LAISSER, E. L. K.; KIPANYULA, M. J.; KARIMURIBO, E. D.; KUSILUKA, L. J. M.; CHENYAMBUGA, S. W.; ROTHSCCHILD, M. F. Determination of genetic structure and signatures of selection in three strains of Tanzania Shorthorn Zebu, Boran and Friesian cattle by genome-wide SNP analyses. **PLoS ONE**, v. 12, n. 1, 2017. . Acesso em: 25 abr. 2021.
- NANDOLO, W.; UTSUNOMIYA, Y. T.; MÉSZÁROS, G.; WURZINGER, M.; KHAYADZADEH, N.; TORRECILHA, R. B. P.; MULINDWA, H. A.; GONDWE, T. N.; WALDMANN, P.; FERENČAKOVIĆ, M.; GARCIA, J. F.; ROSEN, B. D.; BICKHART, D.; VAN TASSELL, C. P.; CURIK, I.; SÖLKNER, J. Misidentification of runs of homozygosity islands in cattle caused by interference with copy number variation or large intermarker distances. **Genetics Selection Evolution**, v. 50, n. 1, p. 43, 22 dez. 2018. . Acesso em: 29 out. 2019.

- ONZIMA, R. B.; UPADHYAY, M. R.; DOEKES, H. P.; BRITO, L. F.; BOSSE, M.; KANIS, E.; GROENEN, M. A. M.; CROOIJMANS, R. P. M. A. Genome-wide characterization of selection signatures and runs of homozygosity in Ugandan goat breeds. **Frontiers in Genetics**, v. 9, n. AUG, p. 318, 14 ago. 2018. Disponível em: <www.frontiersin.org>. Acesso em: 24 jun. 2021.
- OTTO, P. I.; GUIMARÃES, S. E. F.; CALUS, M. P. L.; VANDENPLAS, J.; MACHADO, M. A.; PANETTO, J. C. C.; DA SILVA, M. V. G. B. Single-step genome-wide association studies (GWAS) and post-GWAS analyses to identify genomic regions and candidate genes for milk yield in Brazilian Girolando cattle. **Journal of Dairy Science**, v. 103, n. 11, p. 10347–10360, 1 nov. 2020. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0022030220306470>>. Acesso em: 25 abr. 2021.
- PERIPOLLI, E.; METZGER, J.; DE LEMOS, M. V. A.; STAFUZZA, N. B.; KLUSKA, S.; OLIVIERI, B. F.; FEITOSA, F. L. B.; BERTON, M. P.; LOPES, F. B.; MUNARI, D. P.; LÔBO, R. B.; MAGNABOSCO, C. de U.; DI CROCE, F.; OSTERSTOCK, J.; DENISE, S.; PEREIRA, A. S. C.; BALDI, F. Autozygosity islands and ROH patterns in Nellore lineages: evidence of selection for functionally important traits. **BMC Genomics**, v. 19, n. 1, p. 680, 17 dez. 2018a.
- PERIPOLLI, E.; MUNARI, D. P.; SILVA, M. V. G. B.; LIMA, A. L. F.; IRGANG, R.; BALDI, F. Runs of homozygosity: current knowledge and applications in livestock. **Animal Genetics**, v. 48, n. 3, p. 255–271, jun. 2017. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/27910110>>. Acesso em: 29 out. 2019.
- PERIPOLLI, E.; STAFUZZA, N. B.; AMORIM, S. T.; DE LEMOS, M. V. A.; GRIGOLETTO, L.; KLUSKA, S.; FERRAZ, J. B. S.; ELER, J. P.; MATTOS, E. C.; BALDI, F. Genome-wide scan for runs of homozygosity in the composite Montana Tropical® beef cattle. **Journal of Animal Breeding and Genetics**, v. 137, n. 2, p. 155–165, 8 mar. 2020. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1111/jbg.12428>>. Acesso em: 19 ago. 2020.
- PERIPOLLI, E.; STAFUZZA, N. B.; MUNARI, D. P.; LIMA, A. L. F.; IRGANG, R.; MACHADO, M. A.; PANETTO, J. C. do C.; VENTURA, R. V.; BALDI, F.; DA SILVA, M. V. G. B. Assessment of runs of homozygosity islands and estimates of genomic inbreeding in Gyr (*Bos indicus*) dairy cattle. **BMC Genomics**, v. 19, n. 1, p.

34, 9 dez. 2018b. Disponível em:

<<https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-017-4365-3>>.

Acesso em: 19 ago. 2020.

PITT, D.; SEVANE, N.; NICOLAZZI, E. L.; MACHUGH, D. E.; PARK, S. D. E.; COLLI, L.; MARTINEZ, R.; BRUFORD, M. W.; OROZCO-TERWENGEL, P.

Domestication of cattle: Two or three events? **Evolutionary Applications**, v. 12, n. 1, p. 123–136, 23 jan. 2019. Disponível em:

<<https://onlinelibrary.wiley.com/doi/10.1111/eva.12674>>. Acesso em: 7 jul. 2021.

PURCELL, S.; NEALE, B.; TODD-BROWN, K.; THOMAS, L.; FERREIRA, M. A. R.; BENDER, D.; MALLER, J.; SKLAR, P.; DE BAKKER, P. I. W.; DALY, M. J.; SHAM, P. C. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. **The American Journal of Human Genetics**, v. 81, n. 3, p. 559–575, set. 2007. . Acesso em: 2 mar. 2019.

PURFIELD, D. C.; BERRY, D. P.; MCPARLAND, S.; BRADLEY, D. G. Runs of homozygosity and population history in cattle. **BMC Genetics**, v. 13, n. 1, p. 70, 14 dez. 2012. . Acesso em: 17 fev. 2019.

RANDHAWA, I. A. S.; KHATKAR, M. S.; THOMSON, P. C.; RAADSMA, H. W. A Meta-Assembly of Selection Signatures in Cattle. **PLOS ONE**, v. 11, n. 4, p. e0153013, 5 abr. 2016. Disponível em:

<<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0153013>>. Acesso em: 22 abr. 2021.

REBELATO, A. B.; CAETANO, A. R. Runs of homozygosity for autozygosity estimation and genomic analysis in production animals. **Pesquisa Agropecuária Brasileira**, v. 53, n. 9, p. 975–984, set. 2018. . Acesso em: 24 fev. 2019.

ROSEN, B. D.; BICKHART, D. M.; SCHNABEL, R. D.; KOREN, S.; ELSIK, C. G.; TSENG, E.; ROWAN, T. N.; LOW, W. Y.; ZIMIN, A.; COULDREY, C.; HALL, R.; LI, W.; RHIE, A.; GHURYE, J.; MCKAY, S. D.; THIBAUD-NISSEN, F.; HOFFMAN, J.; MURDOCH, B. M.; SNELLING, W. M.; MCDANELD, T. G.; HAMMOND, J. A.; SCHWARTZ, J. C.; NANDOLO, W.; HAGEN, D. E.; DREISCHER, C.; SCHULTHEISS, S. J.; SCHROEDER, S. G.; PHILLIPPY, A. M.; COLE, J. B.; VAN TASSELL, C. P.; LIU, G.; SMITH, T. P. L.; MEDRANO, J. F. De novo assembly of the cattle reference genome with single-molecule sequencing. **GigaScience**, v. 9, n. 3, p.

1–9, 2020.

SAMUELS, D. C.; WANG, J.; YE, F.; HE, J.; LEVINSON, R. T.; SHENG, Q.; ZHAO, S.; CAPRA, J. A.; SHYR, Y.; ZHENG, W.; GUO, Y. Heterozygosity Ratio, a Robust Global Genomic Measure of Autozygosity and Its Association with Height and Disease Risk. **Genetics**, v. 204, n. 3, p. 893–904, 1 nov. 2016. . Acesso em: 1 dez. 2020.

SANTOS, W.; SCHETTINI, G.; FONSECA, M. G.; PEREIRA, G. L.; CHARDULO, L. A.; NETO, O.; BALDASSINI, W. A.; OLIVEIRA, H.; CURI, R. Fine-scale estimation of inbreeding rates, runs of homozygosity and genome-wide heterozygosity levels in the Mangalarga Marchador horse breed. **Journal of Animal Breeding and Genetics**, v. 138, n. 2, p. 161–173, 19 mar. 2021. . Acesso em: 26 abr. 2021.

SAS INSTITUTE INC. **SAS 9.1.3 Help and Documentation**. Cary: ADABAS, 2013.

SEMPÉRÉ, G.; MOAZAMI-GOUDARZI, K.; EGGEN, A.; LALOË, D.; GAUTIER, M.; FLORI, L. WIDDE: A Web-Interfaced next generation database for genetic diversity exploration, with a first application in cattle. **BMC Genomics**, v. 16, n. 1, p. 1–8, 2015. Disponível em: <<http://dx.doi.org/10.1186/s12864-015-2181-1>>.

SENCZUK, G.; GUERRA, L.; MASTRANGELO, S.; CAMPOBASSO, C.; ZOUBEYDA, K.; IMANE, M.; MARLETTA, D.; KUSZA, S.; KARSLI, T.; GAOUAR, S.; PILLA, F.; CIANI, E. Fifteen Shades of Grey: Combined Analysis of Genome-Wide SNP Data in Steppe and Mediterranean Grey Cattle Sheds New Light on the Molecular Basis of Coat Color. **Genes**, v. 11, n. 8, p. 932, 13 ago. 2020. Disponível em: <<https://www.mdpi.com/2073-4425/11/8/932>>. Acesso em: 25 abr. 2021.

SIEVERT, C. **Interactive web-based data visualization with R, plotly, and shiny**. [s.l.: s.n.]

SILVA, A. A.; SILVA, D. A.; SILVA, F. F.; COSTA, C. N.; SILVA, H. T.; LOPES, P. S.; VERONEZE, R.; THOMPSON, G.; CARVALHEIRA, J. GWAS and gene networks for milk-related traits from test-day multiple lactations in Portuguese Holstein cattle.

Journal of Applied Genetics, v. 61, n. 3, p. 465–476, 1 set. 2020. Disponível em: <<http://link.springer.com/10.1007/s13353-020-00567-3>>. Acesso em: 25 abr. 2021.

SZMATOŁA, T.; GURGUL, A.; ROPKA-MOLIK, K.; JASIELCZUK, I.; ZABEK, T.; BUGNO-PONIEWIERSKA, M. Characteristics of runs of homozygosity in selected cattle breeds maintained in Poland. **Livestock Science**, v. 188, p. 72–80, 1 jun. 2016.

Disponível em:

- <<https://www.sciencedirect.com/science/article/pii/S1871141316300671>>. Acesso em: 24 jul. 2019.
- VANRADEN, P. M. Efficient methods to compute genomic predictions. **Journal of Dairy Science**, v. 91, n. 11, p. 4414–4423, 1 nov. 2008. . Acesso em: 4 jul. 2021.
- VENTURA, R. V.; BRITO, L. F.; OLIVEIRA, G. A.; DAETWYLER, H. D.; SCHENKEL, F. S.; SARGOLZAEI, M.; VANDERVOORT, G.; FONSECA E SILVA, F.; MILLER, S. P.; CARVALHO, M. E.; SANTANA, M. H. A.; MATTOS, E. C.; FONSECA, P.; ELER, J. P.; FERRAZ, J. B. S. A comprehensive comparison of high-density SNP panels and an alternative ultra-high-density panel for genomic analyses in Nellore cattle. **Animal Production Science**, v. 60, n. 3, p. 333, 5 fev. 2020.
- WANG, J.; LIAO, Y.; JAEHNIG, E.; SHI, Z.; SHENG, Q. **Gene Set Analysis Toolkit WebGestaltR**. [s.l: s.n.]. Disponível em: <<https://github.com/bzhanglab/WebGestaltR>>. Acesso em: 3 mar. 2021.
- WELDENEGODGUAD, M.; POPOV, R.; POKHAREL, K.; AMMOSOV, I.; MING, Y.; IVANOVA, Z.; KANTANEN, J. Whole-Genome Sequencing of Three Native Cattle Breeds Originating From the Northernmost Cattle Farming Regions. **Frontiers in Genetics**, v. 9, n. JAN, p. 728, 11 jan. 2019. Disponível em: <<https://www.frontiersin.org/article/10.3389/fgene.2018.00728/full>>. Acesso em: 1 dez. 2020.
- WILLIAMS, J. L.; HALL, S. J. G.; DEL CORVO, M.; BALLINGALL, K. T.; COLLI, L.; AJMONE MARSAN, P.; BISCARINI, F. Inbreeding and purging at the genomic Level: the Chillingham cattle reveal extensive, non-random SNP heterozygosity. **Animal Genetics**, v. 47, n. 1, p. 19–27, 1 fev. 2016. . Acesso em: 26 abr. 2021.
- WRIGHT, S. Coefficients of Inbreeding and Relationship. **The American Naturalist**, v. 56, n. 645, p. 330–338, jul. 1922. Disponível em: <<https://www.jstor.org/stable/pdf/2456273.pdf>>. Acesso em: 21 abr. 2021.
- XIA, X.; ZHANG, S.; ZHANG, H.; ZHANG, Z.; CHEN, N.; LI, Z.; SUN, H.; LIU, X.; LYU, S.; WANG, X.; LI, Z.; YANG, P.; XU, J.; DING, X.; SHI, Q.; WANG, E.; RU, B.; XU, Z.; LEI, C.; CHEN, H.; HUANG, Y. Assessing genomic diversity and signatures of selection in Jiaxian Red cattle using whole-genome sequencing data. **BMC Genomics**, v. 22, n. 1, p. 43, 9 dez. 2021. Disponível em: <<https://bmcbgenomics.biomedcentral.com/articles/10.1186/s12864-020-07340-0>>.

Acesso em: 26 abr. 2021.

YANG, J.; BENYAMIN, B.; MCEVOY, B. P.; GORDON, S.; HENDERS, A. K.; NYHOLT, D. R.; MADDEN, P. A.; HEATH, A. C.; MARTIN, N. G.; MONTGOMERY, G. W.; GODDARD, M. E.; VISSCHER, P. M. Common SNPs explain a large proportion of the heritability for human height. **Nature Genetics**, v. 42, n. 7, p. 565–569, jul. 2010. . Acesso em: 4 jul. 2021.

YANG, J.; LEE, S. H.; GODDARD, M. E.; VISSCHER, P. M. GCTA: A Tool for Genome-wide Complex Trait Analysis. **The American Journal of Human Genetics**, v. 88, n. 1, p. 76–82, jan. 2011. Disponível em: <<http://dx.doi.org/10.1016/j.ajhg.2010.11.011>>.

ZANELLA, R.; LAGO, L. V.; DA SILVA, A. N.; PÉRTILLE, F.; DE CARVALHO, N. S.; PANETTO, J. C. do C.; ZANELLA, G. C.; FACIOLI, F. L.; MARCOS, M. V. Genetic characterization of indubrasil cattle breed population. **Veterinary Sciences**, v. 5, n. 4, p. 98, 3 dez. 2018.

ZAVAREZ, L. B.; UTSUNOMIYA, Y. T.; CARMO, A. S.; NEVES, H. H. R.; CARVALHEIRO, R.; FERENCAKOVIC, M.; PÉREZ O'BRIEN, A. M.; CURIK, I.; COLE, J. B.; VAN TASSELL, C. P.; DA SILVA, M. V. G. B.; SONSTEGARD, T. S.; SÖLKNER, J.; GARCIA, J. F. Assessment of autozygosity in Nellore cows (*Bos indicus*) through high-density SNP genotypes. **Frontiers in Genetics**, v. 5, n. JAN, p. 5, 29 jan. 2015. Disponível em: <<http://journal.frontiersin.org/article/10.3389/fgene.2015.00005/abstract>>. Acesso em: 19 ago. 2020.

ZHANG, Q.; CALUS, M. P.; GULDBRANDTSEN, B.; LUND, M. S.; SAHANA, G. Estimation of inbreeding using pedigree, 50k SNP chip genotypes and full sequence data in three cattle breeds. **BMC Genetics**, v. 16, n. 1, p. 88, 22 dez. 2015. Disponível em: <<http://www.r-project.org/>>. Acesso em: 20 abr. 2021.

CAPÍTULO 2

CARACTERIZAÇÃO GENÉTICA E GENÔMICA DE UMA NOVA RAÇA COMPOSTA DE GADO DE CORTE (PURUNÃ) DESENVOLVIDA PARA PRODUÇÃO EM SISTEMA A PASTO

Artigo publicado na revista *Frontiers in Genetics* – doi: [10.3389/fgene.2022.858970](https://doi.org/10.3389/fgene.2022.858970)

Resumo

Purunã é uma raça composta de bovinos de corte, desenvolvida no sul do Brasil através do cruzamento das raças Angus, Charolês, Canchim e Caracu. O objetivo deste trabalho foi realizar a primeira caracterização genética da raça Purunã, com base tanto no pedigree quanto nas informações genômicas. Para isso, 100 animais selecionados aleatoriamente foram genotipados e 11.205 animais nascidos entre os anos 1997 a 2019 com informações de pedigree foram utilizados. As principais análises genéticas realizadas foram Análise de Componentes Principais (PCA), *Admixture*, Filogenética, estimativas de endogamia via pedigree e genômica, desequilíbrio de ligação (LD), tamanho efetivo populacional (N_e), consistência da fase gamética, corridas de homozigose (ROH), regiões ricas em heterozigose (HER) e análises funcionais das regiões de ROH e HER identificadas. Os achados indicam que Purunã é mais relacionado geneticamente com as raças Charolês, Canchim e Angus do que Caracu ou Nelore. Os níveis de endogamia mostraram-se pequenos em todas as métricas avaliadas e variaram de -0,009 a 0,029. Observou-se uma correlação baixa (-0,12 a 0,31) da endogamia baseada em pedigree em relação a todos os coeficientes de endogamia genômica avaliados. A média de LD foi de 0,031 ($\pm 0,0517$) e a consistência da fase gamética mostrou-se baixa para todos os pares raciais, variando de 0,42 a 0,27 à distância de 20 Mb. O N_e com base no pedigree e nas informações genômicas foi de 158 e 115, respectivamente. Um total de 1.839 ROHs foram encontrados e a maioria delas são de pequeno tamanho (<4Mb). Uma importante região homozigótica foi identificada no BTA5 com vias significativas relacionadas a características comportamentais (por exemplo, percepção sensorial, detecção de estímulo, resposta a estímulo extracelular e atividade do receptor olfativo), bem como genes candidatos relacionados à tolerância ao calor (por exemplo, *MYO1A*), taxa de conversão alimentar (por exemplo, *RDH5*) e reprodução (por exemplo, *AMDHD1*). Um total de 1.799 HERs foram identificados na raça Purunã com 92,3% delas classificadas no grupo de comprimento 0,5-1Mb, e 19 ilhas de HERs foram identificadas no genoma autossômico. Essas ilhas de HERs abrigam genes envolvidos em vias de crescimento, peso da carcaça (por exemplo, *SDCBP*), carne e qualidade da carcaça (por exemplo, *MT2A*) e deposição de marmoreio (por exemplo, *CISH*). Apesar da relação genética entre Purunã e as raças fundadoras, uma avaliação genômica multirracial não é viável devido à estrutura

populacional e baixa consistência da fase gamética entre eles.

Palavras-chave: coeficiente de endogamia, corridas de heterozigose, corridas de homozigose, diversidade genômica, persistência da fase gamética

Abstract

Purunã is a composite beef cattle breed, developed in Southern Brazil through crossing of the Angus, Charolais, Canchim, and Caracu breeds. The goal of this study was to perform the first genetic characterization of the Purunã breed, based on both pedigree and genomic information. For this, 100 randomly-selected animals were genotyped and 11,205 animals born from 1997 to 2019 had pedigree information. The main genetic analyses performed were Principal Component Analysis (PCA), Admixture, Phylogenetic Tree, pedigree and genomic inbreeding, linkage disequilibrium (LD), effective population size (N_e), consistency of gametic phase, runs of homozygosity (ROH), heterozygosity-enriched regions (HER), and functional analyses of the ROH and HER regions identified. Our findings indicate that Purunã is more genetically related to the Charolais, Canchim, and Angus breeds than Caracu or Nelore. The levels of inbreeding were shown to be small based on all the metrics evaluated and ranged from -0.009 to 0.029. A low (-0.12 to 0.31) correlation of the pedigree-based inbreeding compared to all the genomic inbreeding coefficients evaluated was observed. The LD average was 0.031 (± 0.0517) and the consistency of gametic phase was shown to be low for all the breed pairs, ranging from 0.42 to 0.27 to the distance of 20 Mb. The N_e based on pedigree and genomic information was 158 and 115, respectively. A total of 1,839 ROHs were found and the majority of them are of small length (<4Mb). An important homozygous region was identified on BTA5 with significant pathways related to behavioral traits (e.g., sensory perception, detection of stimulus, response to extracellular stimulus, and olfactory receptor activity), as well as candidate genes related to heat tolerance (e.g., MYO1A), feed conversion rate (e.g., RDH5), and reproduction (e.g., AMDHD1). A total of 1,799 HERs were identified in the Purunã breed with 92.3% of them classified within the 0.5-1Mb length group, and 19 HERs islands were identified in the autosomal genome. These HERs islands harbor genes involved in growth pathways, carcass weight (e.g., SDCBP), meat and carcass quality (e.g., MT2A), and marbling deposition (e.g., CISH). Despite the genetic relationship between Purunã and the founder breeds, a multi-breed genomic evaluation is likely not feasible due to their population structure and low consistency of gametic phase among them.

Keywords: beef cattle, genomic diversity, inbreeding coefficient, persistency of

gametic phase, runs of heterozygosity, runs of homozygosity

Introdução

A caracterização da estrutura populacional e da diversidade genética é essencial para a compreensão da base genética de adaptação ambiental e conservação dos recursos genéticos bovinos (XIA *et al.*, 2021). Essa caracterização e avaliação da diversidade precisam ser consideradas ao projetar ou atualizar programas de melhoramento e estratégias de conservação que possam ser aplicadas em populações de raças puras e mestiças.

A raça Purunã é uma população composta desenvolvida no sul do Brasil pelo cruzamento de Angus, Charolês, Canchim e Caracu, em proporções idênticas. Isso feito para melhorar as principais características de interesse e explorar a complementaridade entre as raças (OTTO *et al.*, 2021), especialmente para produção em sistemas a pasto. As pesquisas para a criação da raça Purunã começaram no Instituto Agrônomo do Paraná (IAPAR; Ponta Grossa, Paraná, Brasil) no início da década de 1980, quando pesquisadores do IAPAR estimaram a heterose nas progênes mestiças de Charolês x Caracu e Angus x Canchim (PEROTTO *et al.*, 2000; PEROTTO; MOLETTA; CUBAS, 2000). Quase 15 anos depois, foram obtidos os primeiros resultados desse experimento, onde a heterose mantida desses cruzamentos resultou em maior peso de carcaça quente, rendimento de carcaça quente, área de olho de lombo, melhor conformação de carcaça Charolês x Caracu (PEROTTO; MOLETTA; CUBAS, 2000), e maior ganho de peso médio diário em diferentes idades com os cruzamentos de Angus x Canchim (PEROTTO *et al.*, 2000).

Com base nos primeiros resultados, os pesquisadores do IAPAR realizaram um segundo acasalamento para gerar outro conjunto de animais usando as progênes resultantes da população F1 anterior. O objetivo naquele momento era reunir todas as características favoráveis em um composto que apresentasse um peso elevado e produzisse uma carcaça de alta qualidade. A hipótese de uso das raças citadas acima foi captar uma contribuição particular de cada raça para criar uma população composta com maior desempenho produtivo e adaptada às regiões tropicais e subtropicais do Brasil. A raça Angus forneceu características relacionadas à precocidade, temperamento mais dócil e alta qualidade da carne (CRISTIANA; MIRELA, 2018; TAYE *et al.*, 2018); Charolês proporcionou maior ganho de peso e rendimento de carcaça (JAHUEY-MARTÍNEZ *et al.*, 2019) e, finalmente, Caracu e Canchim contribuíram com rusticidade, tolerância ao

calor e resistência a parasitas (PIRES *et al.*, 2021; URBINATI *et al.*, 2016). Esses animais são muito bem adaptados às condições ambientais tropicais e apresentam bom potencial para ganho de peso (ITO *et al.*, 2010). A caracterização do Purunã, definida pela Associação Brasileira de Criadores de Gado Purunã (Ponta Grossa, Paraná, Brasil), é que os animais devem apresentar pelagem curta com aspecto brilhante, admitindo variação na cor da pelagem (vermelho, branco, preto e baio), tamanho médio a grande e boa distribuição muscular. Além disso, espera-se que os animais sejam dóceis e prolíficos, com precocidade sexual e rápido acabamento de carcaça.

Figura 1 - Animais Purunã do Instituto Agronômico do Paraná (IAPAR, Ponta Grossa, Paraná, Brasil)



Alguns estudos no Brasil avaliaram o desempenho da raça Purunã para características de carcaça (ITO *et al.*, 2010), produção de carne e qualidade (MISSIO *et al.*, 2015), crescimento (MOURA *et al.*, 2014) e peso em diferentes estágios de vida (OTTO *et al.*, 2021). Esses estudos indicam animais abatidos aos 24 meses de idade, com peso médio de 460 kg, espessura de gordura próxima a 3 mm e concentração substancial de ácidos graxos na carne. Além disso, estimativas de parâmetros genéticos para características de crescimento demonstraram estimativas de herdabilidade (0,05-0,21)

para pesos corporais medidos em diferentes idades (OTTO *et al.*, 2021). No entanto, nenhuma pesquisa anterior até o momento avaliou a diversidade genética e a estrutura populacional da raça Purunã. Apesar da premissa de que o cruzamento aumenta a variabilidade genética de uma população, seu histórico de desenvolvimento e manejo populacional ao longo das gerações pode ter impactado na diversidade genética da população formada (PERIPOLLI *et al.*, 2020). Os estudos de diversidade genética desempenham um papel importante na constituição de um programa de cruzamento, pois a forma como a variabilidade é controlada pode interferir na heterose produzida e impactar no vigor híbrido esperado.

Os estudos de diversidade genética são cruciais na fase inicial, chamada de pré-melhoramento, na qual é possível regenerar, caracterizar, explorar e promover a conservação da variabilidade da população (PONTES *et al.*, 2020). Os parâmetros estimados incluem coeficientes de endogamia individuais, relação genética entre os animais e níveis de regiões em homozigose e heterozigose no genoma, bem como sua distribuição ao longo dos cromossomos (BISCARINI *et al.*, 2020). Além disso, o desequilíbrio de ligação precisa ser estimado para melhor implementar a seleção genômica e para identificar segmentos conservados do genoma entre as raças (LARMER; SARGOLZAEI; SCHENKEL, 2014). Todas essas métricas contribuem para um melhor entendimento dos eventos genéticos que ocorrem na população, do impacto de decisões tomadas no passado e das estratégias que serão tomadas no futuro. Nosso objetivo com este estudo foi caracterizar a diversidade genética e genômica e a estrutura populacional de uma nova raça composta de bovinos de corte – Purunã, com base em informações genômicas e de pedigree.

Material e métodos

Cem animais da raça Purunã foram amostrados aleatoriamente e genotipados usando o painel GGP Bovine 100K (NEOGEN) contendo mais de 100.000 polimorfismos de nucleotídeo único (SNPs). O material genético foi cedido pelo Instituto Agrônomo do Paraná (IAPAR, Ponta Grossa, Paraná, Brasil). Para o controle de qualidade do genótipo (QC), apenas os cromossomos autossômicos foram retidos e um QC foi feito separadamente para cada análise. Para as corridas de homozigose (ROH) e regiões ricas em heterozigose (HER), removeu-se SNPs com call-rate a 0,90,

posição duplicada, não autossomais ou sem posição conhecida (BISCARINI *et al.*, 2020; FERENČAKOVIĆ; SÖLKNER; CURIK, 2013). Para as demais análises, os parâmetros de frequência do alelo menor (MAF <0,05) e afastamento extremo do equilíbrio de Hardy-Weinberg (HWE <10⁻⁶) também foram usados para filtrar os SNPs.

Para o banco de pedigree foram consideradas informações de 11.205 animais nascidos entre 1997 e 2019, sendo 5.224 machos e 5.981 fêmeas, sendo a população base formada por 3.999 animais. Esses dados foram usados para criar o banco de dados de pedigree, incluindo informações sobre animais, pai, mãe, sexo e data de nascimento.

Estratificação da população

Análise de componentes principais

Para avaliar as semelhanças entre a raça Purunã e as raças Angus, Canchim, Charolês e Nelore, foi realizada uma análise de componentes principais (PCA) usando o software PLINK v1.9 (PURCELL *et al.*, 2007). Os genótipos da raça Angus, Charolês e Canchim foram fornecidos pela plataforma de dados WIDDE (SEMPÉRÉ *et al.*, 2015), e os genótipos da raça Nelore foram fornecidos pela Katayama Agropecuária Ltda. Os PCAs foram estimados com base na variância padronizada da matriz de relacionamento genômico (**G**) onde a covariância de cada SNP foi dividida pela respectiva variância, utilizando apenas os SNPs em comum para todas as raças (após o QC), conforme a seguinte equação proposta por VanRaden (2008):

$$\mathbf{G} = \frac{(\mathbf{M} - 2\mathbf{P})(\mathbf{M} - 2\mathbf{P})'}{2\sum p_i (1 - p_i)}$$

onde **M** é uma matriz de contagens do alelo A, p_i é a frequência do alelo A do i^{th} SNP e **P** é uma matriz com cada linha contendo os valores de p_i .

Análise de Admixture

A análise de admixture foi realizada por meio do software ADMIXTURE (ALEXANDER *et al.*, 2015) para avaliar a história evolutiva entre a raça Purunã e suas raças fundadoras (Angus, Charolês, Canchim e Nelore). Esta análise estima a

ancestralidade através probabilidade máxima em um modelo paramétrico como (ALEXANDER; LANGE, 2011):

$$\mathcal{L}(\mathbf{Q}, \mathbf{F}) = \sum_{ij} \{n_{ij} | np_{ij} + (2 - n_{ij}) | n(1 - p_{ij})\}$$

onde, o p_{ij} é a probabilidade de sucesso na distribuição binomial $n_{ij} \sim \text{Bin}(2, p_{ij})$ dependendo da fração q_{ik} da ancestralidade de i atribuível à população k e da frequência f_{kj} do alelo 1 na população k . As matrizes $\mathbf{Q} = (q_{ik})$ e $\mathbf{F} = (f_{kj})$.

O sucesso da análise depende da escolha correta de K , que representa o número de populações ancestrais. Foram avaliados K igual a 1 até 20, mas apenas $K=2$ e 3 foram escolhidos para serem mostrados aqui, que têm uma maior interpretação biológica e $K=3$ o menor erro de validação cruzada. O pacote “pong” (BEHR *et al.*, 2016) foi usado para agrupar os resultados e visualizar a estrutura da população.

Árvore de filogenética

Para estimar a distância entre as populações, foi utilizado o software hapFLK (FARIELLO *et al.*, 2013) com base na abordagem descrita por Bonhomme *et al.* (2010). A árvore de junção foi construída a partir das distâncias genéticas de Reynolds, Weir e Cockerham (1983) entre pares de populações. A distância de Reynold foi estimada usando o coeficiente de co-ancestralidade, onde esse coeficiente é a probabilidade de que um par aleatório de genes no mesmo locus dentro de uma população escolhida aleatoriamente seja idêntico por descendência, fornecendo uma medida natural da deriva genética. Assumindo que a frequência do alelo é igual a:

$$\hat{p}_0 = \frac{1' \mathbf{n} \mathcal{F}^{-1} \mathbf{p}}{1' \mathbf{n} \mathcal{F}^{-1} \mathbf{1}_n}$$

Onde, \mathbf{p} é a frequência, \mathcal{F} é a matriz de co-ancestralidade e \hat{p}_0 é a estimativa linear imparcial com variância mínima, com $1' \mathbf{n}$ denotando a matriz de n -vetor é feita de 1's.

Estrutura da população

Medidas de endogamia

Seis estimadores de coeficientes de endogamia foram analisados. O primeiro modelo foi baseado em informações de pedigree (F_{PED}), utilizando o software ENDOG v4.8 (GUTIÉRREZ; GOYACHE; CERVANTES, 2010), seguindo o método proposto por Meuwissen e Luo (1992) em que a média de F de uma determinada geração $t(F_t)$ foi calculado como:

$$F_t = 1 - (1 - \Delta F)^t$$

em que ΔF é a mudança na taxa de endogamia de uma geração para outra, como a seguinte equação:

$$\Delta F = \frac{(F_t - F_{t-1})}{(1 - F_{t-1})}$$

em que F_t e F_{t-1} representam as estimativas médias de endogamia para a geração atual e anterior (FALCONER; MACKAY, 1996).

O segundo método teve como base os genótipos homozigotos observados e esperados (F_{HOM1}), calculados como (PURCELL *et al.*, 2007):

$$F_{HOM1} = \frac{H_{exp} - H_{obs}}{H_{exp}}$$

onde, H_{exp} é o valor esperado (proporção) para genótipos homozigotos e H_{obs} é o valor observado para genótipos homozigotos.

O terceiro método foi baseado na variância aditiva dos genótipos (F_{GRM}), usando o seguinte modelo (VANRADEN, 2008):

$$F_{GRM} = \frac{[x_i - 2p_i]^2}{h_i - 1} \text{ no qual } h_i = 2p_i(1 - p_i)$$

onde x_i é o número de cópias do alelo de referência do i^{th} SNP, p_i é a frequência do alelo de referência na população. Semelhante ao segundo método, a metodologia F_{HOM2} foi baseada em genótipos homozigotos seguindo o modelo:

$$F_{HOM2} = 1 - \frac{x_i * (2 - x_i)}{h_i}$$

Os modelos acima são todos dependentes da frequência do alelo do genótipo e, por esse motivo, um quinto modelo foi um teste baseado na correlação entre união de gametas (F_{UNI}) usando o seguinte modelo (YANG *et al.*, 2010):

$$F_{UNI} = \frac{[x_i^2 - (1 + 2p_i) * x_i + 2p_i^2]}{h_i}$$

O último método foi baseado na soma do comprimento individual de ROH dividido pelo comprimento total do genoma autossômico (F_{ROH}) usando a seguinte equação (MCQUILLAN *et al.*, 2008):

$$F_{ROH} = \frac{\sum_{i=1}^n f(ROH_i)}{\sum_{j=1}^A h(j)}$$

onde $f(ROH_i)$ é o comprimento da ROH do indivíduo i^{th} , n é o número total de regiões genômicas homozigóticas de cada indivíduo, $h(j)$ é o comprimento do cromossomo j^{th} , e A é o número de cromossomos autossômicos ($A = 29$). Ainda assim, para cada classe de ROH (<2 Mb, 2-4 Mb, 4-8 Mb, 4-16 Mb, >16 Mb, <8 Mb e >8 Mb), as estimativas de endogamia foram obtidas dividindo a soma total de segmentos ROH pelo comprimento total do genoma autossômico bovino coberto por SNPs. Todos os coeficientes de endogamia genômica foram calculados usando o software PLINK v1.9 (PURCELL *et al.*, 2007). A opção PROC CORR do software estatístico (SAS INSTITUTE INC., 2013) foi utilizada para correlacionar as estimativas dos coeficientes de endogamia. Foi criado um mapa de calor para melhor visualização dos resultados através do pacote plotly (SIEVERT, 2020).

Desequilíbrio de ligação

O desequilíbrio de ligação (r^2) foi estimado usando o software PLINK v1.9. Para observar a diminuição de r^2 com o aumento da distância dos marcadores, utilizamos a abordagem de bins estimando a média de r^2 de cada distância de 10 kb a 100 kb a cada 10 kb, e após a distância de 100 kb a cada 100 kb até a distância de 1.000 kb (1 Mb). Em análise preliminar, foram definidos que os bins relatados neste estudo deveriam ter pelo menos 50 marcadores pareados para estimar a média de r^2 .

Tamanho efetivo da população

Duas metodologias foram usadas para estimar o tamanho efetivo da população (N_e). O primeiro método utilizava informações de pedigree por meio da seguinte equação:

$$N_e = \frac{1}{2} \Delta F, \text{ onde } \Delta F = \frac{(F_t - F_{t-1})}{(1 - F_{t-1})}$$

onde F_t e F_{t-1} é a endogamia média da prole e seus pais, respectivamente (FALCONER; MACKAY, 1996). A estimativa foi realizada por meio do software POPREP (GROENEVELD *et al.*, 2009).

O segundo método foi realizado usando informações genômicas, investigadas com o método de relação entre as variâncias LD e N_e por meio da seguinte fórmula (CORBIN *et al.*, 2012):

$$N_{e(T)} = \left(4f(c_t)^{-1} \left(E \left[r^{2|ct} \right]^{-1} - \alpha \right) \right)$$

onde N_e é o tamanho efetivo da população na $t^{\text{ésima}}$ geração, c_t é a taxa de recombinação para a distância física entre os marcadores, α é a probabilidade de ocorrência de mutação e r^2 é o valor de LD.

Consistência da fase gamética

A consistência da fase gamética (CGP) foi tomada pela raiz quadrada dos valores

de r^2 somando o sinal da medida de desequilíbrio (D), como:

$$D = p(ab) - p(a)p(b)$$

onde $p(a)$ é a frequência do haplótipo-a, $p(b)$ é a frequência do haplótipo-b e p_{ab} é a frequência do haplótipo com o alelo a no primeiro locus e o alelo b no segundo locus. O CGP foi assumido como a correlação de Person entre cada raça fundadora e Purunã usando os valores de raiz quadrada sinalizada. Para estimar o CGP, apenas os SNPs em comum (após o controle de qualidade) entre cada par de raças foram usados para estimar o CGP com base na mesma distância e bin descrito na seção LD.

Proporção de SNPs polimórficos e distribuições de SNPs por intervalo de MAF

A proporção de SNPs polimórficos, após QC, foi calculada com base no MAF. As distribuições de SNPs foram calculadas em 10 intervalos MAF de 0 até 0,5 definidos a cada 0,05 pontos no MAF.

Corridas de homozigose

O software PLINK v1.9 (PURCELL *et al.*, 2007) foi utilizado para a identificação da ROH com base nos seguintes critérios:

- Um SNP heterozigoto e um ausente foram permitidos;
- A sobreposição de janelas utilizada foi de 0,05;
- A distância entre dois SNPs consecutivos não poderia ser superior a 1.000 kb;
- O comprimento mínimo de uma ROH era de 500 kb;
- O número mínimo de SNPs consecutivos que criam uma ROH deve ser igual ou superior a 30;
- A densidade de 1 SNP usado em pelo menos 50 kb;
- Uma janela genômica deslizante foi usada com 50 SNPs.

As ROHs foram classificadas nas seguintes classes: <2 Mb, 2-4 Mb, 4-8 Mb, 4-16 Mb e >16 Mb. Uma região encontrada em 36% da população foi considerada para análise futura (análise funcional e filogenética).

Regiões ricas em Heterozigose

O pacote *detectRUNS* (BISCARINI *et al.*, 2019) foi usado para a detecção de HER seguindo o método de SNPs consecutivos. Para a análise consecutiva dos SNPs, foram considerados os seguintes parâmetros:

- Um número mínimo de 20 SNPs consecutivos constitui um HER;
- Comprimento mínimo de 500 kb;
- É permitido um mínimo de dois SNP homozigotos e um ausente; e,
- O intervalo máximo entre SNPs consecutivos não pode ser superior a 1.000 kb.

As regiões genômicas que apresentaram pelo menos 10% dos animais com HER foram incluídas nas posteriores análises funcionais e de árvore filogenética.

Análises funcionais

As regiões genômicas consideradas como ilhas de ROH e HER foram usadas para anotações genômicas. Para a anotação dos genes dessas regiões foi utilizado o pacote GALLO (FONSECA *et al.*, 2020b) para *Bos taurus* da base de dados Ensembl (www.ensembl.org/Bos_taurus/Info/Index), versão ARS-UCD1.2 (ROSEN *et al.*, 2020). Posteriormente, o pacote WebGestaltR (WANG *et al.*, 2020) foi utilizado para as análises ontologia gênica (GO) para identificar processos biológicos, funções moleculares e componentes celulares nos quais os genes candidatos posicionais estão envolvidos.

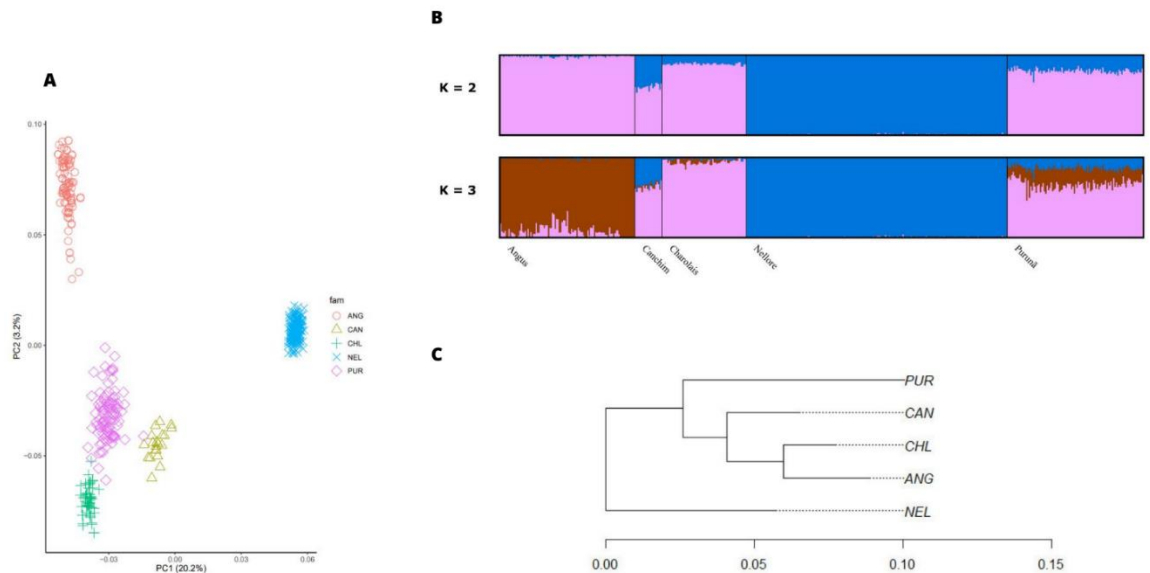
Resultados

Estratificação da população

Análise do componente principal

A PCA entre as populações de Purunã, Angus, Canchim, Charolês e Nelore é apresentado na Figura 2A. O primeiro componente principal (PC1) explicou 20,2% da variação entre as populações, enquanto o segundo componente principal (PC2) foi responsável por 3,2%. Conforme mostrado na Figura 2, os animais são agrupados dentro de raças, não havendo clara mistura entre os grupos, mesmo para populações compostas como Purunã. As raças mais próximas do Purunã são Charolês, Canchim e Angus.

Figura 2 - Estratificação Populacional da raça Purunã. **A** Análise de Componentes Principais (PCA) incluindo animais das raças Purunã, Angus, Canchim, Charolês e Nelore. **B** Análise de Admixture das Raças Purunã, Angus, Canchim, Charolês e Nelore. **C** Árvore de filogenética usando a distância de Reynolds para as populações Purunã (PUR), Angus (ANG), Canchim (CAN), Charolês (CHL) e Nelore (NEL).



Análise de admixture

Figura 2B apresenta a análise de miscigenação para as populações Purunã, Angus, Canchim, Charolês e Nelore para $K = 2$ e 3. Para $K=2$, foram observados dois grupos e a mistura entre esses indicam que duas populações fundadoras distintas (*Bos taurus taurus* e *Bos taurus indicus*) foram utilizados no desenvolvimento da raça Purunã. Em média, Angus teve 99,2% e 0,9%, Canchim 60,6% e 39,4%, Charolês 89,6% e 10,4%, Nelore 99,8% e 0,2%, e Purunã teve 80,8% e 19,2% da população ancestral 1 e 2, respectivamente. Para $K = 3$, três grupos foram observados afetando a miscigenação para a população em estudo. Este resultado ($K = 3$) indica maior contribuição do Charolês e Canchim na raça Purunã, seguindo para a raça Angus e uma pequena proporção da raça Nelore.

Árvore de filogenética

Figura 2C mostra a árvore populacional das raças Purunã (PUR), Angus (ANG),

Charolês (CHL), Canchim (CAN) e Nelore (NEL). Há uma divisão em grupos na árvore, mas a distância de um grupo para outro não é grande (0,05). Nelore aparece em uma seção separada enquanto Purunã, Canchim, Charolês e Angus estão situados em outros três nós, agrupando-se de acordo com clados de proximidade racial.

Estrutura da População

Endogamia

As médias dos coeficientes de endogamia são apresentadas na Tabela 1.

Tabela 1 - Estimativas do coeficiente de endogamia com diferentes metodologias para animais da raça Purunã

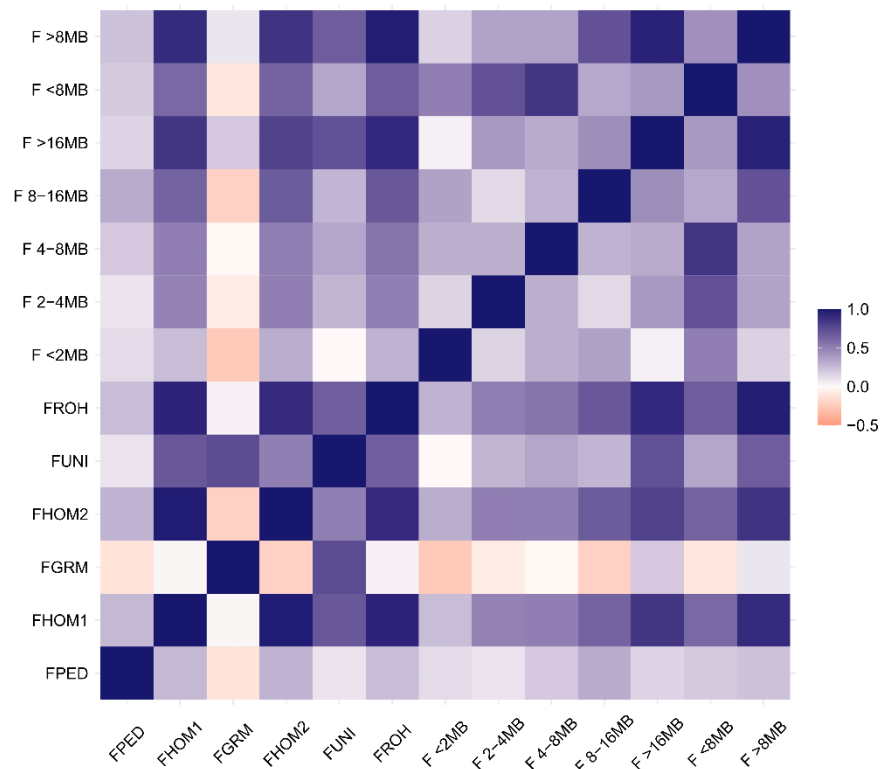
	N	média	D.P.	Mínimo	Máximo
F_{PED1}	11.205	0,002	0,019	0,000	0,375
F_{PED2}	100	0,007	0,023	0,000	0,125
F_{HOM1}	100	-0,009	0,027	-0,052	0,163
F_{GRM}	100	-0,009	0,041	-0,092	0,095
F_{HOM2}	100	-0,009	0,032	-0,060	0,171
F_{UNI}	100	-0,009	0,023	-0,052	0,133
F_{ROH}	100	0,029	0,024	0,004	0,190
F_{<2MB}	100	0,004	0,002	0,000	0,010
F_{2-4MB}	100	0,007	0,003	0,001	0,023
F_{4-8MB}	100	0,007	0,005	0,000	0,021
F_{8-16MB}	100	0,006	0,007	0,000	0,032
F_{>16MB}	100	0,005	0,016	0,000	0,127
F_{<8MB}	100	0,018	0,007	0,002	0,045
F_{>8MB}	100	0,011	0,020	0,000	0,145

N: número de indivíduos analisados; média: média do coeficiente de endogamia; D.P.: desvio padrão; F_{PED1}: Coeficiente de endogamia baseado no pedigree para todos os indivíduos da raça Purunã; F_{PED2}: Coeficiente de endogamia baseado no pedigree para indivíduos genotipados Purunã; F_{HOM1}: Coeficiente de endogamia baseado no número de genótipos homozigotos observados e esperados; F_{GRM}: Coeficiente de endogamia baseado na variância genotípica aditiva; F_{HOM2}: Coeficiente de endogamia baseado na homozigose dos genótipos; F_{UNI}: Coeficiente de endogamia baseado na correlação entre união de gametas. F_{ROH}: Coeficiente de endogamia baseado no comprimento dos ROH e no comprimento total do genoma autossômico.

A média do coeficiente de endogamia estimado com base no pedigree para todos os indivíduos Purunã (F_{PED2}) foi de 0,002. Os métodos F_{HOM1}, F_{HOM2}, F_{UNI} e F_{GRM} foram os métodos que apresentaram os menores valores de média (-0,009), enquanto a maior média do coeficiente de endogamia foi obtida por F_{ROH} (0,029). As correlações entre o

método dos coeficientes de endogamia são apresentadas na Figura 3.

Figura 3 - Correlação entre métodos de estimativa de endogamia



Fortes correlações foram encontradas entre os métodos: $F_{HOM1}-F_{HOM2}$ (0,97), $F_{HOM1}-F_{ROH}$ (0,93), $F_{HOM1}-F_{>16MB}$ (0,85), $F_{HOM1}-F_{>8MB}$ (0,89), $F_{HOM2}-F_{ROH}$ (0,90), $F_{HOM2}-F_{>16MB}$ (0,79) e $F_{HOM2}-F_{>8MB}$ (0,85). Valores baixos foram encontrados para todas as correlações entre F_{PED} e os outros métodos. O método F_{GRM} apresenta correlação muito fraca para quase todos os métodos, exceto para o F_{UNI} , o qual a correlação foi classificada como moderada (0,74). Correlações negativas foram encontradas para os métodos: $F_{GRM}-F_{HOM2}$ (-0,21), $F_{GRM}-F_{<2MB}$ (-0,24), $F_{GRM}-F_{2-4MB}$ (-0,08), $F_{GRM}-F_{4-8MB}$ (-0,01), $F_{GRM}-F_{8-16MB}$ (-0,21), $F_{GRM}-F_{<8MB}$ (-0,11) e $F_{GRM}-F_{PED}$ (-0,12).

Desequilíbrio de ligação

O LD médio variou de 0,43 a 0,04, com distância entre dois marcadores de 10 a 1.000 kb, respectivamente. A média geral de LD foi de 0,031 ($\pm 0,0517$) na distribuição

média dos marcadores 4,856 ($\pm 2,8890$) Mb.

Tamanho Efetivo da População

O tamanho efetivo da população com base no pedigree foi de 158 para a geração atual. Por outro lado, o N_e de base genômica diferiu com base na geração e no software utilizado. O software PLINK possibilitou a estimativa de N_e até a 13^a e 5^a geração, respectivamente. Para o SNeP, na 13^a geração, o N_e foi de 229, enquanto para o PLINK, o resultado para a mesma geração foi de 207. O N_e estimado para a 5^a geração no PLINK foi igual a 115.

Consistência da fase gamética

A consistência da fase gamética entre Purunã e Angus, Canchim, Charolês e Nelore são apresentados na Tabela 2.

Tabela 2 - Consistência da fase gamética baseada na correlação Person, entre a raça Purunã e suas raças fundadoras: Angus, Canchim, Charolês e Nelore

Distância (kb)	Angus	Canchim	Charolês	Nelore
20	0,40	0,42	0,43	0,27
30	0,36	0,40	0,42	0,18
40	0,30	0,36	0,41	0,17
50	0,29	0,33	0,39	0,14
60	0,29	0,33	0,39	0,13
70	0,26	0,33	0,35	0,11
80	0,24	0,30	0,33	0,08
90	0,24	0,30	0,33	0,04
100	0,20	0,27	0,31	0,04
200	0,18	0,24	0,25	0,04
300	0,12	0,20	0,18	0,03
400	0,10	0,16	0,16	0,02
500	0,10	0,13	0,12	0,02
600	0,10	0,13	0,12	0,02
700	0,09	0,11	0,11	0,01
800	0,09	0,11	0,10	0,01
900	0,08	0,11	0,10	0,00
1.000	0,07	0,09	0,10	0,00

A maior correlação, a 20 kb entre pares de SNP, entre Purunã e as demais raças foi encontrada com Charolês (0,43), seguido de Canchim (0,42), Angus (0,40) e Nelore (0,27). A distância de 10kb apresentou menor número pares de marcadores do que o limiar (<50) utilizado como critério. Portanto, esses resultados não foram apresentados.

Proporção de SNPs polimórficos e distribuição de SNPs por faixa de MAF

A proporção de SNPs polimórficos com base nas classes de MAF é apresentado na Tabela 3

Tabela 3 - Proporção de SNPs polimórficos classificados pela frequência do menor alelo (MAF)

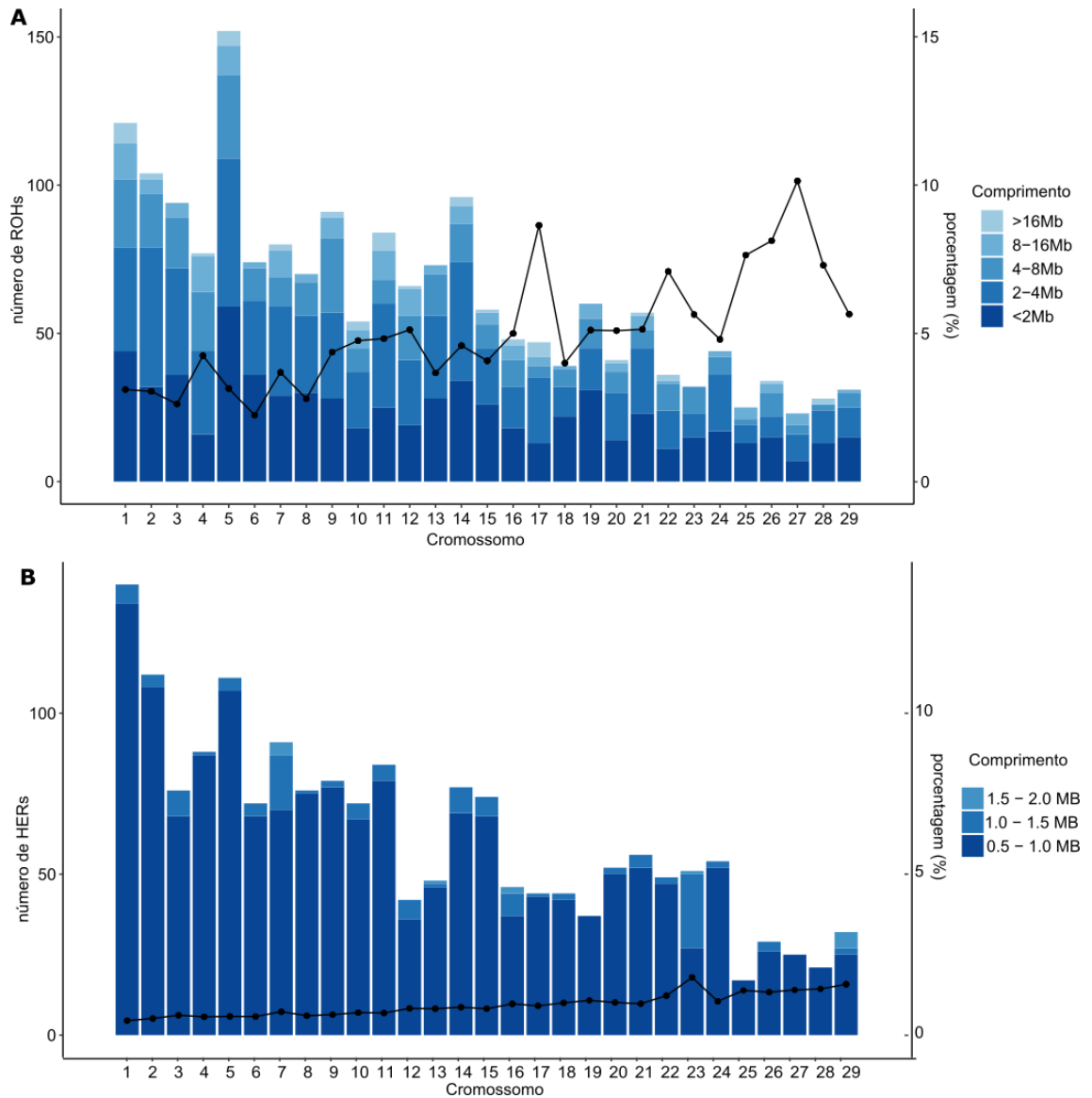
MAF	NSNP	%
0,00 – 0,05	3.232	3,67
0,05 – 0,10	2.967	3,37
0,10 – 0,20	5.792	6,59
0,20 – 0,25	7.255	8,25
0,25 – 0,30	9.314	10,59
0,30 – 0,35	11.287	12,83
0,35 – 0,40	13.283	15,10
0,40 – 0,45	14.932	16,98
0,45 – 0,50	15.652	17,80

MAF: frequência do menor alelo, NSNP: número de SNP na classe; %: porcentagem de SNP na classe.

Corridas de homozigose

Um total de 1.839 ROHs foram encontradas na raça Purunã. A distribuição de ROHs ao longo de todo o genoma autossômico e a divisão do tamanho do comprimento de ROH, podem ser observadas na Figura 4A. O comprimento de ROH observado pode ser classificado como 37,4% para < 2 Mb; 25,3% como 2-4 Mb; 17,1% como 4-8 Mb; 7,6% como 8-16 Mb; e apenas 2,6% ROHs maiores que 16 Mb. O cromossomo que apresentou maior quantidade de ROHs foi BTA5, seguido do BTA1, onde a concentração de ROHs >16Mb foi superior em relação a todos os outros cromossomos autossomais. Os cromossomos que apresentaram o menor número de ROHs foram o BTA27 e o BTA25, representando uma pequena fração de regiões de ROH.

Figura 4 - Classificação das corridas de homozigose (ROH - A) e regiões ricas em heterozigose (HER - B), por cromossomo, de acordo com o tamanho na raça Purunã e a porcentagem média dos cromossomos cobertos por ROHs/HERs.



Regiões ricas em heterozigose

No total foram encontrados 1.799 HERs na raça Purunã. A distribuição do padrão HER ao longo de todo o genoma autossômico é apresentado na Figura 4B. Cerca de 92,3% dos HERs encontrados foram classificados no comprimento de 0,5-1,0 Mb; 7,0% como 1,0-1,5 Mb; e 0,7% como 1,5-2,0 Mb. Nenhum HER maior que 2 Mb foi encontrado na população Purunã. O cromossomo que apresentou maior quantidade de HER foi o BTA1 enquanto o BTA25 apresentou o menor número de HER.

Ilhas ROH e HER e Análises Funcionais

Corridas de homozigose

Na análise de ROH, encontrou-se uma região em homozigose presente em 36% dos animais, apesar de o Purunã ser uma raça composta e recentemente desenvolvida. Essa região está localizada no BTA5 entre 54.304.681 pb e 62.031.799 pb, e tem um comprimento de 7,73 Mb, onde 131 SNPs estão presentes nessa região. Essa região é responsável pela codificação de 220 genes, sendo 181 genes codificadores de proteínas, sete pseudogenes, cinco RNAs longos não codificantes, nove microRNAs, dez RNAs miscelâneos, seis RNAs nucleolares pequenos e dois RNAs nucleares pequenos. Os termos significativos da ontologia gênica (GO) ($p < 0,05$) em que esses genes fazem parte são apresentados na Tabela 4.

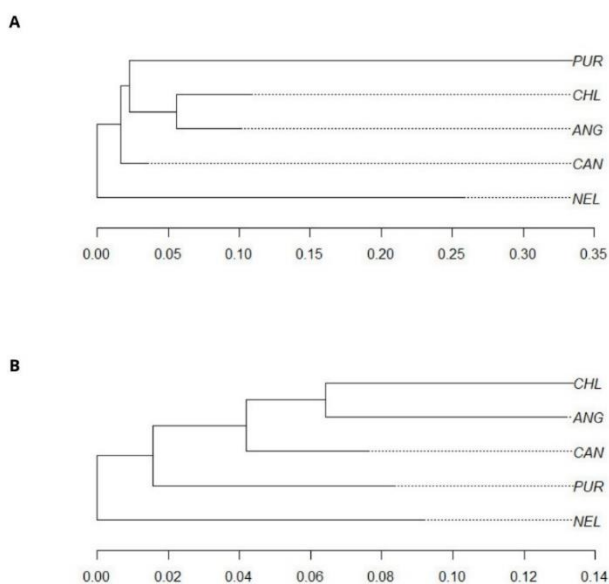
Tabela 4 - Termos significativos ($p < 0,05$) da ontologia gênica (GO) para os genes localizados em corridas de regiões de homozigose na raça Purunã

	Descrição	p-valor	genes
<i>Processo Biológico</i>			
GO:0043648	Processo metabólico do ácido dicarboxílico	0,001	<i>SHMT2; GLS2; AMDHD1 ; HAL</i> <i>MIO1A; MIP; RDH5; OR10P1;</i> <i>ENSBTAG00000047825;</i> <i>ENSBTAG00000046778;</i>
GO:0007600	Percepção sensorial	0,004	<i>ENSBTAG00000048295;</i> <i>ENSBTAG00000002913;</i> <i>OR10A7;</i> <i>ENSBTAG00000037629</i>
GO:0043473	Pigmentação	0,008	<i>DCTN2; PMEL; CD63</i>
GO:0006520	Processo metabólico de aminoácidos celulares	0,014	<i>MARTE1; SHMT2; GLS2;</i> <i>AMDHD1; HAL</i>
GO:0006091	Geração de metabólitos precursores e energia	0,015	<i>NDUFA4L2; SHMT2; PTGES3;</i> <i>CS; COQ10A; BLOC1S1</i>
GO:0044282	Processo catabólico de moléculas	0,019	<i>CYP27B1; SHMT2;</i>

	pequenas		<i>GLS2; AMDHD1; HAL</i>
			<i>OR10P1;</i>
			<i>ENSBTAG00000047825;</i>
			<i>ENSBTAG00000046778;</i>
GO:0051606	Detecção de estímulo	0,022	<i>ENSBTAG00000048295;</i>
			<i>ENSBTAG00000002913;</i>
			<i>OR10A7;</i>
			<i>ENSBTAG00000037629</i>
GO:0007422	Desenvolvimento do sistema nervoso periférico	0,023	<i>NAB2; ERBB3</i>
GO:0006766	Processo metabólico da vitamina	0,041	<i>CYP27B1; SHMT2</i>
GO:0009991	Resposta a estímulos extracelulares	0,047	<i>CYP27B1; DDIT3; MARTE1;</i>
			<i>SLC39A5</i>
<i>Função Molecular</i>			
			<i>OR10P1;</i>
			<i>ENSBTAG00000047825;</i>
			<i>ENSBTAG00000046778;</i>
GO:0004984	Atividade do receptor olfativo	0,005	<i>ENSBTAG00000048295;</i>
			<i>ENSBTAG00000002913;</i>
			<i>OR10A7;</i>
			<i>ENSBTAG00000037629</i>
GO:0016741	Atividade de transferase, transferindo grupos de um carbono	0,018	<i>EEF1AKMT3;</i>
			<i>METTL1; SHMT2; METTL7B</i>
GO:0000049	ligação tRNA	0,032	<i>METTL1; MARTE1</i>
<i>Componente Celular</i>			
GO:0009295	Nucleóide	0,017	<i>SHMT2; ATP5F1B</i>
GO:0016328	Membrana plasmática lateral	0,018	<i>MIO1A; ERBB3</i>
GO:0045177	Parte apical da célula	0,022	<i>MIO1A; MIP; ERBB3; NEDD1</i>
GO:0005759	matriz mitocondrial	0,031	<i>TSMF; SHMT2; ATP5F1B;</i>
			<i>CS; BLOC1S1</i>
GO:0098687	região cromossômica	0,034	<i>DCTN2; MBD6; NABP2; CDK2</i>
GO:0005788	Lúmen do retículo endoplasmático	0,045	<i>OS9; RDH5</i>

Dez processos biológicos, três funções moleculares e seis componentes celulares foram identificados nas vias significativas. Curiosamente, nesta região foram encontradas vias ligadas ao comportamento animal, incluindo percepção sensorial (GO:0007600), detecção de estímulo (GO:0051606), resposta a estímulo extracelular (GO:0009991), atividade do receptor olfativo (GO:0004984), e outros. Para rastrear a origem desta região em homozigose no Purunã realizamos uma análise de árvore de filogenia, utilizando apenas os SNPs alocados nesta região. A Figura 5A mostra a árvore filogenética para a região em homozigose encontrada no BTA5. Nesta região em particular, as raças Purunã, Charolês e Angus estão mais próximas em comparação com as raças Canchim e Nelore, indicando que Charolês e Angus podem ter contribuído com as principais características comportamentais observadas na raça Purunã.

Figura 5 - Árvores de filogenéticas. **A** Árvore filogenética para região em homozigose comparando as populações Purunã (PUR), Angus (ANG), Canchim (CAN), Charolês (CHL) e Nelore (NEL). **B** Árvore de filogenética da via de crescimento em regiões ricas em heterozigose



Regiões ricas em heterozigose

Para a análise de HER, as regiões identificadas em pelo menos 10% dos animais foram consideradas como ilhas de HER e utilizadas para verificar os genes e vias candidatas. A Tabela 5 apresenta as ilhas de HER encontrada na raça Purunã.

Tabela 5 - Regiões enriquecidas em heterozigotos (HER) que aparecem em pelo menos 10% dos indivíduos Purunã

CHR	%	BP1	BP2	nSNP	Comprimento
BTA1	11	26.505.838	29.555.484	22	3.049.646
BTA2	17	42.384.465	43.575.039	23	1.190.574
BTA3	10	8.435.805	9.838.978	23	1.403.173
BTA5	14	70.752.944	72.012.890	23	1.259.946
BTA5	14	75.043.240	75.983.135	26	939.895
BTA6	11	27.154.761	28.275.511	21	1.120.750
BTA7	13	8.562.310	10.432.630	24	1.870.320
BTA10	11	44.820.482	46.032.038	25	1.211.556
BTA11	14	67.243.961	69.096.131	22	1.852.170
BTA12	10	40.237.435	41,970427	25	1.732.992
BTA14	17	24.167.298	25.953.073	24	1.785.775
BTA14	16	50.608.626	51.640.291	23	1.031.665
BTA15	10	1.215.097	1.819.862	30	604.765
BTA18	10	23.515.690	24.470.198	23	954.508
BTA19	12	34.233.799	35.283.135	25	1.049.336
BTA20	10	44.099.958	45.220.153	21	1.120.195
BTA22	10	48.961.009	52.638.988	21	3.677.979
BTA23	27	26.021	1.697.122	27	1.671.101
BTA24	10	40.975.659	41.855.725	21	880.066

CHR: cromossomo; %: porcentagem da população que apresentou a ilha; BP1: posição em par de bases onde começa a HER; BP2: posição em par de bases onde termina a HER; nSNP: número de SNPs que a HER cobre; Comprimento: comprimento da HER.

Foram encontrados 19 HERs distribuídos em 17 cromossomos, sendo o BTA5 e o BTA14 os únicos a apresentarem duas HERs em cada cromossomo. A HER mais frequente (27% da população) foi encontrada na BTA23. O HER mais longa foi encontrado no BTA22 com tamanho de 3,67 Mb e o menor HER foi encontrada no BTA15 com 0,60 Mb.

Todas essas regiões são responsáveis pela codificação de 413 genes, incluindo 363 proteínas codificadoras, seis pseudogenes, 13 RNAs longos não codificantes, sete

microRNAs, dois RNAs miscelâneos, um pequeno RNA nucleolar, 15 pequenos RNAs nucleares, três pseudogenes e três RNAs ribossomais. Os termos GO significativos ($p < 0,05$) e seus genes relacionados são apresentados na Tabela 6.

Tabela 6 - Termos significativos ($p < 0,05$) da Gene Ontology (GO) para regiões enriquecidas com heterozigotos encontrados na raça Purunã

Descrição	p-valor	genes
<i>Processo Biológico</i>		
GO:0051270 Regulação do movimento do componente celular	0,003	<i>SLAMF1; RAC2; ZNF609; SDCBP; FLCN; MAP2K3; IQCF1; HYAL2; HYAL1; SEMA3B; SEMA3F; MST1; DAG1; RHOA; ELP6; PTPRM</i>
GO:0040012 Regulação da locomoção	0,004	<i>SLAMF1; RAC2; ZNF609; SDCBP; FLCN; MAP2K3; IQCF1; HYAL2; HYAL1; SEMA3B; SEMA3F; MST1; DAG1; RHOA; ELP6; PTPRM</i>
GO:0050801 homeostase iônica	0,005	<i>CASQ1; ATP1A4; ATP1A2; KCNJ10; PVALB; KCTD17; PTGDR; MT1E; MT1E; MT2A; MT3; MT4; GRM2; HYAL2; PTH1R</i>
GO:0010563 Regulação negativa do processo metabólico do fósforo	0,012	<i>PWP1; ELFN2; RTRAF; FLCN; HYAL2; INKA1; DAG1; RHOA; QARS1; PRKAR2A</i>
GO:0090407 Processo biossintético de organofosfato	0,013	<i>CD244; PIGM; LPCAT2; PRPSAP2; PEMT; FLCN; IP6K1; IMPDH2; IPGK2; TREX1; NME6</i>
GO:0072521 Processo metabólico composto contendo purina	0,018	<i>ATP1A2; PTGDR; PRPSAP2; SHMT1; FLCN; RHOA; IMPDH2; UQCRC1; TREX1; NME6; NDUFV2</i>
GO:0040007 Crescimento	0,019	<i>PPIB; SDCBP; MT2A; MT3; RAI1; FLCN; DCAF1; CISH; HYAL2; HYAL1; SEMA3B; SEMA3F; ARIH2; IP6K2</i>
GO:0007187 Via de sinalização do receptor acoplado à proteína G, acoplado ao segundo mensageiro nucleotídeo cíclico	0,019	<i>PTGDR; GNAO1; GRM2; GNAI2; GNAT1; PTH1R</i>
GO:0055086 Processo metabólico de moléculas pequenas contendo nucleobases	0,020	<i>ATP1A2; PTGDR; PRPSAP2; SHMT1; NT5M; FLCN; GMPPB; RHOA; IMPDH2; UQCRC1;</i>

			<i>TREX1; NME6; NDUFV2</i>
GO:0098657	Importar para a célula	0,020	<i>ATP1A4; ATP1A2; KCNJ9; KCNJ10; CACNG2; NCF4; CSNK1G1; SNX1; SLC6A2; MAPKAPK3; TREX1; RALBP1</i>
GO:0044283	Processo biossintético de moléculas pequenas	0,032	<i>CD244; CYP7A1; AASDHPPT; SHMT1; NT5M; HYAL1; IP6K1; MST1; IMPDH2; IPGK2; TREX1; NME6</i>
GO:0031647	Regulação da estabilidade da proteína	0,038	<i>PEX19; PPIB; MT3; COPS3; USP4; TREX1</i>
GO:1901657	Processo metabólico do composto glicosílico	0,040	<i>PTGDR; PRPSAP2; IMPDH2; NME6</i>
GO:0017144	Processo metabólico de drogas	0,041	<i>ATP1A2; MPST; PTGDR; LPCAT2; MT3; SHMT1; FLCN; AMT; RHOA; UQCRC1; TREX1; NDUFV2</i>
GO:0062012	Regulação do processo metabólico de moléculas pequenas	0,045	<i>CD244; CYP7A1; FLCN; MST1; RHOA; TREX1</i>
GO:0001505	Regulação dos níveis de neurotransmissores	0,046	<i>ATP1A2; KCNJ10; SYN3; SLC6A2; SHMT1; AMT</i>
GO:0001667	celular do tipo ameboidal	0,048	<i>MAP2K3; HYAL2; HYAL1; SEMA3B; SEMA3F; RHOA; PTPRM</i>
<i>Função Molecular</i>			
GO:0019900	ligação quinase	0,002	<i>USF1; FBXO7; LLGL1; MAP2K3; MAPKAPK3; HYAL2; GNAT1; INKA1; MST1; RHOA; QARS1; PRKAR2A; CDC25A</i>
GO:0019207	Atividade reguladora da quinase	0,003	<i>RAC2; MT3; CISH; HYAL2; INKA1; QARS1; PRKAR2A</i>
GO:0016776	Atividade da fosfotransferase, grupo fosfato como aceptor	0,006	<i>IP6K1; IPGK2; NME6</i>
GO:0008289	ligação lipídica	0,009	<i>ATP1A2; BPIFC; NCF4; SNX22; SNX1; ANXA4; MANF; AMT; ACAMPAMENTO; CATHL4; CATHL2; CATHL5; SCAP</i>
GO:0001882	Ligação de nucleosídeo	0,015	<i>IFT27; RAC2; EHD3; GNAO1; DRG2; RASD1; GNAI2; GNAT1; GMPPB; RHOA</i>
GO:0019001	Ligação do nucleotídeo guanil	0,016	<i>IFT27; RAC2; EHD3; GNAO1; DRG2; RASD1; GNAI2; GNAT1; GMPPB; RHOA</i>
GO:0005516	Ligação de calmodulina	0,035	<i>IQCF2; IQCF1; IQCF5; CAMKV</i>
GO:0004857	Atividade inibidora enzimática	0,046	<i>TIMP3; MT3; HYAL2; INKA1; QARS1; PRKAR2A; NGP</i>
<i>Componente celular</i>			
GO:0005635	Envelope nuclear	0,007	<i>RTCB; RAC2; ANXA4; GMCL1; C11H2orf42; SDCBP; ENSBTAG00000011588; APEH;</i>

			<i>SHISA5; NRADD</i>
GO:1905360	Complexo GTPase	0,012	<i>GNAO1; GNAI2; GNAT1</i>
GO:0016528	Sarcoplasma	0,024	<i>CASQ1; RASD1; MANF</i>
GO:0099503	vesícula secretora	0,031	<i>SYN3; ANXA4; MT3; IQCF1; ACAMPAMENTO; CATHL4; CATHL2; CATHL5; NGP</i>
GO:0005581	Trímero de colágeno	0,032	<i>ENSBTAG00000030683; MGC137211; MGC137014</i>
GO:0030496	Midbody	0,032	<i>FLCN; GNAI2; RHOA; KLHDC8B; MTCL1</i>

No total, foram encontrados 17 termos GO significativos envolvidos em processos biológicos, oito em funções moleculares e seis em componentes celulares. Regiões interessantes relacionadas às vias de crescimento (GO:0040007) foram encontradas em regiões ricas em heterozigoses no BTA10, BTA14, BTA18, BTA19 e BTA20, onde 14 genes estão atuando em maior variabilidade na população. Tais genes são *PPIB* (peptidilprolil isomerase B), *SDCBP* (proteína de ligação ao sindecán), componentes *MT2A* e *MT3* da metalotioneína, *RAI1* (ácido retinóico induzido 1), *FLCN* (foliculina), *DCAF1* (proteína de ligação de dano ao DNA 1), *CISH* (citocina -induzível SH2), componentes *HYAL1* e *HAYAL2* da Hialuronidase, componentes *SEMA3B* e *SEMA3F* da Semaforina, *ARIH2* (Ariadne RBR E3 Ubiquitina Proteína Ligase 2) e *IP6K2* (Inositol Hexakisfosfato Quinase 2). Para rastrear a origem dessas regiões heterozigotas em Purunã, utilizamos uma análise de árvore filogenética. A Figura 5B mostra a árvore filogenética para as regiões relacionadas à via de crescimento encontrada nas regiões ricas em heterozigose.

Discussão

Nosso principal objetivo neste estudo foi caracterizar geneticamente a raça Purunã, estimando a diversidade genética e os parâmetros de estrutura populacional com base em informações genômicas e de pedigree. Essa raça se desenvolveu pelo cruzamento, em mesma proporção, de Charolês, Canchim, Angus e Caracu, enquanto o Canchim também é uma raça composta que tem o Charolês e o Nelore como principais raças fundadoras. Portanto, a proporção genética média para Purunã é 13/32 Charolês, 8/32 Caracu, 8/32 Angus e 3/32 Nelore.

Com base nos resultados da seção de estratificação populacional, a raça Purunã parece ser geneticamente mais próxima das raças Charolês, Canchim e Angus, com

maior contribuição genômica da raça Charolês (Figura 2B, $K = 3$). Sendo assim, os resultados indicam que Purunã está mais próximo dos *Bos taurus taurus* do que das raças *Bos taurus indicus*. Isso era esperado devido à maior contribuição das raças taurinas na formação da raça Purunã.

Medidas de endogamia

A manutenção de baixos níveis de endogamia também é desejável em raças compostas, uma vez que uma vantagem do cruzamento é a heterose. Tal heterose é influenciada pela distância genética entre as raças parentais e pelo nível de endogamia na população, o que pode afetar o grau de retenção da heterose (PERIPOLLI *et al.*, 2020). Conforme mostrado na Tabela 1, o coeficiente de endogamia para todas as métricas estimadas neste estudo variou de $-0,009 (\pm 0,041)$ a $0,029 (\pm 0,024)$. Esses resultados são esperados, pois a raça Purunã é uma raça composta desenvolvida recentemente. O baixo nível de endogamia, com ênfase nos valores negativos, indica que a probabilidade de dois genes homólogos de um indivíduo serem idênticos por descendência é menor do que a probabilidade dois genes homólogos retirados da população de referência terem a ancestralidade compartilhada pela população (WANG, 2014). Ou ainda, em termos de variabilidade de ganho ou perda para uma população base de referência (VILLANUEVA *et al.*, 2021), os valores indicavam que alguma variabilidade possa ter sido adquirida (por migração ou fluxo gênico de outras populações) ou, nos casos em que o coeficiente de endogamia é positivo, uma ligeira perda de variabilidade.

Uma medida precisa de F_{PED} pode ser esperada quando se têm um completo, profundo (muitas gerações registradas) e nenhum (ou poucos) erros no arquivo de pedigree. No caso da raça Purunã, as informações nos arquivos de pedigree chegam, em média, a 2,35 gerações. Como o pedigree de Purunã é raso, o uso de informações genômicas para estimar os coeficientes de endogamia é uma ótima alternativa para acessar os níveis de endogamia dos indivíduos. Além disso, essas métricas de endogamia não dependem de informações de pedigree, levando em consideração a variação de amostragem mendeliana (DOEKES *et al.*, 2019), a natureza estocástica da recombinação (FERENČAKOVIĆ *et al.*, 2013) e corrigindo a falha do pedigree em assumir que os fundadores de uma população são independentes (REBELATO;

CAETANO, 2018). Outro ponto positivo é que algumas métricas não apenas medem os níveis de endogamia geral, mas também fornecem uma estimativa de quando a endogamia foi criada, como no caso de F_{ROH} .

O F_{ROH} captura o nível mais alto de endogamia, especialmente porque a medida F_{ROH} é capaz de capturar endogamia recente e mais antiga (GHOREISHIFAR *et al.*, 2020). Conforme mostrado na Tabela 1, o valor do coeficiente de endogamia antiga ($F_{<8MB}$) é maior do que a endogamia mais recente ($F_{>8MB}$). Essa endogamia antiga poderia ser fornecida por gerações antigas nos primeiros acasalamentos e ainda estar na população passando por gerações. Essa divisão entre endogamia antiga e recente é útil para gerenciar a diversidade da população. Como nem toda endogamia deve ser igualmente desfavorável, espera-se que a endogamia recente tenha efeitos mais negativos do que a endogamia antiga (DOEKES *et al.*, 2019), portanto, manter um baixo nível de coeficiente de endogamia recente é uma meta desejável. A Figura 3 ilustra a correlação entre as medidas do coeficiente de endogamia, todas essas apresentaram baixa correlação com F_{PED} . Alguns autores já mencionaram que as medidas de endogamia genômica são mais precisas na avaliação da endogamia individual (CURIK; FERENČAKOVIĆ; SÖLKNER, 2014; DOEKES *et al.*, 2019; MARRAS *et al.*, 2015). Isso ocorre devido às particularidades mencionadas anteriormente sobre a estimativa de pedigree, mas como o F_{PED} , cada métrica usada para calcular o coeficiente de endogamia genômica tem suas especificidades e capta um tipo diferente de endogamia que foi originalmente definido por Wright (1922) e/ou Malécot (1948).

As medidas genômicas variam de acordo com o peso que cada marcador recebe para na matriz de relacionamento (HOWARD *et al.*, 2017). Isso afeta como a endogamia é calculada para cada indivíduo e a correlação entre as medidas. As estimativas F_{HOM} e F_{ROH} pesam todos os alelos igualmente, enquanto as medidas F_{UNI} e F_{GRM} dão mais pesos aos alelos raros (ALEMU *et al.*, 2021). Isso poderia explicar o porquê F_{HOM} e F_{ROH} , e F_{GRM} e F_{UNI} apresentam correlação moderada a forte, enquanto as classes F_{GRM} e F_{HOM} ou F_{ROH} apresentaram correlação negativa.

Desequilíbrio de ligação, tamanho efetivo da população e consistência da fase gamética

Valores mais altos de LD foram observados para marcadores localizados mais

próximos uns dos outros e uma diminuição mais rápida dos valores de LD foi encontrada à medida que a distância entre os marcadores aumentava, como observado em outras populações cruzadas ou compostas (DENG *et al.*, 2019; PRIEUR *et al.*, 2017). A extensão da LD é fortemente influenciada pela história da população, principalmente em populações de animais domésticos, que sofreram gargalos durante a domesticação e a posterior formação de raças (BRITO *et al.*, 2015). Tal LD está diretamente relacionado à seleção genômica, onde o número de marcadores necessários para prever com precisão os valores genéticos depende desse LD (LARMER; SARGOLZAEI; SCHENKEL, 2014). Seguindo a equação proposta por McKay *et al.* (2007), o número de marcadores necessários para uma seleção genômica precisa será em torno de 95.000 marcadores [2,67 GB/30 kb em LD = 0,2] para a raça Purunã. No entanto, é essencial destacar que para uma implementação de seleção genômica em Purunã é crucial que uma população de treinamento considerável seja gerada, para fornecer previsões genômicas precisas de valores genéticos à seleção.

A análise de LD desempenha um papel central em muitas áreas da genética de populações, incluindo a determinação de mapas genéticos, determinação de níveis de recombinação no nível populacional e estimativa de N_e (D'AMBROSIO *et al.*, 2019). Com base em todas as métricas, as estimativas de N_e para Purunã são superiores a 100 nas gerações atuais, que é um limite proposto por Meuwissen (2009) para garantir a sustentabilidade da população a longo prazo. A estimativa de N_e baseada em LD foi capaz de ser detectada até a 5ª geração usando o software PLINK (PURCELL *et al.*, 2007). Observamos uma pequena divergência entre os resultados dos softwares SNeP e PLINK, mas não tão alta quanto relatada por Barbato *et al.* (2015).

Compreender os níveis de LD, a estrutura populacional e o CGP entre as raças é crucial para implementar a seleção genômica (BRITO *et al.*, 2015). O CGP para todas as raças avaliadas, incluindo o Purunã, resultou em baixa correlação, conforme mostrado na Tabela 2. Esses resultados indicam que a fase dos marcadores (ou a fase entre marcadores e QTL) não é consistente entre os pares raciais. Nesse contexto, a possível utilização de uma população de treinamento multirracial para avaliações genômicas usando essas raças (Purunã, Charolês, Canchim e Angus) pode não resultar em valores genéticos precisos. Como os marcadores não estão na mesma fase entre as raças, a capacidade de usar uma raça para determinar os efeitos do SNP e ajudar na

seleção em outra população torna-se menos provável (LARMER; SARGOLZAEI; SCHENKEL, 2014).

Corridas de homozigose

Figura 4A mostra o número de ROHs encontrado por cromossomo em Purunã. O BTA5 apresentou o maior número de ROHs, como também observado em outros estudos com bovinos de corte (PERIPOLLI *et al.*, 2018a, 2020). A maioria das ROH encontradas (62,7%) foram classificadas como ROHs curtas, e como o comprimento das ROH está negativamente correlacionado com a coancestralidade (MASTRANGELO *et al.*, 2018a, as ROH encontradas no Purunã foram concebidas em uma geração mais antigas. Tomando o comprimento da ROH e usando os estudos que estimam a ROH e correlacionam com a geração, como o trabalho de Howrigan *et al.* (2011), a maioria das ROH encontradas neste estudo foram geradas entre 10 e 20 gerações atrás.

A ROH pode ser usada para uma caracterização do genoma e melhor compreensão das implicações da pressão de seleção (MARRAS *et al.*, 2018). Uma região interessante foi identificada no BTA5, que contém vias significativas relacionadas a características comportamentais. A primeira via foi a via de percepção sensorial associada a uma série de eventos necessários para que um organismo receba um estímulo sensorial, converta-o em um sinal molecular e reconheça e caracterize o sinal (AmiGO 2: Term Details for “sensory perception” (GO:0007600)). A segunda foi a detecção de uma via de estímulo relacionada a um estímulo recebido por uma célula ou organismo. Essa via converte um sinal em resposta a um estímulo extracelular, associando qualquer movimento, secreção, produção enzimática ou expressão gênica em um estímulo extracelular (AmiGO 2: Term Details for “detection of stimulus” (GO:0051606)). A terceira via foi a via de atividade do receptor olfativo, uma via relacionada à combinação de um odorante e transmissão do sinal de um lado da membrana para o outro para iniciar uma mudança na atividade celular em resposta à detecção do cheiro (DEMARIA; NGAI, 2010).

Para rastrear a origem dessa região, uma árvore filogenética (Figura 5A) foi realizada para avaliar qual raça poderia ter fornecido tal região. Como apresentado, os animais Purunã, Charolês e Angus parecem ser geneticamente mais próximos e, portanto, Angus e Charolês podem ter contribuído para essa região. Alguns estudos

relataram anteriormente a mesma região com alta incidência de corridas de homozigose (FABBRI *et al.*, 2021; SZMATOŁA *et al.*, 2019). Alguns genes interessantes, já citados na literatura como genes candidatos, foram identificados nesta região, como *MYOIA* (Miosina IA) relacionado à tolerância ao calor de bovinos (JIA *et al.*, 2019), *RDH5* (11-cis retinol desidrogenase 5) associado a conversão alimentar (DE ALMEIDA SANTANA *et al.*, 2016) e *AMDHDI* (Amidohydrolase Domain Containing 1) relacionado à reprodução (MORAVČÍKOVÁ *et al.*, 2019).

Região rica em heterozigose

A manutenção da diversidade em um locus pode ser vantajosa para a aptidão e pode estar sujeita a balanço de seleção (WILLIAMS *et al.*, 2016). Essas regiões ricas em heterozigose são diferenças de nucleotídeo único observadas entre os cromossomos paternos e maternos e podem revelar muito sobre a estrutura populacional e a história demográfica da população (SANTOS *et al.*, 2021a). Conforme demonstrado na Figura 4B, a maioria das HER encontradas neste estudo foram classificadas como HER mais curta. Curiosamente, uma região já mencionada como região conservada para bovinos de corte e leite na BTA14 (ZHAO *et al.*, 2015), foi identificada em HER. Essa região é variável em pelo menos 17% dos indivíduos Purunã, demonstrando que mesmo em regiões mais conservadoras, os cruzados poderiam proporcionar alguma variabilidade aos animais.

Embora alguns estudos tenham mostrado que a maioria das ilhas de HER estão relacionados à imunidade a doenças (WILLIAMS *et al.*, 2016), taxa de sobrevivência e fertilidade (BISCARINI *et al.*, 2020), uma interessante via foi encontrada nesse estudo relacionado à crescimento. A via de crescimento é um processo biológico relacionado ao aumento de tamanho ou massa de um organismo inteiro, parte de um organismo ou uma célula (“AmiGO 2: Term Details for ‘growth’ (GO:0040007)”). Esse é o caso do gene *PPIB*, gene utilizado como gene de referência em estudos de expressão gênica (COSTA *et al.*, 2013; DA COSTA *et al.*, 2013) ou do gene *SDCBP*, um possível gene candidato relacionado ao peso de carcaça em Hanwoo, uma raça nativa coreana (LEE *et al.*, 2013) e o composto Montana Tropical, uma população composta de bovinos de corte desenvolvida no Brasil (GRIGOLETTO *et al.*, 2020). Outro gene citado como um possível gene candidato para qualidade de carne e rendimento de carcaça foi o *MT2A*,

que está envolvido na resposta a glicocorticoide e com respostas biológicas a metais e antioxidantes (HAEGEMAN *et al.*, 2003). Embora o gene *CISH* esteja diretamente relacionado ao metabolismo da insulina, Fonseca *et al.* (2020a) indicaram que este gene poderia desempenhar um papel essencial na deposição de marmoreio. Nos resultados apresentados para esta via não foi possível definir uma única raça responsável por fornecer tal região de HER para a raça Purunã, conforme mostra a Figura 5B. Isso significa que tal variabilidade não é fornecida por um único ou pequeno grupo de raças, mas pela mistura das raças utilizadas na criação da raça Purunã.

Conclusões

Conforme observado nas análises de admixture, a raça Purunã recebeu contribuição genética mais significativa para sua formação de Charolês, Canchim e Angus. Os níveis de endogamia para Purunã foram pequenos com base em múltiplos métodos de endogamia. Valores maiores de LD foram observados para marcadores com distâncias pequenas e uma diminuição mais rápida associada ao aumento da distância entre os marcadores (variando de 0,43 a 0,04 com distância de 10 a 1000 Kb), indicando que um painel mais denso de marcadores é necessário para alcançar maiores níveis de acurácia em uma seleção genômica de Purunã. Um alto N_e (>100) foi observado em todas as métricas avaliadas e a consistência de fase gamética resultou em uma pequena correlação entre todas as raças, o que determina que uma avaliação genética multirracial para Purunã pode não ser vantajosa. Uma região homozigose interessante foi encontrada no BTA5 com vias significativas relacionadas ao comportamento e genes relacionados a características como tolerância ao calor (*MYO1A*), taxa de conversão alimentar (*RDH5*) e reprodução (*AMDHD1*). Isso poderia indicar uma possível pressão de seleção nessa região. Para a heterozigose, o número de HER foi elevado, mas isso era esperado já que o Purunã é uma raça composta. Entre as regiões HER, foi identificada uma via interessante relacionada ao crescimento com maior variabilidade, contendo genes previamente associados ao peso da carcaça (*SDCBP*), carne e qualidade da carcaça (*MT2A*) e deposição de marmoreio (*CISH*).

Referências

- ALEMU, S. W.; KADRI, N. K.; HARLAND, C.; FAUX, P.; CHARLIER, C.; CABALLERO, A.; DRUET, T. An evaluation of inbreeding measures using a whole-genome sequenced cattle pedigree. **Heredity**, v. 126, n. 3, p. 410–423, 6 mar. 2021. Disponível em: <<http://www.nature.com/articles/s41437-020-00383-9>>.
- ALEXANDER, D. H.; LANGE, K. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. **BMC Bioinformatics** 2011 12:1, v. 12, n. 1, p. 1–6, 18 jun. 2011.
- ALEXANDER, D. H.; SHRINGARPURE, S. S.; NOVEMBRE, J.; LANGE, K. Admixture 1.3 Software Manual. 2015.
- AmiGO 2: Term Details for “detection of stimulus” (GO:0051606)**. Disponível em: <<http://amigo.geneontology.org/amigo/term/GO:0051606>>. Acesso em: 18 nov. 2021.
- AmiGO 2: Term Details for “growth” (GO:0040007)**. Disponível em: <<http://amigo.geneontology.org/amigo/term/GO:0040007>>. Acesso em: 24 nov. 2021.
- AmiGO 2: Term Details for “sensory perception” (GO:0007600)**. Disponível em: <<http://amigo.geneontology.org/amigo/term/GO:0007600>>. Acesso em: 18 nov. 2021.
- BARBATO, M.; OROZCO-TERWENGEL, P.; TAPIO, M.; BRUFORD, M. W. SNeP: a tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. **Frontiers in Genetics**, v. 6, 20 mar. 2015. Disponível em: <http://www.frontiersin.org/Livestock_Genomics/10.3389/fgene.2015.00109/abstract>. Acesso em: 2 mar. 2021.
- BEHR, A. A.; LIU, K. Z.; LIU-FANG, G.; NAKKA, P.; RAMACHANDRAN, S. pong: fast analysis and visualization of latent clusters in population genetic data. **Bioinformatics**, v. 32, n. 18, p. 2817–2823, 15 set. 2016.
- BISCARINI, F.; COZZI, P.; GASPA, G.; MARRAS, G. **detectRUNS: an R package to detect runs of homozygosity and heterozygosity in diploid genomes**.
- BISCARINI, F.; MASTRANGELO, S.; CATILLO, G.; SENCZUK, G.; CIAMPOLINI, R. Insights into Genetic Diversity, Runs of Homozygosity and Heterozygosity-Rich Regions in Maremmana Semi-Feral Cattle Using Pedigree and Genomic Data. **Animals**, v. 10, n. 12, p. 2285, 3 dez. 2020.
- BONHOMME, M.; CHEVALET, C.; SERVIN, B.; BOITARD, S.; ABDALLAH, J.; BLOTT, S.; SANCRISTOBAL, M. Detecting Selection in Population Trees: The

- Lewontin and Krakauer Test Extended. **Genetics**, v. 186, n. 1, p. 241–262, 1 set. 2010.
- BRITO, L. F.; JAFARIKIA, M.; GROSSI, D. A.; KIJAS, J. W.; PORTO-NETO, L. R.; VENTURA, R. V.; SALGORZAEI, M.; SCHENKEL, F. S. Characterization of linkage disequilibrium, consistency of gametic phase and admixture in Australian and Canadian goats. **BMC genetics**, v. 16, n. 1, 25 jun. 2015.
- CORBIN, L. J.; LIU, A. Y. H.; BISHOP, S. C.; WOOLLIAMS, J. A. Estimation of historical effective population size using linkage disequilibria with marker data. **Journal of Animal Breeding and Genetics**, v. 129, n. 4, p. 257–270, 2012.
- COSTA, A. S. H.; SILVA, M. P.; ALFAIA, C. P. M.; PIRES, V. M. R.; FONTES, C. M. G. A.; BESSA, R. J. B.; PRATES, J. A. M. Genetic Background and Diet Impact Beef Fatty Acid Composition and Stearoyl-CoA Desaturase mRNA Expression. **Lipids**, v. 48, n. 4, p. 369–381, 7 abr. 2013.
- CRISTIANA, A. D.; MIRELA, C. S. Study on technological aspects related to the growth of the aberdeen angus breed in Romania. **Annals Food Science and Technology**, v. 19, n. 3, p. 1–5, 2018.
- CURIK, I.; FERENČAKOVIĆ, M.; SÖLKNER, J. Inbreeding and runs of homozygosity: A possible solution to an old problem. **Livestock Science**, v. 166, n. 1, p. 26–34, ago. 2014. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1871141314003060>>. Acesso em: 26 maio. 2019.
- DA COSTA, A. S. H.; PIRES, V. M. R.; FONTES, C. M. G. A.; MESTRE PRATES, J. A. Expression of genes controlling fat deposition in two genetically diverse beef cattle breeds fed high or low silage diets. **BMC Veterinary Research**, v. 9, n. 1, p. 118, 17 dez. 2013.
- D'AMBROSIO, J.; PHOCAS, F.; HAFFRAY, P.; BESTIN, A.; BRARD-FUDULEA, S.; PONCET, C.; QUILLET, E.; DECHAMP, N.; FRASLIN, C.; CHARLES, M.; DUPONT-NIVET, M. Genome-wide estimates of genetic diversity, inbreeding and effective size of experimental and commercial rainbow trout lines undergoing selective breeding. **Genetics Selection Evolution**, v. 51, n. 1, p. 1–15, 6 jun. 2019.
- DE ALMEIDA SANTANA, M. H.; JUNIOR, G. A. O.; CESAR, A. S. M.; FREUA, M. C.; DA COSTA GOMES, R.; DA LUZ E SILVA, S.; LEME, P. R.; FUKUMASU, H.; CARVALHO, M. E.; VENTURA, R. V.; COUTINHO, L. L.; KADARMIDEEN, H. N.;

- FERRAZ, J. B. S. Copy number variations and genome-wide associations reveal putative genes and metabolic pathways involved with the feed conversion ratio in beef cattle. **Journal of Applied Genetics**, v. 57, n. 4, p. 495–504, 21 nov. 2016.
- DEMARIA, S.; NGAI, J. **The cell biology of smell** *Journal of Cell Biology* 1 nov. 2010.
- DENG, T.; LIANG, A.; LIU, J.; HUA, G.; YE, T.; LIU, S.; CAMPANILE, G.; PLASTOW, G.; ZHANG, C.; WANG, Z.; SALZANO, A.; GASPARRINI, B.; CASSANDRO, M.; RIAZ, H.; LIANG, X.; YANG, L. Genome-Wide SNP Data Revealed the Extent of Linkage Disequilibrium, Persistence of Phase and Effective Population Size in Purebred and Crossbred Buffalo Populations. **Frontiers in genetics**, v. 9, n. JAN, 2019.
- DOEKES, H. P.; VEERKAMP, R. F.; BIJMA, P.; DE JONG, G.; HIEMSTRA, S. J.; WINDIG, J. J. Inbreeding depression due to recent and ancient inbreeding in Dutch Holstein–Friesian dairy cattle. **Genetics Selection Evolution**, v. 51, n. 1, p. 54, 27 dez. 2019. Disponível em: <<https://doi.org/10.1186/s12711-019-0497-z>>. Acesso em: 1 nov. 2019.
- FABBRI, M. C.; DADOUSIS, C.; TIEZZI, F.; MALTECCA, C.; LOZADA-SOTO, E.; BIFFANI, S.; BOZZI, R. Genetic diversity and population history of eight Italian beef cattle breeds using measures of autozygosity. **PLoS ONE**, v. 16, n. 10 October, p. e0248087, 25 out. 2021. Disponível em: <<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0248087>>. Acesso em: 18 nov. 2021.
- FALCONER, D. S.; MACKAY, T. F. C. **Introduction to quantitative genetics**. 4th. ed. [s.l.: s.n.]463 p.
- FARIELLO, M. I.; BOITARD, S.; NAYA, H.; SANCRISTOBAL, M.; SERVIN, B. Detecting Signatures of Selection Through Haplotype Differentiation Among Hierarchically Structured Populations. **Genetics**, v. 193, n. 3, p. 929–941, 1 mar. 2013.
- FERENČAKOVIĆ, M.; HAMZIĆ, E.; GREDLER, B.; SOLBERG, T. R.; KLEMETSDAL, G.; CURIK, I.; SÖLKNER, J. Estimates of autozygosity derived from runs of homozygosity: Empirical evidence from selected cattle populations. **Journal of Animal Breeding and Genetics**, v. 130, n. 4, p. 286–293, ago. 2013.
- FERENČAKOVIĆ, M.; SÖLKNER, J.; CURIK, I. Estimating autozygosity from high-

throughput information: effects of SNP density and genotyping errors. **Genetics Selection Evolution**, v. 45, n. 1, p. 42, 29 dez. 2013. . Acesso em: 13 abr. 2021.

FONSECA, L. F. S.; DOS SANTOS SILVA, D. B.; GIMENEZ, D. F. J.; BALDI, F.; FERRO, J. A.; CHARDULO, L. A. L.; DE ALBUQUERQUE, L. G. Gene expression profiling and identification of hub genes in Nellore cattle with different marbling score levels. **Genomics**, v. 112, n. 1, p. 873–879, jan. 2020a.

FONSECA, P. A. S.; SUÁREZ-VEGA, A.; MARRAS, G.; CÁNOVAS, Á. GALLO: An R package for genomic annotation and integration of multiple data sources in livestock for positional candidate loci. **GigaScience**, v. 9, n. 12, p. 1–9, 30 dez. 2020b. . Acesso em: 2 mar. 2021.

GHOREISHIFAR, S. M.; MORADI-SHAHRBABA, H.; FALLAHI, M. H.; JALIL SARGHALE, A.; MORADI-SHAHRBABA, M.; ABDOLLAHI-ARPANAHI, R.; KHANSEFID, M. Genomic measures of inbreeding coefficients and genome-wide scan for runs of homozygosity islands in Iranian river buffalo, *Bubalus bubalis*. **BMC Genetics**, v. 21, n. 1, p. 16, 10 dez. 2020. Disponível em: <<https://bmckenet.biomedcentral.com/articles/10.1186/s12863-020-0824-y>>. Acesso em: 19 abr. 2021.

GRIGOLETTO, L.; FERRAZ, J. B. S.; OLIVEIRA, H. R.; ELER, J. P.; BUSSIMAN, F. O.; ABREU SILVA, B. C.; BALDI, F.; BRITO, L. F. Genetic Architecture of Carcass and Meat Quality Traits in Montana Tropical® Composite Beef Cattle. **Frontiers in Genetics**, v. 11, 27 fev. 2020.

GROENEVELD, E.; WESTHUIZEN, B. D.; MAIWASHE, A.; VOORDEWIND, F.; FERRAZ, J. B. POPREP: a generic report for population management. **Genetics and molecular research : GMR**, v. 8, n. 3, p. 1158–1178, 2009.

GUTIÉRREZ, J.; GOYACHE, F.; CERVANTES, I. Endog v4.8 – a computer program for monitoring genetic variability of populations using pedigree information. **User's Guide. Universidad Complutense de Madrid. Madrid, España**, p. 45p, 2010.

HAEGEMAN, A.; WILLIAMS, J. L.; LAW, A.; VAN ZEVEREN, A.; PEELMAN, L. J. Mapping and SNP analysis of bovine candidate genes for meat and carcass quality. **Animal Genetics**, v. 34, n. 5, p. 349–353, out. 2003.

HOWARD, J. T.; PRYCE, J. E.; BAES, C.; MALTECCA, C. Invited review: Inbreeding in the genomics era: Inbreeding, inbreeding depression, and management of

- genomic variability. **Journal of Dairy Science**, v. 100, n. 8, p. 6009–6024, ago. 2017.
- HOWRIGAN, D. P.; SIMONSON, M. A.; KELLER, M. C. Detecting autozygosity through runs of homozygosity: A comparison of three autozygosity detection algorithms. **BMC Genomics**, v. 12, n. 1, p. 460, 23 dez. 2011. Disponível em: <<http://www.biomedcentral.com/1471-2164/12/460>>. Acesso em: 12 abr. 2021.
- ITO, R. H.; PRADO, I. N. do; VISENTAINER, J. V.; PRADO, R. M. do; FUGTA, C. A.; PIRES, M. C. de O. Carcass characteristics, chemical and fatty acid composition of Longissimus muscle of Purunã bulls slaughtered at 18 or 24 months of age. **Acta Scientiarum. Animal Sciences**, v. 32, n. 3, p. 299–307, 3 set. 2010.
- JAHUEY-MARTÍNEZ, F. J.; PARRA-BRACAMONTE, G. M.; SIFUENTES-RINCÓN, A. M.; MORENO-MEDINA, V. R. Signatures of selection in Charolais beef cattle identified by genome-wide analysis. **Journal of Animal Breeding and Genetics**, v. 136, n. 5, p. 378–389, 1 set. 2019.
- JIA, P.; CAI, C.; QU, K.; CHEN, N.; JIA, Y.; HANIF, Q.; LIU, J.; ZHANG, J.; CHEN, H.; HUANG, B.; LEI, C. Four novel SNPs of MYO1A gene associated with heat-tolerance in Chinese cattle. **Animals**, v. 9, n. 11, 1 nov. 2019.
- LARMER, S. G.; SARGOLZAEI, M.; SCHENKEL, F. S. Extent of linkage disequilibrium, consistency of gametic phase, and imputation accuracy within and across Canadian dairy breeds. **Journal of dairy science**, v. 97, n. 5, p. 3128–3141, 2014.
- LEE, S. H.; CHOI, B. H.; LIM, D.; GONDRO, C.; CHO, Y. M.; DANG, C. G.; SHARMA, A.; JANG, G. W.; LEE, K. T.; YOON, D.; LEE, H. K.; YEON, S. H.; YANG, B. S.; KANG, H. S.; HONG, S. K. Genome-Wide Association Study Identifies Major Loci for Carcass Weight on BTA14 in Hanwoo (Korean Cattle). **PLoS ONE**, v. 8, n. 10, p. e74677, 7 out. 2013.
- MALÉCOT, G. **Les mathématiques de l'hérédité**. [s.l.] Masson, 1948. 63 p.
- MARRAS, G.; GASPA, G.; SORBOLINI, S.; DIMAURO, C.; AJMONE-MARSAN, P.; VALENTINI, A.; WILLIAMS, J. L.; MACCIOTTA, N. P. P. Analysis of runs of homozygosity and their relationship with inbreeding in five cattle breeds farmed in Italy. **Animal Genetics**, v. 46, n. 2, p. 110–121, 1 abr. 2015.
- MARRAS, G.; WOOD, B. J.; MAKANJUOLA, B.; MALCHIODI, F.; PEETERS, K.; AS, P. van; BAES, C. F.; BISCARINI, F.; TURKEYS, H. Characterization of runs of

homozygosity and heterozygosity-rich regions in a commercial turkey (*Meleagris gallopavo*) population. Em: Proceedings of the 11th World Congress of Genetics Applied to Livestock Production, Szwaczkowski 2017., 2018, Auckland. [...].

Auckland: 2018. v. 11, p. 763–767.

MASTRANGELO, S.; CIANI, E.; SARDINA, M. T.; SOTTILE, G.; PILLA, F.; PORTOLANO, B. Runs of homozygosity reveal genome-wide autozygosity in Italian sheep breeds. **Animal genetics**, v. 49, n. 1, p. 71–81, 1 fev. 2018.

MCKAY, S. D.; SCHNABEL, R. D.; MURDOCH, B. M.; MATUKUMALLI, L. K.; AERTS, J.; COPPIETERS, W.; CREWS, D.; NETO, E.; GILL, C. A.; GAO, C.; MANNEN, H.; STOTHARD, P.; WANG, Z.; VAN TASSELL, C. P.; WILLIAMS, J. L.; TAYLOR, J. F.; MOORE, S. S. Whole genome linkage disequilibrium maps in cattle. **BMC genetics**, v. 8, 25 out. 2007.

MCQUILLAN, R.; LEUTENEGGER, A.-L.; ABDEL-RAHMAN, R.; FRANKLIN, C. S.; PERICIC, M.; BARAC-LAUC, L.; SMOLEJ-NARANCIC, N.; JANICIJEVIC, B.; POLASEK, O.; TENESA, A.; MACLEOD, A. K.; FARRINGTON, S. M.; RUDAN, P.; HAYWARD, C.; VITART, V.; RUDAN, I.; WILD, S. H.; DUNLOP, M. G.; WRIGHT, A. F.; CAMPBELL, H.; WILSON, J. F. Runs of Homozygosity in European Populations. **The American Journal of Human Genetics**, v. 83, n. 3, p. 359–372, 12 set. 2008. Disponível em:

<<https://linkinghub.elsevier.com/retrieve/pii/S000292970800445X>>. Acesso em: 31 out. 2019.

MEUWISSEN, T. H. Accuracy of breeding values of “unrelated” individuals predicted by dense SNP genotyping. **Genetics Selection Evolution**, v. 41, n. 1, p. 35, 11 dez. 2009. Disponível em: <www.ncbi.nlm.nih.gov/sites/entrez?db=genome>. Acesso em: 21 abr. 2021.

MEUWISSEN, T. H. E.; LUO, Z. Computing inbreeding coefficients in large populations. **Genetics, Selection, Evolution**, v. 24, n. 4, p. 305–313, 1992.

MISSIO, R. L.; RESTLE, J.; MOLETTA, J. L.; KUSS, F.; NEIVA, J. N. M.; ELEJALDE, D. A. G.; MOURA, I. C. F.; PRADO, I. N.; MIOTTO, F. R. C. Slaughter weights on animal performance, carcass commercial cuts and meat characteristics of cull cows. **Semina: Ciências Agrárias**, v. 36, n. 6, p. 3827, 9 dez. 2015.

MORAVČÍKOVÁ, N.; TRAKOVICKÁ, A.; KADLEČÍK, O.; KASARDA, R.

Genomic signatures of selection in cattle through variation of allele frequencies and linkage disequilibrium. **Journal of Central European Agriculture**, v. 20, n. 2, p. 576–580, 2019.

MOURA, I. C. F.; KUSS, F.; MOLETTA, J. L.; DE MENEZES, L. F. G.; HENRIQUE, D. S.; LIPINSKI, L. C.; MARTINS, A. de S. Performance of purunã beef cows subjected to different nursing managements. **Pesquisa Agropecuaria Brasileira**, v. 49, n. 1, p. 49–56, jan. 2014.

NEOGEN. **GeneSeek® Genomic Profiler™ Bovine 100K**. [s.l: s.n.].

OTTO, P. I.; DOS SANTOS, A. L.; PEROTTO, D.; DE OLIVEIRA, S. N.; GRANZOTTO, F.; GOBO, D. O. R.; DE SOUZA, F. E. Z.; THOMAZINI, G. Estimation of genetic parameters for weaning and yearling weights in a composite population used to form the Purunã breed. **Revista Brasileira de Zootecnia**, v. 50, p. 1–11, 2021.

PERIPOLLI, E.; METZGER, J.; DE LEMOS, M. V. A.; STAFUZZA, N. B.; KLUSKA, S.; OLIVIERI, B. F.; FEITOSA, F. L. B.; BERTON, M. P.; LOPES, F. B.; MUNARI, D. P.; LÔBO, R. B.; MAGNABOSCO, C. de U.; DI CROCE, F.; OSTERSTOCK, J.; DENISE, S.; PEREIRA, A. S. C.; BALDI, F. Autozygosity islands and ROH patterns in Nellore lineages: evidence of selection for functionally important traits. **BMC Genomics**, v. 19, n. 1, p. 680, 17 dez. 2018.

PERIPOLLI, E.; STAFUZZA, N. B.; AMORIM, S. T.; DE LEMOS, M. V. A.; GRIGOLETTO, L.; KLUSKA, S.; FERRAZ, J. B. S.; ELER, J. P.; MATTOS, E. C.; BALDI, F. Genome-wide scan for runs of homozygosity in the composite Montana Tropical® beef cattle. **Journal of Animal Breeding and Genetics**, v. 137, n. 2, p. 155–165, 8 mar. 2020. Disponível em:

<<https://onlinelibrary.wiley.com/doi/abs/10.1111/jbg.12428>>. Acesso em: 19 ago. 2020.

PEROTTO, D.; CUBAS, A. C.; MOLETTA, J. L.; LESSKIU, C. Heterosis upon weights in Canchim and Aberdeen Angus calves and in their reciprocal crosses. **Pesquisa Agropecuaria Brasileira**, v. 35, n. 12, p. 2511–2520, 2000.

PEROTTO, D.; MOLETTA, J. L.; CUBAS, A. C. Características quantitativas da carcaça de bovinos Charolês, Caracu e cruzamentos recíprocos terminados em confinamento. **Revista Brasileira de Zootecnia**, v. 29, n. 1, p. 117–124, jan. 2000.

PIRES, B. V.; STAFUZZA, N. B.; DE FREITAS, L. A.; MERCADANTE, M. E. Z.;

- RAMOS, E. S.; PAZ, C. C. P. Expression of candidate genes for residual feed intake in tropically adapted *Bos taurus* and *Bos indicus* bulls under thermoneutral and heat stress environmental conditions. **Journal of Thermal Biology**, v. 99, p. 102998, 1 jul. 2021.
- PONTES, D. S.; ROSADO, R. D. S.; CRUZ, C. D.; NASCIMENTO, M.; OLIVEIRA, A. M. C.; PENSKY, S. M. Trait selection using procrustes analysis for the study of genetic diversity in conilon coffee. **Acta Scientiarum - Agronomy**, v. 42, n. 1, p. 1–12, 27 maio 2020.
- PRIEUR, V.; CLARKE, S. M.; BRITO, L. F.; MCEWAN, J. C.; LEE, M. A.; BRAUNING, R.; DODDS, K. G.; AUVRAY, B. Estimation of linkage disequilibrium and effective population size in New Zealand sheep using three different methods to create genetic maps. **BMC Genetics**, v. 18, n. 1, 2017.
- PURCELL, S.; NEALE, B.; TODD-BROWN, K.; THOMAS, L.; FERREIRA, M. A. R.; BENDER, D.; MALLER, J.; SKLAR, P.; DE BAKKER, P. I. W.; DALY, M. J.; SHAM, P. C. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. **The American Journal of Human Genetics**, v. 81, n. 3, p. 559–575, set. 2007. . Acesso em: 2 mar. 2019.
- REBELATO, A. B.; CAETANO, A. R. Runs of homozygosity for autozygosity estimation and genomic analysis in production animals. **Pesquisa Agropecuária Brasileira**, v. 53, n. 9, p. 975–984, set. 2018. . Acesso em: 24 fev. 2019.
- REYNOLDS, J.; WEIR, B. S.; COCKERHAM, C. C. ESTIMATION OF THE COANCESTRY COEFFICIENT: BASIS FOR A SHORT-TERM GENETIC DISTANCE. 1983.
- ROSEN, B. D.; BICKHART, D. M.; SCHNABEL, R. D.; KOREN, S.; ELSIK, C. G.; TSENG, E.; ROWAN, T. N.; LOW, W. Y.; ZIMIN, A.; COULDREY, C.; HALL, R.; LI, W.; RHIE, A.; GHURYE, J.; MCKAY, S. D.; THIBAUD-NISSEN, F.; HOFFMAN, J.; MURDOCH, B. M.; SNELLING, W. M.; MCDANELD, T. G.; HAMMOND, J. A.; SCHWARTZ, J. C.; NANDOLO, W.; HAGEN, D. E.; DREISCHER, C.; SCHULTHEISS, S. J.; SCHROEDER, S. G.; PHILLIPPY, A. M.; COLE, J. B.; VAN TASSELL, C. P.; LIU, G.; SMITH, T. P. L.; MEDRANO, J. F. De novo assembly of the cattle reference genome with single-molecule sequencing. **GigaScience**, v. 9, n. 3, p. 1–9, 2020.
- SANTOS, W.; SCHETTINI, G.; FONSECA, M. G.; PEREIRA, G. L.; CHARDULO, L.

- A.; NETO, O.; BALDASSINI, W. A.; OLIVEIRA, H.; CURI, R. Fine-scale estimation of inbreeding rates, runs of homozygosity and genome-wide heterozygosity levels in the Mangalarga Marchador horse breed. **Journal of Animal Breeding and Genetics**, v. 138, n. 2, p. 161–173, 19 mar. 2021. . Acesso em: 26 abr. 2021.
- SAS INSTITUTE INC. **SAS 9.1.3 Help and Documentation**. Cary: ADABAS, 2013.
- SEMPÉRÉ, G.; MOAZAMI-GOUDARZI, K.; EGGEN, A.; LALOË, D.; GAUTIER, M.; FLORI, L. WIDDE: A Web-Interfaced next generation database for genetic diversity exploration, with a first application in cattle. **BMC Genomics**, v. 16, n. 1, p. 1–8, 2015. Disponível em: <<http://dx.doi.org/10.1186/s12864-015-2181-1>>.
- SIEVERT, C. **Interactive web-based data visualization with R, plotly, and shiny**. [s.l: s.n.]
- SZMATOŁA, T.; GURGUL, A.; JASIELCZUK, I.; ZĄBEK, T.; ROPKA-MOLIK, K.; LITWIŃCZUK, Z.; BUGNO-PONIEWIERSKA, M. A comprehensive analysis of runs of homozygosity of eleven cattle breeds representing different production types. **Animals**, v. 9, n. 12, p. 1024, 25 nov. 2019.
- TAYE, M.; YOON, J.; DESSIE, T.; CHO, S.; OH, S. J.; LEE, H. K.; KIM, H. Deciphering signature of selection affecting beef quality traits in Angus cattle. **Genes and Genomics**, v. 40, n. 1, p. 63–75, 2018.
- URBINATI, I.; STAFUZZA, N. B.; OLIVEIRA, M. T.; CHUD, T. C. S.; HIGA, R. H.; REGITANO, L. C. de A.; DE ALENCAR, M. M.; BUZANSKAS, M. E.; MUNARI, D. P. Selection signatures in Canchim beef cattle. **Journal of Animal Science and Biotechnology**, v. 7, n. 1, p. 1–9, 5 maio 2016.
- VANRADEN, P. M. Efficient methods to compute genomic predictions. **Journal of Dairy Science**, v. 91, n. 11, p. 4414–4423, 1 nov. 2008. . Acesso em: 4 jul. 2021.
- VILLANUEVA, B.; FERNÁNDEZ, A.; SAURA, M.; CABALLERO, A.; FERNÁNDEZ, J.; MORALES-GONZÁLEZ, E.; TORO, M. A.; PONG-WONG, R. The value of genomic relationship matrices to estimate levels of inbreeding. **Genetics Selection Evolution**, v. 53, n. 1, p. 1–17, 1 dez. 2021.
- WANG, J. Marker-based estimates of relatedness and inbreeding coefficients: An assessment of current methods. **Journal of Evolutionary Biology**, v. 27, n. 3, p. 518–530, mar. 2014.
- WANG, J.; LIAO, Y.; JAEHNIG, E.; SHI, Z.; SHENG, Q. **Gene Set Analysis Toolkit**

WebGestaltR. [s.l.: s.n.]. Disponível em: <<https://github.com/bzhanglab/WebGestaltR>>.

Acesso em: 3 mar. 2021.

WILLIAMS, J. L.; HALL, S. J. G.; DEL CORVO, M.; BALLINGALL, K. T.; COLLI, L.; AJMONE MARSAN, P.; BISCARINI, F. Inbreeding and purging at the genomic Level: the Chillingham cattle reveal extensive, non-random SNP heterozygosity.

Animal Genetics, v. 47, n. 1, p. 19–27, 1 fev. 2016. . Acesso em: 26 abr. 2021.

WRIGHT, S. Coefficients of Inbreeding and Relationship. **The American Naturalist**, v. 56, n. 645, p. 330–338, jul. 1922. Disponível em:

<<https://www.jstor.org/stable/pdf/2456273.pdf>>. Acesso em: 21 abr. 2021.

XIA, X.; ZHANG, S.; ZHANG, H.; ZHANG, Z.; CHEN, N.; LI, Z.; SUN, H.; LIU, X.; LYU, S.; WANG, X.; LI, Z.; YANG, P.; XU, J.; DING, X.; SHI, Q.; WANG, E.; RU, B.; XU, Z.; LEI, C.; CHEN, H.; HUANG, Y. Assessing genomic diversity and

signatures of selection in Jiaxian Red cattle using whole-genome sequencing data. **BMC Genomics**, v. 22, n. 1, p. 43, 9 dez. 2021. Disponível em:

<<https://bmcbgenomics.biomedcentral.com/articles/10.1186/s12864-020-07340-0>>.

Acesso em: 26 abr. 2021.

YANG, J.; BENYAMIN, B.; MCEVOY, B. P.; GORDON, S.; HENDERS, A. K.; NYHOLT, D. R.; MADDEN, P. A.; HEATH, A. C.; MARTIN, N. G.;

MONTGOMERY, G. W.; GODDARD, M. E.; VISSCHER, P. M. Common SNPs explain a large proportion of the heritability for human height. **Nature Genetics**, v. 42, n. 7, p. 565–569, jul. 2010. . Acesso em: 4 jul. 2021.

ZHAO, F.; MCPARLAND, S.; KEARNEY, F.; DU, L.; BERRY, D. P. Detection of selection signatures in dairy and beef cattle using high-density genomic information. **Genetics Selection Evolution**, v. 47, n. 1, p. 49, 19 dez. 2015

CAPÍTULO 3

AVALIAÇÃO DE PARÂMETROS QUE INFLUENCIAM A IDENTIFICAÇÃO DE REGIÕES RICAS EM HETEROZIGOSE EM REBANHO HOLANDÊS COM BASE EM PAINEL DE SNP OU DADOS DE SEQUÊNCIA DO GENOMA COMPLETO

Resumo

Regiões ricas em heterozigose (HERs - do termo em inglês “*Heterozygous-enriched regions*”) são regiões genômicas com alta variabilidade gerada por fatores como balanço de seleção, introgressão e admixture. Neste estudo, foi avaliado o impacto de diferentes parâmetros (número mínimo de SNPs em uma HER, distância máxima entre dois SNPs consecutivos, comprimento mínimo de uma HER, número máximo de homozigotos permitido em uma HER) e cenários (diferentes densidades de painel SNP) na detecção de HERs. Também foram caracterizados HERs em rebanho Holandês com base em dados de sequência do genoma completo (WGS) e comparados com HERs identificadas em animais Angus (ANG), Jersey (JER) e Norwegian Red (RDC). Os parâmetros utilizados para a identificação do HER impactaram significativamente na sua detecção. A distância máxima entre dois SNPs consecutivos avaliados não afetou a detecção de HER, pois a mesma média de HERs ($269,31 \pm 787,00$) foi observada entre os cenários. No entanto, o número mínimo de marcadores, marcadores homozigotos máximos permitidos dentro de uma HER e o tamanho mínimo do comprimento de uma HER variaram de acordo com os parâmetros selecionados. Para o tamanho de comprimento mínimo, o cenário de 10Kb apresentou o maior número médio de HER ($1.364,69 \pm 1.483,64$). O número de HERs diminuiu com o aumento do número mínimo de marcadores ($621,31 \pm 1.271,83$ para $6,08 \pm 21,94$), e um padrão oposto foi observado ao máximo de marcadores homozigotos permitidos dentro de uma HER ($54,47 \pm 195,51$ para $494,89 \pm 1.169,35$). Quarenta e cinco ilhas localizadas em 23 cromossomos com altos valores D de Tajima foram detectadas em todos os cenários avaliados. No total, 3.440 marcadores e 28 genes previamente relacionados à fertilidade (ex. *TP63*, *ZSCAN23*, *NEK5*, *ARHGAP44*), consumo alimentar residual (ex. *MAYO9A*), sensibilidade ao estresse (ex. *SERPINA6*) e o percentual de gordura do leite (ex. *NOLA*) foram identificados. Ao comparar as ilhas de HER entre as raças, houve uma sobreposição substancial entre Holandês e Angus (95,28%), Jersey (94,28%) e Norwegian Red (97,14%), indicando HERs conservados entre as raças taurinas. A detecção de HER variou de acordo com os parâmetros utilizados, mas algumas HER foram consistentes em todos os diferentes cenários. Genótipos heterozigotos através de gerações e raças parecem ser conservados em regiões de HER. Os resultados apresentados servirão de guia para a definição dos parâmetros de detecção de HER em estudos futuros.

palavra-chave: corrida de heterozigose, diversidade, heterozigoto, Tajima

Abstract

Heterozygous-enriched regions (HER) are genomic regions with high variability generated by factors such as balancing selection, introgression, and admixture processes. In this study, we evaluated the impact of different parameters (e.g., minimum number of SNPs in a HER, maximum distance between two consecutive SNPs, minimum length of a HER, maximum number of homozygous allowed in a HER), and scenarios (e.g., different SNP panel densities) on the detection of HERs. We also characterized HERs in Holstein cattle based on whole-genome sequence (WGS) data and compared the HERs identified with those from Angus (ANG), Jersey (JER), and Norwegian Red cattle (RDC). The parameters used for the identification of HER significantly impact their detection. The maximum distance between two consecutive SNPs evaluated did not impact HER detection as the same average of HERs (269.31 ± 787.00) was observed across scenarios. However, the minimum number of markers, maximum homozygous markers allowed inside a HER, and the minimum length size changed according to the parameters selected. For the minimum length size, the 10Kb scenario showed the highest average number of HER ($1,364.69 \pm 1,483.64$). The number of HERs decreased as the minimum number of markers increased ($621.31 \pm 1,271.83$ to 6.08 ± 21.94), and an opposite pattern was observed to the maximum homozygous markers allowed inside a HER (54.47 ± 195.51 to $494.89 \pm 1,169.35$). Forty-five islands located in 23 chromosomes with high Tajima's D values and differential were detected in all evaluated scenarios. In total, 3,440 markers and 28 genes previously related to fertility (e.g., TP63, ZSCAN23, NEK5, ARHGAP44), immunity (e.g., TP63, IGC, ARHGAP44), residual feed intake (e.g., MAYO9A), stress sensitivity (e.g., SERPINA6), and milk fat percentage (e.g., NOL4) were identified. When comparing HER islands among breeds, there were substantial overlap between Holstein and Angus (95.28%), Jersey (94.28%), and Norwegian Red cattle (97.14%), indicating conserved HERs across taurine breeds. The detection of HER varied according to the parameters used, but some HER were consistent across all different scenarios. Heterozygous genotypes across generations and breeds appear to be conserved in HER regions. The results presented will serve as a guide for defining HER detection parameters in future studies.

keywords: diversity, heterozygous, runs of heterozygosity, Tajima

Introdução

Avaliar e desenvolver estratégias para manter a diversidade genômica em populações de gado leiteiro tornou-se uma atividade importante em programas de melhoramento devido ao acúmulo mais rápido de endogamia por ano e redução no tamanho efetivo da população como resultado de práticas de reprodução intensiva (por exemplo, inseminação artificial) e esquemas de seleção genômica (BAES *et al.*, 2019; FORUTAN *et al.*, 2018; LOZADA-SOTO *et al.*, 2022). Numerosos estudos caracterizaram regiões homozigóticas (por exemplo, corridas de homozigose – ROH) em várias espécies, incluindo bovinos (GURGUL *et al.*, 2016; LOZADA-SOTO *et al.*, 2021; SZMATOŁA *et al.*, 2019), suínos (GROSSI *et al.*, 2017; HOWARD *et al.*, 2017b), ovinos (SELLI *et al.*, 2021), caprinos (BRITO *et al.*, 2017), equinos (SANTOS *et al.*, 2021a) e aves (MARRAS *et al.*, 2018). Sendo exemplo, a endogamia genômica antiga e recente podendo ser estimada com base em ROH (MULIM *et al.*, 2022a). Por outro lado, trechos de genótipos heterozigotos, também conhecidos como “regiões ricas em heterozigotos (HER)” ou “corridas de heterozigose (ROHet)”, são muito menos caracterizados do que ROH em espécies de produção (BISCARINI *et al.*, 2020) e podem indicar regiões genômicas com alta variabilidade genética.

O primeiro estudo relatando HER em animais de produção foi publicado por Williams *et al.* (2016) em bovinos Chillingham. Esta raça não havia sido submetida a seleção artificial e permaneceu como um rebanho fechado por pelo menos 350 anos (WILLIAMS *et al.*, 2016). Apesar desta particularidade, algumas regiões genômicas ainda apresentaram altas concentrações de genótipos heterozigotos, especialmente regiões contendo loci que influenciam características de aptidão e sobrevivência. Após esta primeira caracterização da HER, estudos foram publicados em ovinos (SELLI *et al.*, 2021), bovinos (MULIM *et al.*, 2022b), equinos (SANTOS *et al.*, 2021a) e suínos (RUAN *et al.*, 2022). Tais resultados indicam fortes evidências de que regiões ligadas a algumas características essenciais mantêm alta heterozigosidade ao longo das gerações. A manutenção de diversidade haplotípica em determinados loci pode conferir uma vantagem de aptidão e estar sujeita à balanço de seleção (WILLIAMS *et al.*, 2016).

Balanço de seleção descreve qualquer processo de seleção que resulte na manutenção de múltiplas variantes de loci específicos em frequências intermediárias dentro de uma população (LLAURENS; WHIBLEY; JORON, 2017). Tal seleção afirma

que a heterozigose é generalizada e persiste na população por meio da vantagem do heterozigoto (FIJARCZYK; BABIK, 2015). O balanço de seleção não é o único processo ligado à ocorrência de HER em uma população. Outros processos, como introgressão (a transferência de variantes genéticas de uma espécie para outra, por exemplo, hibridização), admixture (mistura de linhagens ou raças genéticas) e regiões hipervariáveis (porções do genoma com níveis muito mais altos de variação do que outras áreas semelhantes devido a mutações, taxa de recombinação e rearranjos cromossômicos) (MARRAS *et al.*, 2018; WILLIAMS *et al.*, 2016) contribuem para o aparecimento e manutenção de HER. Uma alta concentração de alelos heterozigotos em regiões genômicas específicas em uma grande proporção da população é definida como ilhas HER e pode informar a pressão de seleção para heterozigoto sofrida por uma população em um momento específico.

A avaliação precisa de polimorfismos em regiões genômicas com alta variabilidade apresenta desafios adicionais e pode levar à subestimação dos resultados (FIJARCZYK; BABIK, 2015). Um dos principais desafios quando se pretende quantificar HERs em uma população é a definição dos parâmetros a serem utilizados nas análises. Por exemplo, a detecção de HER pode ser afetada pela densidade dos painéis SNP, distribuição de marcadores em todo o genoma, qualidade da genotipagem e consistência das informações (taxas de erro e frequências do alelo menor) (CEBALLOS *et al.*, 2018). Além desses parâmetros, a identificação correta dos HERs também depende de fatores como o tamanho mínimo do comprimento de uma HER e o número de homozigotos permitidos (REBELATO; CAETANO, 2018). Apesar da influência de todos esses fatores na identificação de HER e ilhas de HER, não existem estudos avaliando o impacto de diferentes parâmetros utilizados nas análises de detecção de HER. Portanto, os principais objetivos deste estudo foram: 1) avaliar o impacto de diferentes parâmetros (número mínimo de SNPs em um HER, distância máxima entre dois SNPs consecutivos, comprimento mínimo de um HER, número máximo de homozigotos permitido em um HER) e cenários (densidades de painel SNP e dados de sequência do genoma completo – WGS) na detecção de HER; 2) caracterizar HERs no rebanho Holandês com base em dados de sequência do genoma completo (WGS), seguido de uma análise genômica funcional das ilhas HER identificadas; e, 3) avaliar a sobreposição dos HER encontrados na raça Holandesa com os de outras raças taurinas (*Bos taurus taurus*),

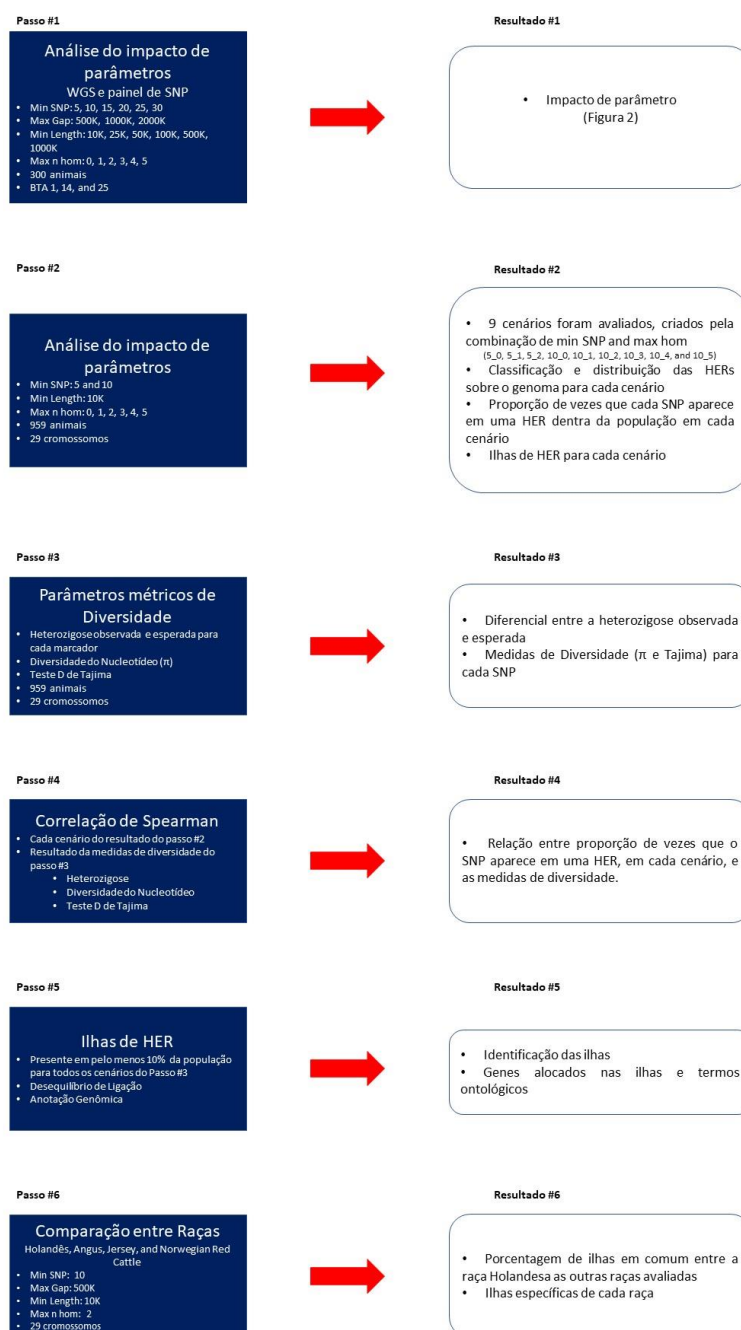
incluindo Angus (ANG), Jersey (JER) e Norwegian Red (RDC).

Material e Métodos

Dados e controle de qualidade

Novecentos e cinquenta e nove (959) animais da raça Holandesa do Projeto 1.000 Bull Genomes (HAYES; DAETWYLER, 2019) foram usados para este estudo. Os dados do WGS para esses indivíduos continham 47.379.463 marcadores distribuídos pelos 29 cromossomos autossômicos. O controle de qualidade (CQ) foi realizado seguindo os critérios propostos por Ferenčaković, Sölkner e Curik (2013) e Biscarini et al. (2020) em estudos de HER e ROH. Em resumo, o QC removeu SNPs com baixa taxa de chamada ($<0,95$), posições duplicadas, localizadas em cromossomos não autossômicos ou com posições desconhecidas. A Figura 1 apresenta um resumo das análises realizadas.

Figura 1 - Roteiro das análises realizadas. Min SNP: número mínimo de SNPs em um HER; Max Gap: distância máxima entre dois SNPs consecutivos (Kb); Min Length: comprimento mínimo de um HER (Kb); Max n hom: Número máximo de homozigotos permitido em um HER



Identificação de Regiões Ricas em Heterozigotos

O pacote detectRUNs (BISCARINI *et al.*, 2019) foi utilizado para identificar HERs, aplicando a abordagem consecutiva nas análises, que escaneiam diretamente o genoma SNP por SNP, conforme proposto por Marras *et al.* (2015). O impacto de diferentes conjuntos de parâmetros na identificação de HERs foi avaliado. Esses parâmetros incluíam:

- Número mínimo de SNPs em uma HER: 5, 10, 15, 20, 25 e 30;
- Distância máxima entre dois SNPs consecutivos (GAP em Kb): 500Kb, 1.000Kb e 2.000Kb;
- Comprimento mínimo de uma HER (Kb): 10Kb, 25Kb, 50Kb, 100Kb, 500Kb e 1.000Kb;
- Número máximo de homozigotos permitido em uma HER: 0, 1, 2, 3, 4 e 5;

A combinação de todos os conjuntos de parâmetros resultou em 648 análises. Portanto, os seguintes critérios foram empregados para otimizar o tempo e a utilização dos processadores computacionais. Um subconjunto de 300 animais da raça Holandesa foi escolhido aleatoriamente dentre os 959 disponíveis. Além disso, o impacto dos parâmetros na identificação de HERs foi avaliado em três cromossomos selecionados com 5.351.067 marcadores (BTA1 – 2.987.435 marcadores, BTA14 – 1.544.553 marcadores, BTA25 – 819.079 marcadores). Esses cromossomos foram escolhidos para representar cromossomos de comprimento longo, médio e curto. Como o número de HERs detectados é baseado em uma combinação de parâmetros (ou seja, não baseado em um único parâmetro), a média e o desvio padrão do número de HERs detectados, correspondentes a cada parâmetro, foram calculados para avaliar o impacto de cada parâmetro na detecção de HER. Por exemplo, o efeito do número mínimo de SNPs igual a 5 na detecção de HER foi medido pela média e desvio padrão do número total de HERs detectados para cada cenário onde o número mínimo de SNPs foi igual a 5.

Foram avaliados os efeitos de todos os parâmetros em HERs ao usar dados de painéis de SNP (menor densidade de SNP em comparação com WGS). Três painéis SNP foram derivados dos dados WGS retendo seletivamente variantes presentes em painéis SNP comerciais: 50K (BovineSNP50), 100K (GGP Bovine 100K) e HD (700K – BovineHD BeadChip).

Por fim, como os resultados não mostraram uma diferenciação relevante com base

no número mínimo de SNPs e homozigotos permitidos, cenários adicionais foram criados. Portanto, foram considerados o número mínimo de marcadores igual a 5 e 10, combinando os homozigotos permitidos dentro do HER variando de 0 a 5 (0-2 para marcadores mínimos iguais a 5, e 0-5 para marcadores mínimos iguais a 10), e fixando o tamanho mínimo de comprimento de uma HER em 10Kb (totalizando 9 cenários). Essa abordagem foi usada para observar a combinação dos melhores parâmetros, usando todos os cromossomos autossômicos, na detecção de HERs.

Estimativa de heterozigose, diversidade de nucleotídeos e teste estatístico D de Tajima

Os níveis de heterozigose observados e esperados para cada marcador foram obtidos usando as estatísticas do teste de Hardy-Weinberg implementadas no software PLINK v.1.09 (PURCELL *et al.*, 2007). O programa VCFtools (DANECEK *et al.*, 2011) foi usado para estimar a diversidade de nucleotídeos de cada marcador (π) e para realizar o teste D de Tajima (TAJIMA, 1989). A diferença entre a heterozigose observada e esperada, a diversidade de nucleotídeos e a estatística D de Tajima foram usadas coletivamente para correlacionar com os resultados da proporção de vezes que o SNP aparece em uma HER dentro da população, usando a correlação de Spearman a fim de observar a relação entre cenários criados no segundo passo e as métricas de diversidade populacional.

Identificação de ilhas HER, desequilíbrio de ligação e anotação genômica

As regiões de HERs presentes em pelo menos 10% dos indivíduos e em todos os cenários do passo dois avaliados foram consideradas como ilhas de HERs. O desequilíbrio de ligação (LD) para todos os marcadores dentro dessas ilhas de HER foi estimado usando o pacote LDheatmap [25]. A anotação genômica dessas regiões foi realizada usando o pacote GALLO (FONSECA *et al.*, 2020b) com os dados anotados para *Bos taurus* do banco de dados Ensembl (www.ensembl.org/Bos_taurus/Info/Index), versão ARS-UCD1.2 (ROSEN *et al.*, 2020). Posteriormente, o pacote WebGestaltR (WANG *et al.*, 2020) foi usado para minerar a Ontologia Gênica (GO) e identificar potenciais processos biológicos, funções moleculares, componentes celulares e vias metabólicas nos quais os genes candidatos posicionais podem estar envolvidos.

Comparação das ilhas HER da raça Holandesa com outras raças taurina

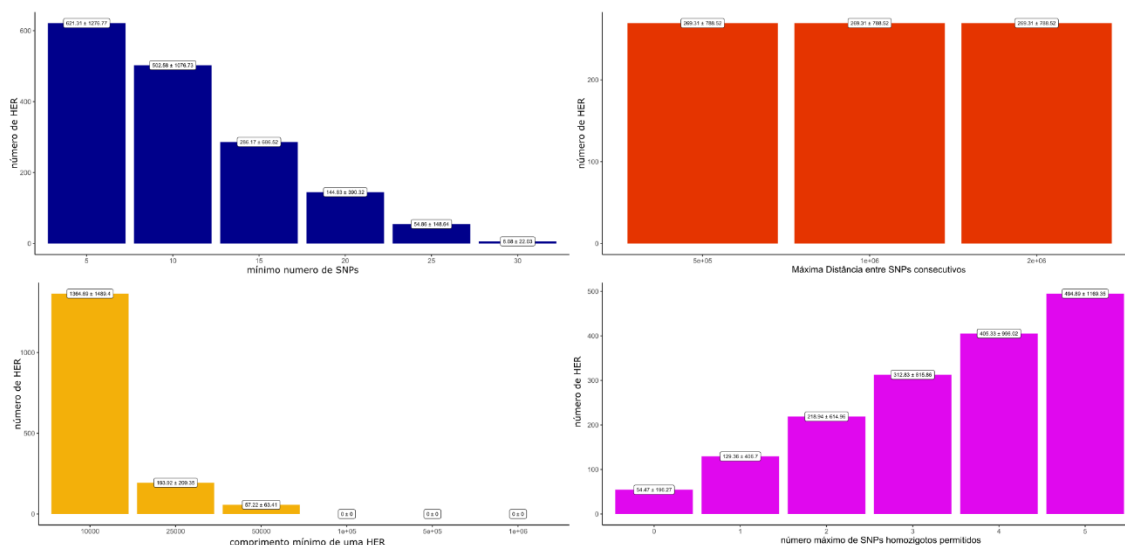
As ilhas HER identificadas na raça Holandesa foram comparadas com as de outras três raças taurinas [ANG (317 animais), JER (179 animais) e RDC (179 animais)] do Projeto 1000 Bull Genomes (HAYES; DAETWYLER, 2019b). Os parâmetros utilizados para a identificação de HERs em todas as raças foram:

- Número mínimo de SNPs em um HER igual a 10;
- Distância máxima entre dois SNPs consecutivos (GAP em Kb) igual a 500 Kb;
- Comprimento mínimo de um HER (Kb) igual a 10 Kb;
- O número máximo de genótipos homozigotos permitidos em um HER foi igual a dois.

Resultados

O impacto dos parâmetros usados na identificação da HER

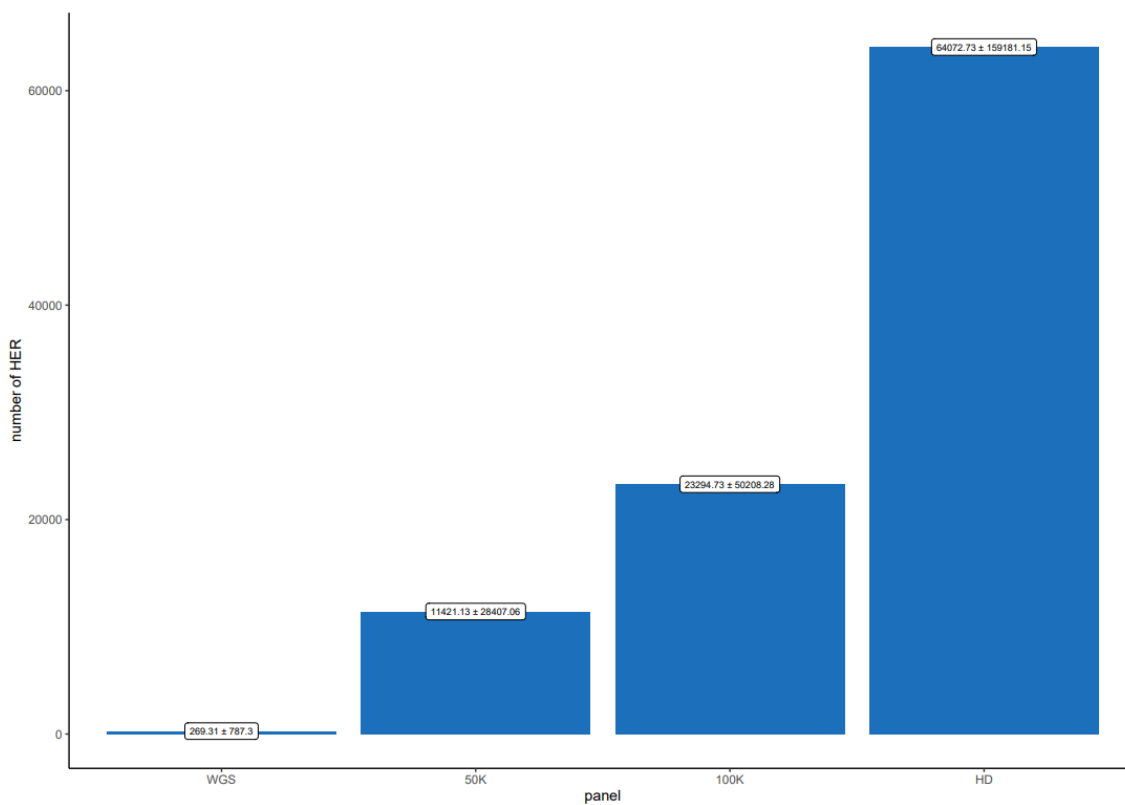
O efeito dos parâmetros investigados no número de HERs identificados nos dados WGS é relatado na Figura 2. A distância máxima de SNPs consecutivos não impactou o número médio de HERs detectados. Em contraste, aumentar o número mínimo de SNP em um HER diminuiu significativamente o número de HERs de $621,31(\pm 1.276,77)$, ao usar 5 SNPs, para $6,08(\pm 6,08)$, usando 30 SNPs. O mesmo padrão foi observado para o tamanho do comprimento mínimo do HER, onde uma diminuição significativa na detecção do HER foi observada com o aumento do tamanho, mas com maior diferenciação quando o parâmetro foi igual a 10Kb para os outros comprimentos avaliados. O tamanho mínimo de 10Kb apresentou um aumento no número de HER's de 603% em relação a 25Kb e 2.283% em relação a 50Kb. O tamanho mínimo de comprimento para seções maiores que 100Kb não detectou HERs, conforme mostrado na Figura 2.

Figura 2 - Efeito do parâmetro na detecção de regiões ricas em heterozigose

Ao considerar o número máximo de marcadores homozigotos permitidos dentro da HER, o padrão foi o oposto em relação ao número mínimo de marcadores e tamanho do comprimento da HER. Maior número de HERs foram observados quando mais marcadores homozigotos foram permitidos em uma HER. A variação desse parâmetro foi de 54,47 ($\pm 196,27$) para 0 marcadores homozigotos a 494,89 ($\pm 1169,35$) para 5 marcadores homozigotos, um aumento em média de 808% na detecção de HERs. Um padrão semelhante foi observado nos resultados do painel HD para todos os parâmetros nos painéis comerciais. Porém, nos painéis SNP de 50K e 100K, quanto ao comprimento mínimo de HER, observou-se maior diferença entre os tamanhos escolhidos a partir do 100K com faixa de 16.542,94 ($\pm 34.229,63$) em 10Kb a 102,97 ($\pm 194,07$) em 1.000Kb em o painel de 50Kb e 35.365,22 ($\pm 60.859,43$) a 93,36 ($\pm 111,62$) no painel de 100K.

O número médio de HERs detectados por cada um dos painéis de menor densidade avaliados é ilustrado na Figura 3. O número de HERs identificadas nos painéis SNP foi maior do que o observado nos dados WGS e com um aumento nos painéis de 50K (11.421,13 \pm 28.407,06) para o painel HD (64.072,73 \pm 159.181,15), com divergência de 4.141%, 8.550%, 23.692%, para 50, 100K e HD, respectivamente, em comparação com os dados WGS.

Figura 3 - Média e desvio padrão da detecção de regiões ricas em heterozigose em sequência do genoma completo (WGS) e painel comercial



Classificação e distribuição de HERs em todo o genoma

O número mínimo de marcadores que constituem uma HER e o número de marcadores homocigotos permitidos dentro de uma HER não apresentaram uma diferenciação clara de um único parâmetro como melhor modelo na detecção de HERs. Assim, o número mínimo de marcadores de 5 e com o número máximo de homocigotos de 0 a 2, e um número mínimo de marcadores igual a 10, com o máximo de homocigotos de 0 a 5 foram escolhidos para observar a distribuição de HER pelo genoma, bem como a classificação das HERs de acordo com o tamanho de seu comprimento. Para todos esses novos cenários, recriados no passo #2 (Figura 1), como o comprimento mínimo de 10Kb apresentou uma diferença substancial no número e HERs detectadas, esse parâmetro foi fixado nas análises.

A Tabela 1 mostra a classificação dos HERs com base no tamanho do comprimento para os cenários como o número mínimo de marcadores igual a 5 e 10. O número de HERs detectadas aumentou com o aumento de marcadores homocigotos em ambos os cenários. No cenário 5, o número de HERs variou de 36.024 a 99.811,

com os marcadores homozigotos iguais a 0 a 2, respectivamente. No cenário 10, o intervalo foi de 17.942 a 115.568 HERs detectados com os marcadores homozigotos iguais a 0 a 5, respectivamente. A maioria dos HERs detectadas em todos os cenários foram classificadas como 10-20 Kb, representando 63,30% a 75,10% no cenário 5 e 53,40% a 72,00% no cenário 10.

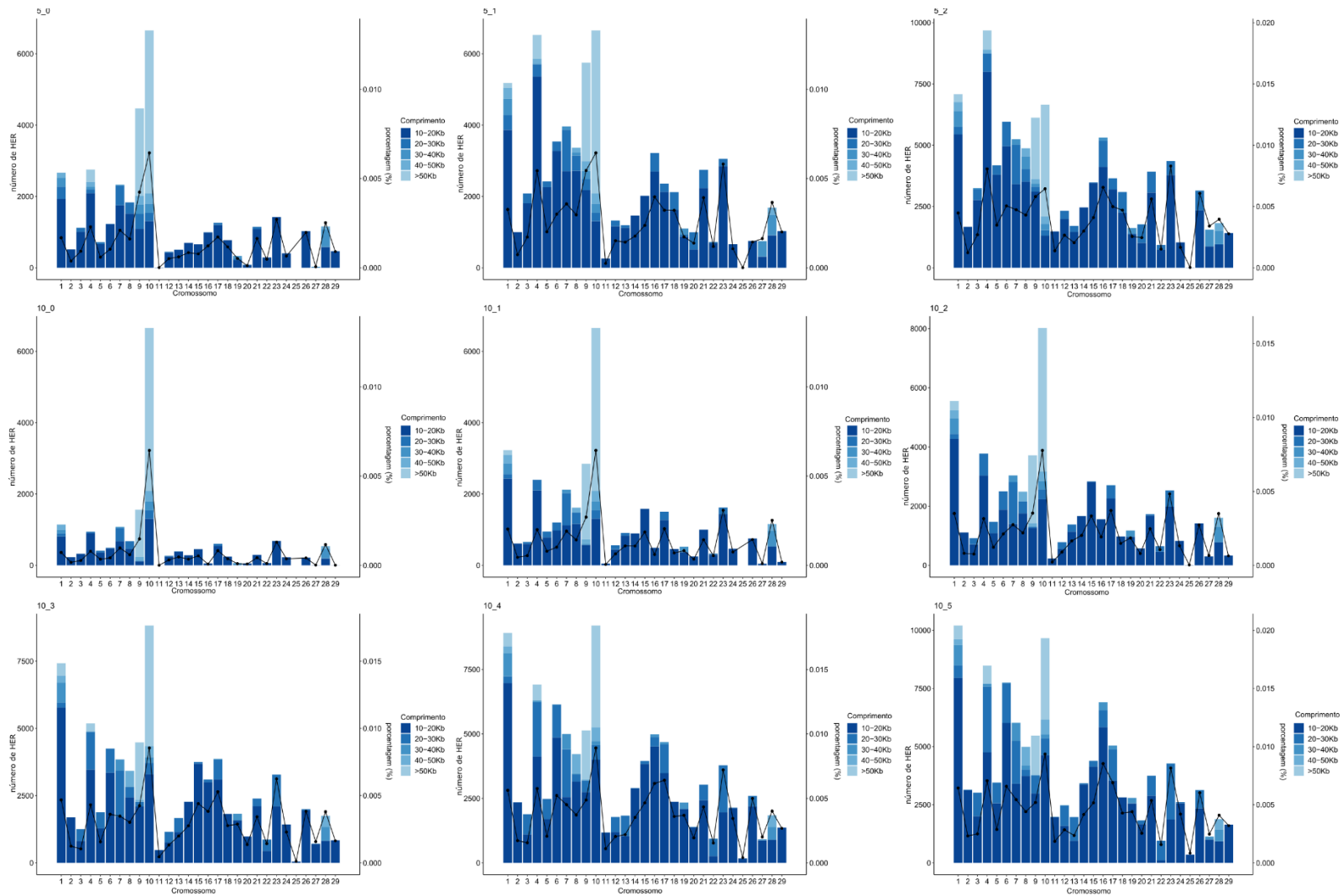
Tabela 1 - Número de regiões ricas em heterozigose detectadas e classificação com base no comprimento para os cenários com o número mínimo de marcadores iguais a 5 e 10

	5_0	5_1	5_2	10_0	10_1	10_2	10_3	10_4	10_5
10-20Kb	23.888	49.777	74.945	9.578	22.895	39.236	55.166	69.741	83.251
20-30Kb	2.426	6.396	11.624	1.131	3.007	6.059	10.015	14.941	20.015
30-40Kb	1.443	2.301	3.415	730	1.255	1.780	2.222	2.765	3.535
40-50Kb	901	1.349	1.653	613	992	1.348	1.731	2.034	2.201
>50Kb	7.366	8.130	8.174	5.890	6.821	7.471	7.381	6.998	6.566
Total	36.024	67.953	99.811	17.942	34.970	55.894	76.515	96.479	115.568

5_0: cenário número mínimo de marcadores igual a 5 e 0 homozigoto permitido dentro de um HER;
 5_1: cenário número mínimo de marcadores igual a 5 e 1 homozigoto permitido dentro de um HER;
 5_2: cenário número mínimo de marcadores igual a 5 e 2 homozigotos permitidos dentro de um HER;
 10_0: cenário número mínimo de marcadores igual a 10 e 0 homozigoto permitido dentro de um HER;
 10_1: cenário número mínimo de marcadores igual a 10 e 1 homozigoto permitido dentro de um HER;
 10_2: cenário número mínimo de marcadores igual a 10 e 2 homozigotos permitidos dentro de um HER;
 10_3: cenário número mínimo de marcadores igual a 10 e 3 homozigotos permitidos dentro de um HER;
 10_4: cenário número mínimo de marcadores igual a 10 e 4 homozigotos permitidos dentro de um HER;
 10_5: cenário número mínimo de marcadores igual a 10 e 5 homozigotos permitidos dentro de um HER;

A Figura 4 mostra a classificação de HER por cromossomo e a porcentagem de cobertura do genoma por HERs. BTA10 apresentou a maior concentração de HERs de tamanho grande (>50Kb). Além disso, o BTA10 apresentou mais HERs nos cenários 5, com homozigotos 0 e 1, e 10, com homozigotos 0 a 4. Em relação à cobertura do genoma, o percentual de HER não passou da marca de 0,010%, com o cromossomo 10 apresentando a maior taxa de cobertura.

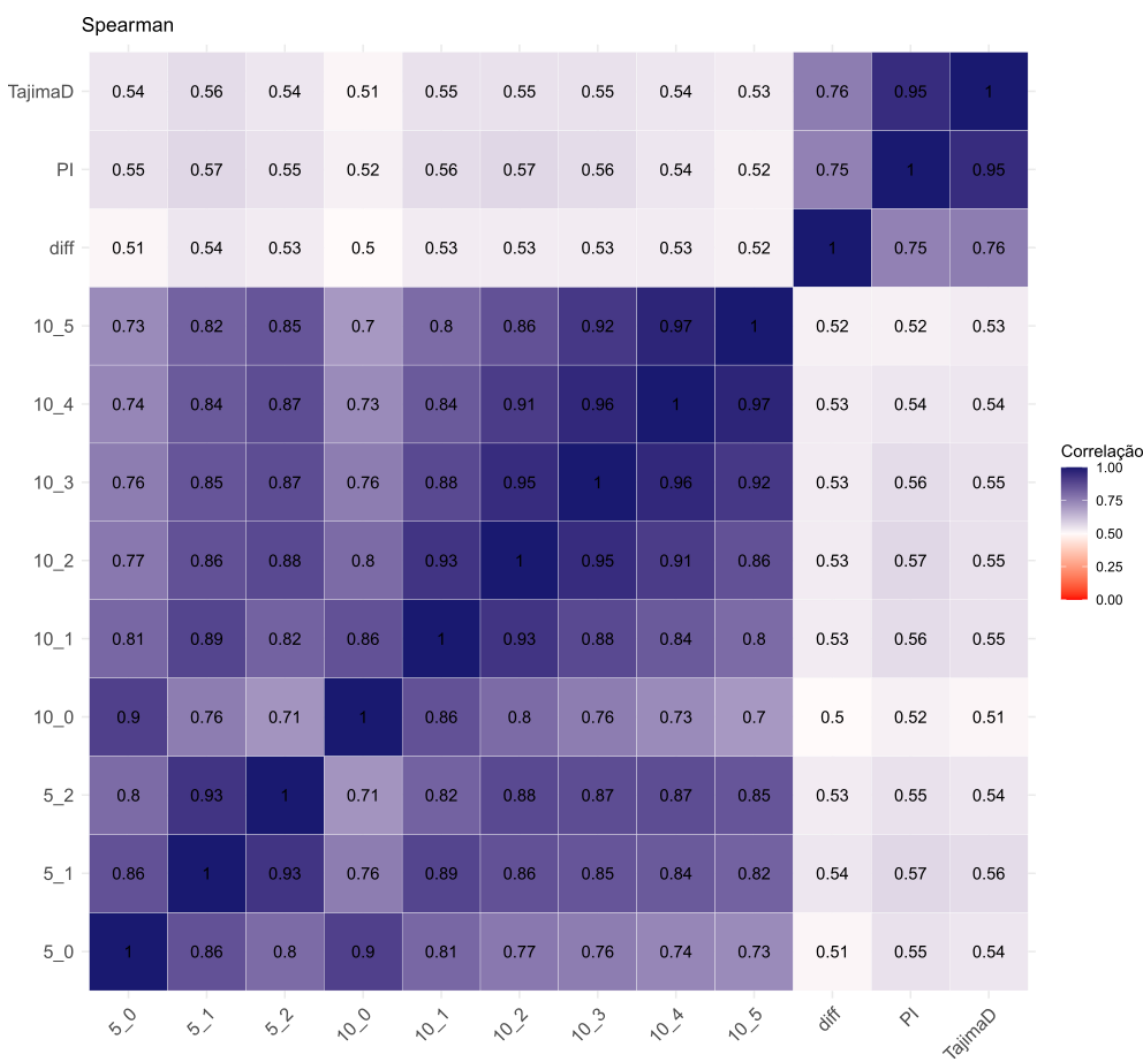
Figure 4 - Classificação das regiões ricas em heterozigose (HER), por cromossomo, de acordo com o comprimento, e a percentagem média de cobertura do cromossomo por HERs em cada um dos cenários com o número mínimo de marcadores iguais a 5 e 10



Correlações

A correlação de Spearman entre os cenários variou de moderadas a altas correlações, considerando a proporção de vezes que o SNP aparece dentro de um HER. A correlação foi classificada como moderada entre todos os cenários e as estimativas de diversidade populacional (diferencial calculado pela heterozigose esperada e observada, a diversidade de nucleotídeos e o teste D de Tajima). As correlações gerais de Spearman são apresentadas na Figura 5.

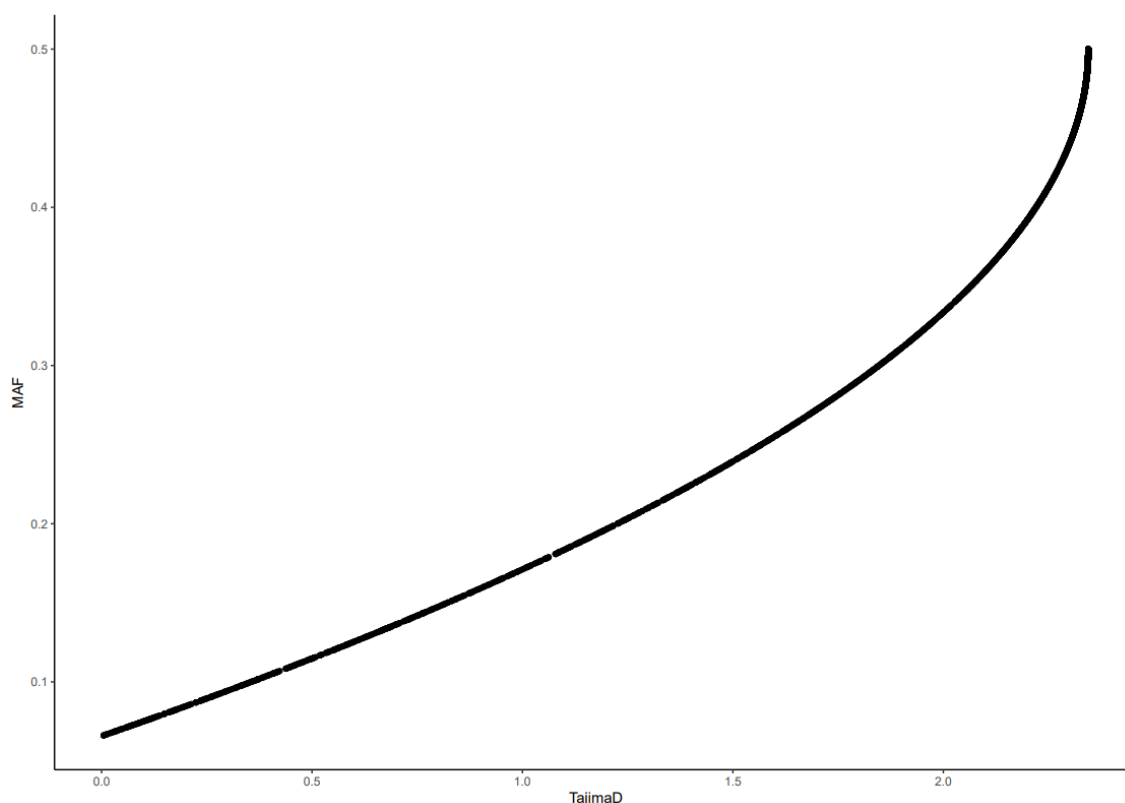
Figura 5 - Correlação de Spearman entre os cenários com o número mínimo de marcadores iguais a 5 e 10, diferencial baseado na heterozigose observada e esperada (diff), diversidade de nucleotídeos (PI) e teste D de Tajima



Estatística D de Tajima

O teste D de Tajima foi utilizado para observar marcadores em balanço de seleção (valores positivos). No total, 10.444.609 marcadores (22,04%) apresentaram valores positivos D do teste Tajima, sendo que a frequência do alelo menor (MAF) de tais marcadores em contraste com os valores de Tajima podem ser acessados na Figura 6.

Figure 6 - Gráfico de dispersão da frequência do alelo menor (MAF) para cada marcador e estatística D de Tajima



Ilhas de regiões enriquecidas em heterozigotos, desequilíbrio de ligação e análises funcionais

As ilhas de HER sobrepostas são apresentadas na Figura 7. Quarenta e cinco ilhas com marcadores em altos valores D de Tajima e alto diferencial foram encontradas em todos os cenários avaliados. Essas ilhas estavam presentes em 23 cromossomos, contendo 3.440 marcadores. A porcentagem que cada ilha aparece em cada um dos cenários e o desequilíbrio de ligação médio são apresentados na Tabela 2.

Figure 7 - Distribuição das ilhas de regiões ricas em heterozigose através do genoma da raça Holandesa

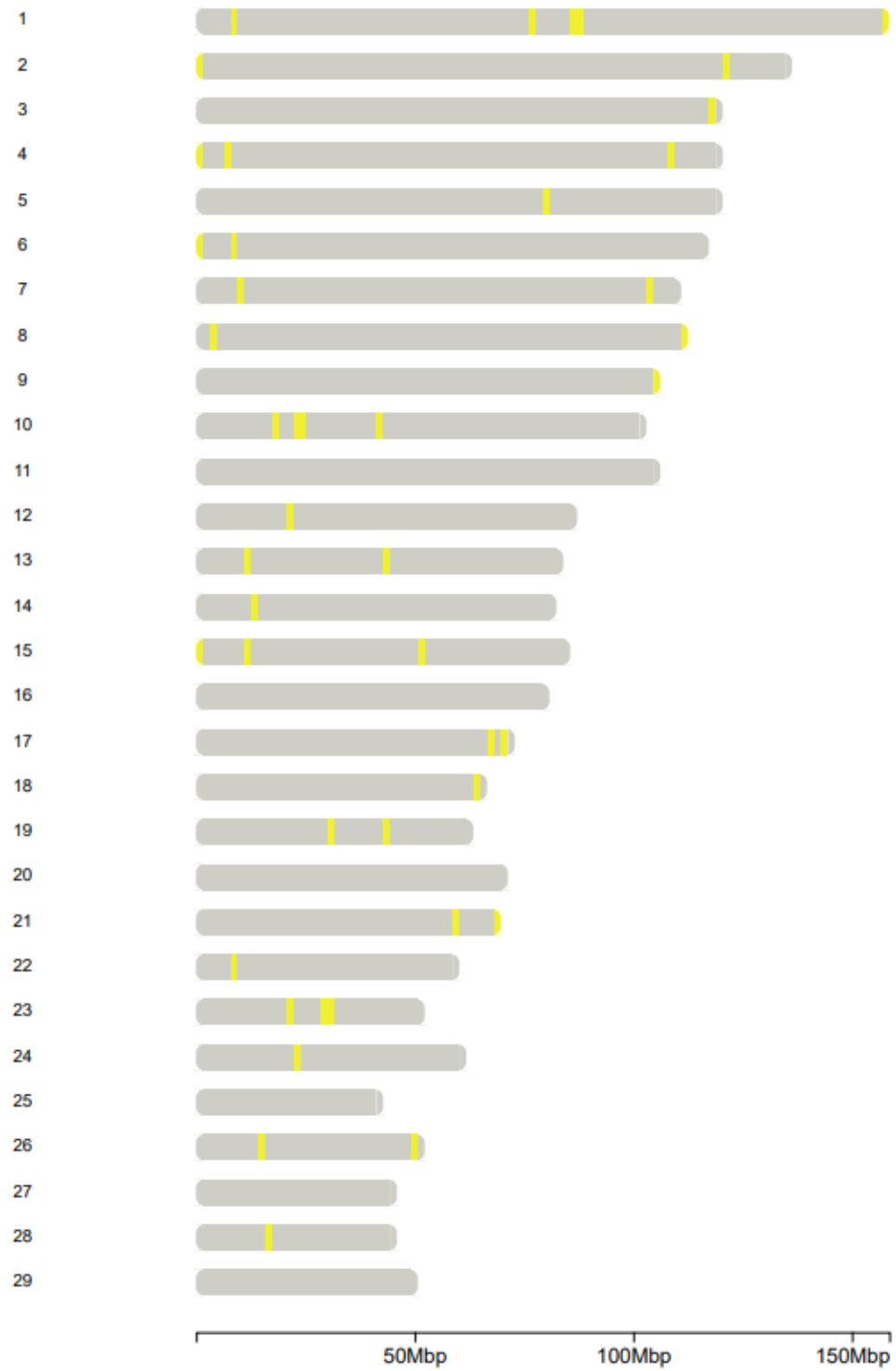


Tabela 2 - Ilhas de regiões rica em heterozigose, média do desequilíbrio de ligação (R2) dos marcadores dentro das ilhas e porcentagem em que que essas aparecem em cada cenário.

CRM	COMEÇO	FIM	NSNP	R2	5_0	5_1	5_2	10_0	10_1	10_2	10_3	10_4	10_5
1	8.947.612	8.984.143	54	0,010	10,14	35,97	61,94	10,11	32,86	58,74	77,79	88,23	94,93
1	77.541.594	77.574.520	114	0,143	24,20	48,00	67,19	17,75	41,52	64,06	74,97	78,78	84,88
1	86.455.167	86.476.612	40	0,062	10,00	10,00	14,77	10,00	10,00	12,34	23,67	32,36	32,87
1	88.572.745	88.595.480	149	0,021	10,00	10,00	10,00	10,00	10,00	10,00	12,92	19,51	25,28
1	158.026.308	158.204.765	2064	0,173	46,31	75,37	87,27	18,18	45,38	70,30	85,03	91,26	93,40
2	11.729	23.106	13	0,001	10,00	10,95	26,02	10,00	10,95	26,02	43,45	55,63	63,23
2	121.358.897	121.376.671	23	0,010	33,05	56,62	70,09	17,84	41,76	62,04	73,44	76,09	75,48
3	119.398.854	119.418.253	22	0,004	34,63	56,51	71,32	24,38	49,64	69,09	80,00	78,71	76,62
4	114.921	131.012	29	0,001	11,17	40,27	70,22	11,17	40,27	70,22	88,15	94,99	98,03
4	7.748.093	7.786.109	61	0,001	39,61	68,96	83,21	27,85	55,59	73,49	83,41	88,04	92,01
4	108.267.668	108.297.255	35	0,004	28,34	57,05	64,85	21,52	50,88	63,06	72,90	82,38	90,95
5	79.520.323	79.544.963	23	0,002	39,11	71,75	89,56	28,25	57,91	80,25	91,35	96,19	98,20
6	39.744	67.663	32	0,003	32,00	46,50	69,91	17,76	31,63	60,84	84,15	95,38	95,93
6	8.997.519	9.019.722	29	0,014	19,78	38,10	56,98	19,57	37,63	56,89	71,05	83,75	89,99
7	9.986.201	9.996.885	10	0,133	53,60	84,25	91,97	10,22	28,68	48,38	69,13	81,96	90,93
7	104.258.582	104.281.037	14	0,001	49,34	79,67	92,09	24,80	55,07	80,54	92,39	95,66	97,55
8	4.462.538	4.479.090	20	0,001	38,93	60,33	70,98	19,52	44,73	63,41	78,73	90,41	96,83
8	111.993.409	112.004.112	16	0,001	14,86	39,06	66,67	14,86	37,75	64,39	85,79	96,15	88,00
9	104.119.378	104.141.024	13	0,002	29,58	63,91	84,23	10,32	35,01	64,23	84,81	95,70	98,46
10	18.852.596	18.940.485	99	0,101	23,20	63,54	75,32	14,38	49,14	75,32	89,89	96,52	98,24
10	23.775.405	24.071.948	47	0,009	67,00	87,97	88,53	42,16	72,19	88,53	94,83	96,73	97,42
10	24.100.254	24.459.318	44	0,001	68,02	88,26	88,46	43,29	72,75	88,46	94,40	96,17	96,98
10	42.169.769	42.201.656	32	0,002	29,98	57,85	74,20	20,15	50,29	74,20	79,39	74,09	76,30

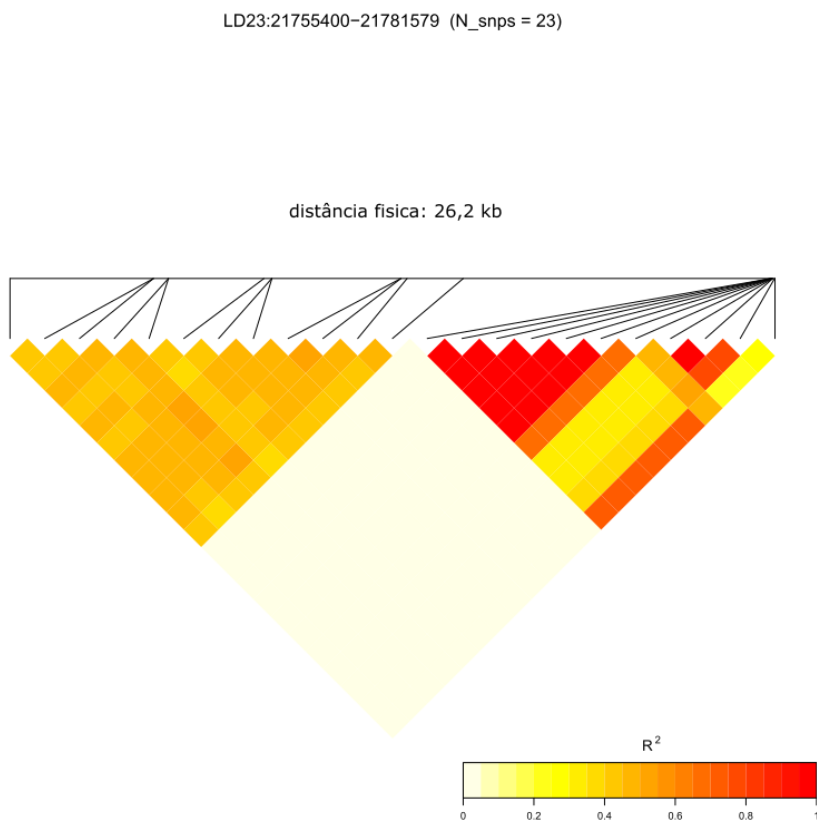
12	21.476.167	21.501.639	42	0,006	25,52	49,65	66,89	21,92	45,33	63,26	77,92	84,18	89,73
13	11.310.352	11.334.273	40	0,028	22,71	44,06	58,87	21,32	42,72	58,24	70,72	77,30	81,55
13	43.331.311	43.358.950	22	0,003	15,59	51,88	81,02	10,29	37,05	67,87	86,79	95,53	96,96
14	13.489.163	13.506.285	13	0,001	47,68	74,43	90,80	21,87	56,03	81,83	81,15	89,10	95,45
15	7.225	21.935	40	0,002	13,54	40,19	65,25	13,05	39,08	64,56	79,05	72,18	63,03
15	12.265.252	12.276.097	11	0,002	33,28	68,21	88,58	21,58	57,92	82,08	93,63	66,95	77,78
15	51.457.087	51.470.355	15	0,002	10,00	30,59	54,46	10,00	28,37	52,39	65,55	76,03	81,23
17	68.057.715	68.070.742	11	0,001	32,07	52,70	71,96	16,49	39,07	63,25	81,00	92,69	88,78
17	71.143.576	71.169.181	36	0,001	36,07	59,83	78,10	27,79	56,11	77,38	90,09	95,91	97,75
18	63.626.958	63.643.160	17	0,001	22,49	38,40	58,38	22,49	38,40	58,38	83,92	84,09	90,50
19	31.258.759	31.290.592	27	0,002	10,00	12,64	21,25	10,00	10,66	20,16	33,49	48,13	61,26
19	43.263.869	43.277.164	10	0,001	19,34	53,57	79,55	10,00	21,47	56,75	81,73	91,62	94,29
21	58.941.106	58.962.580	11	0,007	20,53	49,73	68,81	10,00	30,53	55,44	76,78	87,90	91,34
21	69.831.448	69.844.336	18	0,039	29,53	63,65	84,69	18,87	51,74	78,87	91,47	91,19	93,04
22	8.699.801	8.723.909	18	0,001	12,38	40,16	65,00	10,00	25,38	53,58	76,12	85,55	89,24
23	21.755.400	21.781.579	23	0,191	31,61	54,95	75,63	25,04	40,82	63,21	86,17	94,46	95,79
23	30.203.633	30.220.658	20	0,001	36,33	73,85	89,66	21,46	57,35	81,93	91,83	94,93	96,25
23	30.303.856	30.326.074	25	0,001	19,08	49,33	71,44	13,90	36,24	62,50	81,51	89,13	93,70
24	22.868.170	22.890.804	36	0,035	26,95	47,28	63,89	18,59	38,13	59,57	68,92	80,36	85,13
26	15.215.724	15.230.752	13	0,001	32,61	43,57	84,13	15,28	43,57	71,63	88,80	95,25	97,08
26	50.115.341	50.125.621	11	0,015	10,00	24,67	55,64	10,00	24,67	53,02	78,12	83,11	85,62
28	16.525.956	16.567.307	16	0,001	44,52	66,83	79,83	28,34	55,69	76,22	89,45	93,82	95,44

CRM: cromossomo; COMEÇO: região em pares de bases onde começa a ilha; FIM: região em pares de bases onde termina a ilha; NSPN: número de marcadores dentro da região rica em heterozigose; 5_0: cenário número mínimo de marcadores igual a 5 e 0 homozigoto permitido dentro de um HER; 5_1: cenário número mínimo de marcadores igual a 5 e 1 homozigoto permitido dentro de um HER; 5_2: cenário número mínimo de marcadores igual a 5 e 2 homozigotos permitidos dentro de um HER; 10_0: cenário número mínimo de marcadores igual a 10 e 0 homozigoto permitido dentro de um HER; 10_1: cenário número mínimo de marcadores igual a 10 e 1 homozigoto permitido dentro de um HER; 10_2: cenário número mínimo de marcadores igual a 10 e 2 homozigotos permitidos dentro de um HER; 10_3: cenário número mínimo de marcadores igual a 10 e 3 homozigotos permitidos dentro de um HER; 10_4: cenário número mínimo de marcadores igual a 10 e 4 homozigotos permitidos dentro de um HER; 10_5: cenário número mínimo de marcadores igual a 10 e 5 homozigotos permitidos dentro de um HER;

A maior ilha HER foi identificada no cromossomo 10, na região entre 24.100.254 pb a 24.459.318 pb, com comprimento de 359,06 Kb. A menor ilha HER foi identificada no cromossomo 26, nas regiões entre 50.115.341 pb a 50.125.621 pb. O tamanho médio das ilhas foi de 40,06 Kb ($\pm 68,37$ Kb).

O nível de desequilíbrio de ligação foi geralmente pequeno em todas as ilhas HER avaliadas, variando de 0,001 a 0,191. Uma ilha em particular mostra parte dos marcadores com forte desequilíbrio de ligação, conforme mostrado no mapa de calor da Figura 8. Essa ilha está situada no cromossomo 23 na posição 21.755.400 pb a 21.781.579 pb. Esta região é responsável pela codificação do gene *ENSBTAG00000054235*; um gene ligado à infecção por paratuberculose em bovinos (CANIVE *et al.*, 2021).

Figure 8 - Desequilíbrio de ligação do mapa de calor para a ilha localizada no cromossomo 23



A Tabela 3 apresenta a anotação dos genes para as regiões das ilhas HER. O significativo ($p < 0,05$) processo biológico, função molecular, componentes celulares e vias para os genes encontrados nas ilhas HER são apresentados na Tabela 4.

Tabela 3 - Anotação dos genes para as Ilhas de regiões ricas em heterozigose

CHR	COMEÇO	FIM	comprimento (Kb)	ID do gene	nome do gene	biotipo genético
1	77.541.594	77.574.520	32.926	ENSBTAG00000015460	TP63	Codificação de proteínas
1	8.947.612	8.984.143	36.531	ENSBTAG00000052332	H4C3	Codificação de proteínas
1	158.026.308	158.204.765	178.457	ENSBTAG00000053325		Codificação de proteínas
1	158.026.308	158.204.765	178.457	ENSBTAG00000049601	OR2B28	Codificação de proteínas
1	158.026.308	158.204.765	178.457	ENSBTAG00000037965	ZSCAN23	Codificação de proteínas
1	158.026.308	158.204.765	178.457	ENSBTAG00000050787	5S_rRNA	rRNA
4	7.748.093	7.786.109	38.016	ENSBTAG00000050410		pseudogene
6	39.744	67.663	27.919	ENSBTAG00000048454	U6	snRNA
8	111.993.409	112.004.112	10.703	ENSBTAG00000003540	MYT1L	Codificação de proteínas
10	18.852.596	18.940.485	87.889	ENSBTAG00000007433	MYO9A	Codificação de proteínas
10	18.852.596	18.940.485	87.889	ENSBTAG00000052861	bta-mir-2285dh	miRNA
12	21.476.167	21.501.639	25.472	ENSBTAG00000019134	NEK5	Codificação de proteínas
13	11.310.352	11.334.273	23.921	ENSBTAG00000051963		Codificação de proteínas
17	71.143.576	71.169.181	25.605	ENSBTAG00000017305	IG_C_gene	Codificação de proteínas
19	43.263.869	43.277.164	13.295	ENSBTAG00000049347		Codificação de proteínas
19	43.263.869	43.277.164	13.295	ENSBTAG00000049357	U2	snRNA
19	43.263.869	43.277.164	13.295	ENSBTAG00000045620	U2	snRNA
19	31.258.759	31.290.592	31.833	ENSBTAG00000021938	ARHGAP44	Codificação de proteínas
21	58.941.106	58.962.580	21.474	ENSBTAG00000039808	SERPINA6	Codificação de proteínas
23	21.755.400	21.781.579	26.179	ENSBTAG00000054235		Codificação de proteínas
23	30.203.633	30.220.658	17.025	ENSBTAG00000051232		Codificação de

23	30.303.856	30.326.074	22.218	ENSBTAG00000051232		proteínas Codificação de proteínas
23	30.203.633	30.220.658	17.025	ENSBTAG00000051628		Codificação de proteínas
23	30.303.856	30.326.074	22.218	ENSBTAG00000008943	ZSCAN12	Codificação de proteínas
24	22.868.170	22.890.804	22.634	ENSBTAG00000010299	NOL4	Codificação de proteínas
26	15.215.724	15.230.752	15.028	ENSBTAG00000055018		pseudogene
28	16.525.956	16.567.307	41.351	ENSBTAG00000010109	CDK1	Codificação de proteínas
28	16.525.956	16.567.307	41.351	ENSBTAG00000035206		Codificação de proteínas

Tabela 4 - Termos significativos ($p < 0,05$) da Ontologia Gênica (GO) e vias para os genes localizados em ilhas de regiões ricas com heterozigose.

geneSet	Descrição	valor p	ID do gene
Processo Biológico			
GO:0051301	divisão celular	0,002478152	ENSBTAG00000015460; ENSBTAG00000010109
GO:0030855	diferenciação de células epiteliais	0,003220027	ENSBTAG00000015460; ENSBTAG00000010109
GO:0030162	regulação da proteólise	0,005199694	ENSBTAG00000015460; ENSBTAG00000039808
GO:0006325	organização da cromatina	0,005583408	ENSBTAG00000015460; ENSBTAG00000010109
GO:0006974	resposta celular ao estímulo de dano ao DNA	0,006067112	ENSBTAG00000015460; ENSBTAG00000010109
GO:0007164	estabelecimento da polaridade do tecido	0,006975946	ENSBTAG00000015460
GO:0048483	desenvolvimento do sistema nervoso autônomo	0,007638625	ENSBTAG00000015460
GO:1904888	desenvolvimento do sistema esquelético craniano	0,009624892	ENSBTAG00000015460
GO:0030104	homeostase da água	0,010617031	ENSBTAG00000015460
GO:0042303	ciclo de muda	0,016885194	ENSBTAG00000015460
GO:0055123	desenvolvimento do sistema digestivo	0,020173567	ENSBTAG00000015460
GO:0007498	desenvolvimento do mesoderma	0,020830361	ENSBTAG00000015460
GO:0018210	modificação peptidil-treonina	0,020830361	ENSBTAG00000010109
GO:0007568	envelhecimento	0,024109926	ENSBTAG00000015460
GO:0072331	transdução de sinal pelo mediador da classe p53	0,024109926	ENSBTAG00000015460
GO:0098727	manutenção de número de celular	0,025092365	ENSBTAG00000015460
GO:0048736	desenvolvimento de apêndice	0,028362402	ENSBTAG00000015460
GO:0007219	Via de sinalização de entalhe	0,033253733	ENSBTAG00000015460
GO:0048863	diferenciação de células-tronco	0,03455531	ENSBTAG00000015460
GO:0001763	morfogênese de uma estrutura ramificada	0,034880521	ENSBTAG00000015460
GO:0043588	desenvolvimento da pele	0,037804139	ENSBTAG00000015460
GO:0030522	via de sinalização do receptor intracelular	0,040073978	ENSBTAG00000015460
GO:0070997	morte de neurônio	0,042663705	ENSBTAG00000015460
GO:0008544	desenvolvimento da epiderme	0,043310409	ENSBTAG00000015460
GO:0008202	processo metabólico de esteroides	0,046539556	ENSBTAG00000039808
Função Molecular			

GO:0003682	ligação da cromatina	0,003386756	ENSBTAG00000015460; ENSBTAG00000010109
GO:0002039	ligação p53	0,012013386	ENSBTAG00000015460
GO:0003684	ligação de DNA danificada	0,015001623	ENSBTAG00000015460
GO:0031072	ligação de proteína de choque térmico	0,024301018	ENSBTAG00000010109
GO:0061134	atividade reguladora da peptidase	0,049663249	ENSBTAG00000039808
Componentes Celulares			
GO:0030496	meio do corpo	0,046970942	ENSBTAG00000010109
Vias			
bta04115	via de sinalização p53	0,018444927	ENSBTAG00000010109
bta04914	Maturação oocitária mediada por progesterona	0,021065901	ENSBTAG00000010109
bta04540	Junção gap	0,021542065	ENSBTAG00000010109
bta04114	meiose oocitária	0,027484344	ENSBTAG00000010109
bta04110	Ciclo de célula	0,029382051	ENSBTAG00000010109
bta04218	senescência celular	0,039550472	ENSBTAG00000010109
bta05168	Infecção por herpes simples	0,047082919	ENSBTAG00000010109

Para as análises de QTL, foram identificadas 31 regiões relacionadas a 11 características. Destes, 40% estão relacionados ao leite, 30% à produção, 13,33% ao exterior, 10% à saúde e 6,67% à reprodução. Os resultados do enriquecimento de QTL são apresentados na Tabela 5.

Tabela 5 - QTLs (quantitative trait loci) significativo ($p < 0,05$) para os genes localizados em ilhas de regiões ricas em heterozigose.

QTL	CHR	N_QTLs	pvalor	adj.pval	tipo de QTL
Peso corporal metabólico	1	8	3.61241E-12	3.97365E-11	Produção
Porcentagem de kappa-caseína do leite	1	8	0,000126131	0,000693719	Leite
Porcentagem de kappa-caseína não glicosilada do leite	1	4	0,03090028	0,037767008	Leite
Mastite clínica	6	2	0,001357478	0,003733064	Saúde
Fenda do úbere	6	1	0,00704102	0,012908537	Exterior
Profundidade do úbere	6	1	0,014912737	0,020505013	Exterior
Altura do úbere	6	1	0,008794805	0,013820407	Exterior
Estrutura do úbere	6	1	0,000294204	0,001078749	Exterior
suscetibilidade a <i>M. paratuberculosis</i>	10	1	0,006578947	0,012908537	Saúde

Comparação entre raças

O número de ilhas HER encontradas para cada raça, o número de ilhas em comum com a raça Holandesa, as regiões específicas da raça e da raça Holandesa e o conjunto de genes para a região específica da raça Holandesa são apresentados na Tabela 6.

Tabela 6. Comparação de ilhas de regiões ricas em heterozigose entre animais das raças Holandesa, Angus (ANG), Jersey (JER) e Norwegian Red Cattle (RDC).

	n ilha	% em comum	n ilha específica	Ilha específica da raça	Ilha específica de Holstein
JER	68	94,28	2	chr4: 7.754.341-7.779.125 chr16: 4.805.929-4.981.727	chr11: 29.406-32.150 chr16: 48.405.929-49.810.727 chr17: 30.241.885-30.254.164 chr24: 46.608.950-46.619.501
ANG	72	95,72	4	chr10: 8.947.684-8.979.955 chr19: 566.224.233-566.316.890 chr22: 86.844.336-87.184.460 chr29: 48.220.837-48.234.421	chr1: 8.947.673-8.982.695 chr19: 56.624.233-56.631.689 chr22: 8.699.807-8.723.909
RDC	72	97,14	2	chr4: 7.756.346-7.779.125 chr18: 5.717.899-5.727.914	chr10: 10.036.762-10.051.565 chr14: 49.089.801-55.444.216

A raça com maior sobreposição de ilhas HER foi a NRC, com 97,14% das mesmas ilhas encontradas em bovinos da raça Holandesa. A raça Jersey apresenta o menor percentual de ilhas em comum com a raça Holandesa (94,28%), seguida pela raça Angus (95,72%).

Discussão

Impacto dos parâmetros utilizados na identificação de HERs

Primeiramente, avaliou-se o impacto de diferentes parâmetros na identificação de HER e ilhas de HER. Conforme apresentado na Figura 2, três parâmetros (número mínimo de SNPs, número máximo de homozigotos permitidos e comprimento mínimo de uma HER) influenciaram significativamente na detecção de HERs. O tamanho mínimo de 10 Kb foi considerado o melhor parâmetro para o comprimento mínimo de uma HER, baseado no maior número de HER identificadas usando este parâmetro. Isso corrobora a hipótese de que os HERs são pequenas regiões espalhadas pelo genoma que ocorrem aleatoriamente como consequência de diferentes processos que atuam na população (por exemplo, seleção, mutação, migração) (MARRAS *et al.*, 2018). Essa descoberta afeta diretamente a escolha dos painéis SNP a serem usados para essas análises, pois a detecção precisa de HERs mais curtos em todo o genoma requer painéis SNP mais densos (CEBALLOS; HAZELHURST; RAMSAY, 2018).

O alto número de marcadores no conjunto de dados WGS permitiu a identificação de pequenas HER em todo o genoma. Também, para estudos em regiões ricas em heterozigose, uma boa alternativa uma vez que, como apresentado neste trabalho, o tamanho do comprimento das HERs, na maioria dos casos, são pequenos trechos do genoma (Tabela 1 e Figura 4). Outro ponto a ser destacado, conforme demonstrado por Ceballos, Hazelhurst e Ramsay (2018) é que o número de marcadores heterozigotos presentes em WGS é maior do que nos painéis SNP. Portanto, as chances de capturar verdadeiras regiões ricas em heterozigose são maiores usando dados WGS do que painéis comerciais.

Painéis comerciais geralmente são projetados com base em SNPs que ocorrem em mais de 1% da população utilizada para seu design (YOUSEFI *et al.*, 2018). Esses polimorfismos geralmente têm frequência alélica mais alta e estão localizados perto de

QTLs associados a características importantes (ALILLOO *et al.*, 2018). Na análise de varredura do genoma com tais painéis SNP, assume-se que os marcadores ausentes entre dois SNP heterozigotos consecutivos são heterozigotos. Portanto, pode-se criar viés de apuração nas informações fornecidas, principalmente no tamanho do comprimento de HER, pois pequenas HERs podem não ser detectadas ou inflar o tamanho do comprimento destas, ou ainda identificar HERs falso-positivas.

Conforme apresentado na Figura 3, um número substancialmente inflado de HERs foi obtido ao usar painéis SNP em comparação com WGS. Curiosamente, entre os painéis SNP, a maior densidade mostra uma maior taxa de detecção. Isso pode estar relacionado às particularidades do painel, uma vez que os painéis GGP (100K e HD) possuem um número maior de marcadores concentrados em regiões de telômeros do que o BovineSNP (50K) (TORO OSPINA *et al.*, 2021) e este possui vários marcadores distribuídos uniformemente pelo genoma bovino (GERSHONI *et al.*, 2022). Como as regiões teloméricas têm uma taxa de recombinação mais alta (BAILEY, 2004), o que poderia potencialmente aumentar a variação de haplótipos e a heterozigose (JARAMILLO-CORREA; VERDÚ; GONZÁLEZ-MARTÍNEZ, 2010), as chances de marcadores heterozigotos aparecerem e estarem em um HER são maiores. Outro ponto que pode acontecer são as HERs maiores identificados em um painel menos denso, na verdade serem múltiplas HERs lado a lado em painéis mais densos. Como as chances de detectar HERs falso-positivas ou falso-negativas em painéis SNP são maiores, análises adicionais, como diversidade de nucleotídeos e/ou estatística D de Tajima, são necessárias para confirmar a veracidade da detecção.

Em relação ao número mínimo de marcadores que compõem uma HER e ao número máximo de marcadores homozigotos permitidos dentro de uma HER, esses parâmetros parecem mais relacionados ao quão menos ou mais conservadoras são as análises. Ambos os parâmetros funcionam na direção oposta e têm um impacto maior na detecção de HERs. Pode-se afirmar que esses parâmetros precisam considerar informações prévias sobre a população, histórico de seleção, densidade de painéis e qualidade da genotipagem. Em populações que sofreram um longo processo de seleção direta, pequena diversidade entre os indivíduos e pequeno tamanho efetivo da população, espera-se que a autozigosidade nessas populações seja alta (PURFIELD *et al.*, 2012; REBELATO; CAETANO, 2018). Em resposta, parâmetros mais flexíveis poderiam ser

usados para detectar HERs nessa população. Em relação aos marcadores homozigotos permitidos dentro das HERs, esta é uma forma de considerar possíveis erros que podem quebrar erroneamente uma HER longa, afetado pela qualidade da genotipagem (CEBALLOS; HAZELHURST; RAMSAY, 2018).

Ilhas da região ricas em heterozigose

Regiões ricas em heterozigose podem aparecer em todo o genoma, mas quando estão concentradas em uma região específica podem indicar um padrão de eventos de seleção na população (NANDOLO *et al.*, 2018). Quarenta e cinco ilhas foram encontradas em todos os cenários, e todas essas ilhas com valor D de Tajima positivo indicam uma seleção para alelos heterozigotos (balanço de seleção). Este teste D de Tajima considera a diversidade de nucleotídeos (π) e uma expectativa para π com base na média de marcadores mais o número total de mutações (POLLEY; CONWAY, 2001). Como resultado, se os valores forem positivos, o marcador está sob balanço de seleção (TAJIMA, 1989). Conforme apresentado na Figura 6, o valor mais alto de Tajima foi acompanhado por uma frequência alélica de 50% e, aumentando o valor Tajima, aumentasse a frequência alélica na população. Embora nenhum dos cenários tenha apresentado uma correlação de Spearman alta com a estimativa D de Tajima (Figura 5), esta ainda pode ser usado como uma análise complementar nos estudos de HER para endossá-las.

Com relação ao desequilíbrio de ligação entre os marcadores dentro de uma HER, a maioria dos marcadores apresentou baixo desequilíbrio de ligação. Comparando nossos resultados com Qanbari e Wittenburg (2020), parte das ilhas de HER encontradas estão localizadas em regiões definidas como intervalos de hotspot de recombinação, onde a taxa de recombinação excede 2,5 desvios padrão da taxa de recombinação média de todo o genoma. Essa recombinação quebra e recombina diferentes combinações de alelos. Nesse processo, novas HERs são criados principalmente em HERs próximos às regiões dos telômeros. Mas, este não parece ser o único processo que propicia o surgimento das ilhas de HERs. Vinte e oito genes e 11 QTLs foram encontrados relacionados as HERs estabelecidas (Tabelas 3 e 5) e podem estar, de alguma forma, associados a uma vantagem do alelo heterozigoto. Alguns dos genes encontrados nessas regiões estão relacionados à resposta imune, como esperado, uma vez que níveis mais elevados de heterozigose podem

levar à proteção contra doenças (SAMUELS *et al.*, 2016) e, conseqüentemente, maior chance de adaptação a estresses. Por outro lado, descobriu-se que muitos genes estão relacionados a características como fertilidade e produção. Por exemplo, a *TP63* ou proteína tumoral p63, que tem a função de se ligar à certas regiões do DNA e controlar as atividades de determinados genes, e também previamente associada à puberdade em bovinos (DIAS *et al.*, 2017). O gene *ZSCAN23*, envolvido na regulação da transcrição da RNA polimerase III e associado à fertilidade masculina (MEKATA; YAMAMOTO, 2022). O gene *ARHGAP44*, que afeta a polaridade celular, o tráfego vesicular, o ciclo celular e a transcrição, também influencia a fertilidade de vacas (KLABNIK *et al.*, 2022). Os genes *TP63* e *CDK1* foram relacionados a um maior número de processos e vias, como pode ser observado na Tabela 4. Tais genes impactam mecanismos que controlam processos como a multiplicação de células e/ou transcrição dos genes, e como participantes de múltiplas funções no organismo, foram envolvidos em múltiplos termos ontológicos.

Outras características relacionadas aos genes encontrados nas ilhas de HERs são os ácidos graxos presente no leite e o percentual de gordura do leite (por exemplo, *MYTIL*, (IBEAGHA-AWEMU *et al.*, 2016)); conversão alimentar residual (por exemplo, *MAYO9A* (SERÃO *et al.*, 2013); temperatura de vitrificação de oócitos bovinos (por exemplo, *NEK5* (ZHANG *et al.*, 2020); e, homeostase hormonal e níveis de progesterona (por exemplo, *SERPINA6* (WATHES *et al.*, 2021)). Esses achados sugerem que genótipos heterozigotos podem ser vantajosos nessa área, uma vez que essas regiões foram conservadas ao longo de gerações. Mais estudos comparando a vantagem de alelos heterozigotos para tais características precisam ser desenvolvidos para confirmar esta hipótese.

Com relação à comparação entre as raças, pelo menos 94,28% das ilhas de HER estão presentes nas diferentes raças avaliadas. Parece que as HERs são compartilhadas entre as raças mesmo quando o processo de seleção foi diferente ao longo do tempo. Isso pode estar relacionado a algum processo evolutivo ligado à adaptação dos animais, uma vez que os HERs estão relacionados a tais processos (SANTOS *et al.*, 2021b; WILLIAMS *et al.*, 2016).

Conclusões

A identificação de HER depende dos parâmetros usados para medir as regiões em níveis de heterozigose. O comprimento mínimo de 10Kb apresentou a maior detecção de HERs, confirmando que os HERs são pequenas regiões espalhadas por todo o genoma. O número mínimo de marcadores que constituem um HER e o número máximo de homozigotos permitidos dentro do HER não apresentaram uma medida destaque entre os cenários, estando tais parâmetros mais ligados à estrutura populacional e à qualidade da genotipagem. Quarenta e cinco ilhas foram identificadas em todos os cenários avaliados com altos valores D de Tajima indicando que tais regiões estão em balanço de seleção. Em geral, essas regiões apresentam um pequeno desequilíbrio de ligação e estão relacionadas a características como fertilidade, produção e respostas imunes. Quanto à comparação entre as raças, existem muitos HERs em comum entre as raças avaliadas. Independentemente dos processos de seleção, que cada raça sofreu ao longo das gerações.

Referências

- ALILLOO, H.; MRODE, R.; OKEYO, AM; NI, G.; GODDARD, EU; GIBSON, JP A viabilidade do uso de painéis de marcadores de baixa densidade para imputação de genótipos e previsão genômica de gado leiteiro mestiço da África Oriental. **Journal of Dairy Science** , v. 101, n. 10, pág. 9108–9127, fora. 2018.
- BAES, CF; MAKANJUOLA, BO; MIGLIOR, F.; MARRAS, G.; HOWARD, JT; FLEMING, A.; MALTECCA, C. Revisão do Simpósio: A arquitetura genômica da endogamia: Como a homozigose afeta a saúde e o desempenho. **Journal of Dairy Science** , v. 102, n. 3, pág. 2807–2817, 1 mar. 2019. . Acesso em: 30 out. 2019.
- BAILEY, SM A recombinação frequente no DNA telomérico pode prolongar a vida proliferativa de células negativas para telomerase. **Nucleic Acids Research** , v. 32, n. 12, pág. 3743–3751, 7 jul. 2004. . Acesso em: 1 nov. 2022.
- BISCARINI, F.; COZZI, P.; GASPA, G.; MARRAS, G. **detectRUNS: um pacote R para detectar execuções de homozigose e heterozigosidade em genomas diplóides** .
- BISCARINI, F.; MASTRANGELO, S.; CATILLO, G.; SENCZUK, G.; CIAMPOLINI, R. Insights sobre Diversidade Genética, Corridas de Homozigosidade e Regiões Ricas em Heterozigosidade em Gado Semi-Selvagem Maremmana Usando Pedigree e Dados Genômicos. **Animais** , v. 10, n. 12, pág. 2285, 3 dez. 2020.
- BRITO, LF; KIJAS, JW; VENTURA, R. v.; SARGOLZAEI, M.; PORTO-NETO, LR; CÁNOVAS, A.; FENG, Z.; JAFARIKIA, M.; SCHENKEL, FS Diversidade genética e assinaturas de seleção em várias raças caprinas reveladas por marcadores SNP em todo o genoma. **BMC Genomics** , v. 18, n. 1, pág. 229, 14 dez. 2017. Disponível em: <<http://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-017-3610-0>>. Acesso em: 2 nov. 2019.
- CANIVE, M.; GONZÁLEZ-RECIO, O.; FERNÁNDEZ, A.; VÁZQUEZ, P.; BADIA-BRINGUÉ, G.; LAVÍN, JL; GARRIDO, JM; JUSTE, RA; ALONSO-HEARN, M. Identificação de loci associados à suscetibilidade a *Mycobacterium avium* subsp. paratuberculosis em gado Holandês usando combinações de testes diagnósticos e dados imputados da sequência do genoma completo. **PLOS UM** , v. 16, n. 8, pág. e0256091, 27 atrás. 2021. . Acesso em: 13 dez. 2022.
- CEBALLOS, FC; HAZELHURST, S.; RAMSAY, M. Avaliação de corridas de homozigose: uma comparação de SNP Array e dados de baixa cobertura de toda a

sequência do genoma. **BMC Genomics** , v. 19, n. 1, pág. 106, 30 dez. 2018. . Acesso em: 3 mar. 2021.

CEBALLOS, FC; JOSHI, PK; CLARK, DW; RAMSAY, M.; WILSON, JF Corridas de homozigose: janelas para a história da população e arquitetura de traços. **Nature Reviews Genetics** , v. 19, n. 4, pág. 220–234, 15 abr. 2018. . Acesso em: 29 out. 2019.

DANECEK, P.; AUTON, A.; ABECASIS, G.; ALBERS, CA; BANKS, E.; DEPRISTO, MA; HANDSAKER, RE; LUNTER, G.; MARTH, GT; SHERRY, ST; MCVEAN, G.; DURBIN, R. O formato de chamada variante e VCFtools.

Bioinformática , v. 27, n. 15 de 2011.

DIAS, MM; CÁNOVAS, A.; MANTILLA-ROJAS, C.; RILEY, DG; LUNA-NEVAREZ, P.; COLEMAN, SJ; SPEIDEL, SE; ENNS, RM; ISLAS-TREJO, A.; MEDRANO, JF; MOORE, SS; FORTES, SRA; NGUYEN, LT; VÊNUS, B.; DIAZ, ISDP; SOUZA, FRP; FONSECA, LFS; BALDI, F.; ALBUQUERQUE, LG; THOMAS, MG; OLIVEIRA, HN Detecção de SNP usando seqüências de RNA de genes candidatos associados à puberdade em bovinos. **Genética e Pesquisa Molecular** , v. 16, n. 1, pág. 16019522, 2017. . Acesso em: 8 nov. 2022.

FERENČAKOVIĆ, M.; SÖLKNER, J.; CURIK, I. Estimando a autozigosidade a partir de informações de alto rendimento: efeitos da densidade de SNP e erros de genotipagem. **Evolução da Seleção Genética** , v. 45, n. 1, pág. 42, 29 dez. 2013. . Acesso em: 13 abr. 2021.

FIJARCZYK, A.; BABIK, W. Detectando seleção balanceada em genomas: limites e perspectivas. **Ecologia Molecular** , v. 24, n. 14, pág. 3529–3545, jul. 2015. . Acesso em: 14 nov. 2022.

FONSECA, PAS; SUÁREZ-VEGA, A.; MARRAS, G.; CÁNOVAS, Á. GALLO: Um pacote R para anotação genômica e integração de múltiplas fontes de dados na pecuária para loci candidatos posicionais. **GigaScience** , v. 9, n. 12, pág. 1–9, 30 dez. 2020. . Acesso em: 2 mar. 2021.

FORUTAN, M.; ANSARI MAHYARI, S.; BAES, C.; MELZER, N.; SCHENKEL, FS; SARGOLZAEI, M. Endogamia e corridas de homozigose antes e depois da seleção genômica em bovinos da raça Holandesa Norte-Americana. **BMC Genomics** , v. 19, n. 1, pág. 98, 27 dez. 2018. . Acesso em: 14 fev. 2019.

GERSHONI, M.; SHIRAK, A.; RAZ, R.; SEROUSSI, E. Comparando BeadChip e

Genotipagem WGS: Chamada de Falha Não Técnica é Atribuível a Variação Adicional dentro da Sequência Alvo da Sonda. **Genes** , v. 13, n. 3, pág. 485, 9 mar. 2022. . Acesso em: 1 nov. 2022.

GROSSI, DA; JAFARIKIA, M.; BRITO, LF; BUZANSKAS, EU; SARGOLZAEI, M.; SCHENKEL, FS Diversidade genética, extensão do desequilíbrio de ligação e persistência da fase gamética em suínos canadenses. **BMC Genetics** , v. 18, n. 1, 2017. . Acesso em: 1 ago. 2021.

GURGUL, A.; SZMATOŁA, T.; TOPOLSKI, P.; JASIELCZUK, I.; ŻUKOWSKI, K.; BUGNO-PONIEWIERSKA, M. O uso de corridas de homozigose para estimativa de endogamia recente em bovinos da raça Holandesa. **Journal of Applied Genetics** , v. 57, n. 4, pág. 527–530, 23 nov. 2016. . Acesso em: 29 out. 2019.

HAYES, BJ; DAETWYLER, HD 1000 Projeto Genomas de Touros para Mapear Características Genéticas Simples e Complexas em Bovinos: Aplicações e Resultados. **Revisão Anual de Biociências Animais** , v. 7, n. 1, pág. 89–102, 15 fev. 2019. . Acesso em: 13 nov. 2022.

HOWARD, JT; TIEZZI, F.; HUANG, Y.; GRAY, KA; MALTECCA, C. Um método heurístico para identificar corridas de homozigose associadas a desempenho reduzido em rebanhos. **Journal of Animal Science** , v. 95, n. 10, pág. 4318–4332, 1 fora. 2017. . Acesso em: 13 dez. 2022.

IBEAGHA-AWEMU, EM; PETERS, SO; AKWANJI, KA; IMUMORIN, IG; ZHAO, X. A genotipagem ampla por sequenciamento e associação do genoma de alta densidade identifica SNPs comuns e de baixa frequência e novos genes candidatos que influenciam as características do leite de vaca. **Relatórios Científicos** , v. 6, n. 1, pág. 31109, 10 atrás. 2016. . Acesso em: 9 nov. 2022.

JARAMILLO-CORREA, JP; VERDÚ, M.; GONZÁLEZ-MARTÍNEZ, SC A contribuição da recombinação para a heterozigosidade difere entre linhagens evolutivas de plantas e formas de vida. **BMC Evolutionary Biology** , v. 10, n. 1, pág. 22, 25 dez. 2010. . Acesso em: 1 nov. 2022.

KLABNIK, JL; CHRISTENSON, LK; GUNWARDENA, SSA; POHLER, KG; RISPOLI, LA; PAYTON, RR; MOOREY, SE; NEAL SCHRICK, F.; EDWARDS, JL Aumentos induzidos pelo calor na temperatura corporal em vacas leiteiras em lactação: impacto no transcriptoma das células do cúmulo e da granulosa do folículo

periovulatório. **Journal of Animal Science** , v. 100, n. 7, 1 jul. 2022.

LLAURENS, V.; WHIBLEY, A.; JORON, M. Arquitetura genética e seleção balanceadora: a vida e a morte de variantes diferenciadas. **Ecologia Molecular** , v. 26, n. 9, pág. 2430–2448, 1 de maio de 2017. . Acesso em: 14 nov. 2022.

LOZADA-SOTO, EA; MALTECCA, C.; LU, D.; MILLER, S.; COLE, JB; TIEZZI, F. Tendências na diversidade genética e o efeito da endogamia em bovinos American Angus sob seleção genômica. **Evolução da Seleção Genética** , v. 53, n. 1, pág. 50, 16 dez. 2021. . Acesso em: 13 dez. 2022.

LOZADA-SOTO, EA; TIEZZI, F.; JIANG, J.; COLE, JB; VANRADEN, PM; MALTECCA, C. Caracterização genômica de autozigosidade e tendências recentes de endogamia em todas as principais raças de gado leiteiro dos EUA. **Journal of Dairy Science** , v. 105, n. 11, pág. 8956–8971, 1 nov. 2022. . Acesso em: 13 dez. 2022.

MARRAS, G.; GASPA, G.; SORBOLINI, S.; DIMAURO, C.; AJMONE-MARSAN, P.; VALENTINI, A.; WILLIAMS, JL; MACCIOTTA, NPP Análise de corridas de homozigose e sua relação com a endogamia em cinco raças bovinas criadas na Itália. **Genética Animal** , v. 46, n. 2, pág. 110–121, 1 abr. 2015. . Acesso em: 3 nov. 2019.

MARRAS, G.; MADEIRA, BJ; MAKANJUOLA, B.; MALCHIODI, F.; PEETERS, K.; AS, P. van; BAES, CF; BISCARINI, F.; TURKEYS, H. Caracterização de sequências de regiões ricas em homozigose e heterozigotidade em uma população comercial de perus (*Meleagris gallopavo*). Em: Proceedings of the 11th World Congress of Genetics Applied to Livestock Production, Szwaczkowski 2017., 2018, Auckland. [...]. Auckland: 2018. v. 11, p. 763–767.

MULIM, HA; BRITO, LF; BATISTA PINTO, LF; MOLETTA, JL; DA SILVA, LR; PEDROSA, VB Caracterização Genética e Genômica de uma Nova Raça Composta de Bovinos de Corte (Purunã) Desenvolvida para Produção em Sistemas a Pasto. **Frontiers in Genetics** , v. 13, 2022a.

MULIM, HA; BRITO, LF; PINTO, LFB; FERRAZ, JBS; GRIGOLETTO, L.; SILVA, MR; PEDROSA, VB Caracterização de sequências de homozigose, regiões enriquecidas em heterozigotidade e estrutura populacional em populações bovinas selecionadas para diferentes objetivos de reprodução. **BMC Genomics** , v. 23, n. 1, 2022b.

NANDOLO, W.; UTSUNOMIYA, YT; MÉSZÁROS, G.; WURZINGER, M.;

- KHAYADZADEH, N.; TORRECILHA, RBP; MULINDWA, HA; GONDWE, TN; WALDMANN, P.; FERENČAKOVIĆ, M.; GARCIA, JF; ROSEN, BD; BICKHART, D.; VAN TASSELL, CP; CURIK, I.; SÖLKNER, J. Identificação incorreta de sequências de ilhas de homozigose em bovinos causadas por interferência na variação do número de cópias ou grandes distâncias entre marcadores. **Evolução da Seleção Genética**, v. 50, n. 1, pág. 43, 22 dez. 2018. . Acesso em: 29 out. 2019.
- POLLEY, SD; CONWAY, DJ Seleção diversificada forte em domínios do *Plasmodium falciparum* Apical Membrane Antigen 1 Gene. **Genética**, v. 158, n. 4, pág. 1505–1512, 1 atrás. 2001. . Acesso em: 6 nov. 2022.
- PURCELL, S.; NEALE, B.; TODD-BROWN, K.; THOMAS, L.; FERREIRA, MAR; BENDER, D.; MALLER, J.; SKLAR, P.; DE BAKKER, PIW; DALY, MJ; SHAM, PC PLINK: Um conjunto de ferramentas para análises de associação de todo o genoma e de ligação com base na população. **The American Journal of Human Genetics**, v. 81, n. 3, pág. 559–575, conjunto. 2007. . Acesso em: 2 mar. 2019.
- PURFIELD, DC; BERRY, DP; MCPARLAND, S.; BRADLEY, DG Corridas de homozigose e histórico populacional em bovinos. **BMC Genetics**, v. 13, n. 1, pág. 70, 14 dez. 2012. . Acesso em: 17 fev. 2019.
- QANBARI, S.; WITTENBURG, D. Mapa de recombinação masculina do genoma autossômico em alemão Holstein. **Evolução da Seleção Genética**, v. 52, n. 1, pág. 73, 14 dez. 2020. . Acesso em: 6 nov. 2022.
- REBELATO, AB; CAETANO, AR Rodadas de homozigose para estimativa de autozigosidade e análise genômica em animais de produção. **Pesquisa Agropecuária Brasileira**, v. 53, n. 9, pág. 975–984, conjunto. 2018. . Acesso em: 24 fev. 2019.
- ROSEN, BD; BICKHART, DM; SCHNABEL, RD; KOREN, S.; ELSIK, CG; TSENG, E.; ROWAN, TN; BAIXO, WY; ZIMIN, A.; COULDREY, C.; HALL, R.; LI, W.; RHIE, A.; GHURYE, J.; MCKAY, SD; THIBAUD-NISSEN, F.; HOFFMAN, J.; MURDOCH, BM; SNELLING, WM; MCDANELD, TG; HAMMOND, JA; SCHWARTZ, JC; NANDOLO, W.; HAGEN, DE; DREISCHER, C.; SCHULTHEISS, SJ; SCHROEDER, SG; PHILLIPY, AM; COLE, JB; VAN TASSELL, CP; LIU, G.; SMITH, TPL; MEDRANO, JF Montagem de novo do genoma de referência bovino com sequenciamento de molécula única. **GigaScience**, v. 9, n. 3, pág. 1–9, 2020.
- RUAN, D.; YANG, J.; ZHUANG, Z.; DING, R.; HUANG, J.; QUAN, J.; GU, T.;

HONG, L.; ZHENG, E.; LI, Z.; CAI, G.; WANG, X.; WU, Z. Avaliação da Heterozigosidade e Análise Genômica de Regiões de Heterozigosidade em Duas Populações de Porcos Duroc. **Frontiers in Genetics** , v. 12, p. 2755, 27 jan. 2022. . Acesso em: 14 nov. 2022.

SAMUEL, DC; WANG, J.; YE, F.; HE, J.; LEVINSON, RT; SHENG, Q.; ZHAO, S.; CAPRA, JA; SHYR, Y.; ZHENG, W.; GUO, Y. Relação de heterozigosidade, uma medida genômica global robusta de autozigosidade e sua associação com altura e risco de doença. **Genética** , v. 204, n. 3, pág. 893–904, 1 nov. 2016. . Acesso em: 1 dez. 2020.

SANTOS, W.; SCHETTINI, G.; FONSECA, MG; PEREIRA, GL; CHARDULO, LA; NETO, O.; BALDASSINI, WA; OLIVEIRA, H.; CURI, R. Estimativa em escala fina de taxas de endogamia, séries de homozigose e níveis de heterozigosidade genômica na raça de cavalos Mangalarga Marchador. **Journal of Animal Breeding and Genetics** , v. 138, n. 2, pág. 161–173, 19 mar. 2021a. . Acesso em: 26 abr. 2021.

SANTOS, W.; SCHETTINI, G.; FONSECA, MG; PEREIRA, GL; CHARDULO, LA; NETO, O.; BALDASSINI, WA; OLIVEIRA, H.; CURI, R. Estimativa em escala fina de taxas de endogamia, séries de homozigose e níveis de heterozigosidade genômica na raça de cavalos Mangalarga Marchador. **Journal of Animal Breeding and Genetics** , v. 138, n. 2, pág. 161–173, 19 mar. 2021b. . Acesso em: 2 ago. 2021.

SELLI, A.; VENTURA, R.v; FONSECA, PAS; BUZANSKAS, EU; ANDRIETTA, LT; BALIEIRO, JCC; BRITO, LF Detecção e Visualização de Regiões Ricas em Heterozigosidade e Corridas de Homozigose em Populações Ovinos Mundiais. v. 11, pág. 2696, 2021. . Acesso em: 14 nov. 2022.

SERÃO, NV; GONZÁLEZ-PEÑA, D.; BEEVER, JE; FAULKNER, DB; SOUTHEY, BR; RODRIGUEZ-ZAS, SL Polimorfismos e haplótipos de nucleotídeo único associados à eficiência alimentar em bovinos de corte. **BMC Genetics** , v. 14, n. 1, pág. 94, 25 dez. 2013. . Acesso em: 8 nov. 2022.

SZMATOŁA, T.; GURGUL, A.; JASIELCZUK, I.; ZĄBEK, T.; ROPKA-MOLIK, K.; LITWIŃCZUK, Z.; BUGNO-PONIEWIERSKA, M. Uma análise abrangente de corridas de homozigose de onze raças de gado representando diferentes tipos de produção. **Animais** , v. 9, n. 12, pág. 1024, 25 nov. 2019. . Acesso em: 18 nov. 2021.

TAJIMA, F. Método estatístico para testar a hipótese de mutação neutra por

polimorfismo de DNA. **Genética** , v. 123, n. 3, pág. 585–595, 1 nov. 1989. . Acesso em: 6 nov. 2022.

TORO OSPINA, AM; AGUILAR, I.; VARGAS DE OLIVEIRA, MS; CRUZ DOS SANTOS CORREIA, LE; VERCESI FILHO, AE; ALBUQUERQUE, LG; DE VASCONCELOS SILVA, JAI Avaliação da precisão da imputação na raça Gir usando diferentes painéis SNP. **Genoma** , v. 64, n. 10, pág. 893-899, fora. 2021. . Acesso em: 1 nov. 2022.

WANG, J.; LIAO, Y.; JAEHNIG, E.; SHI, Z.; SHENG, Q. **Gene Set Analysis Toolkit WebGestaltR** . [sl: sn].

WATHES, DC; CHENG, Z.; SALAVATI, M.; BUGGIOTTI, L.; TAKEDA, H.; TANG, L.; BECKER, F.; INGVARTSEN, KI; FERRIS, C.; HOSTENS, M.; CROWE, MA Relações entre perfis metabólicos e expressão gênica em fígado e leucócitos de vacas leiteiras no início da lactação. **Journal of Dairy Science** , v. 104, n. 3, pág. 3596–3616, 1 mar. 2021. . Acesso em: 9 nov. 2022.

WILLIAMS, JL; HALL, SJG; DEL CORVO, M.; BALLINGALL, KT; COLLI, L.; AJMONE MARSAN, P.; BISCARINI, F. Endogamia e purga no nível genômico: o gado Chillingham revela extensa heterozigidade SNP não aleatória. **Genética Animal** , v. 47, n. 1, pág. 19–27, 1 fev. 2016. . Acesso em: 26 abr. 2021.

YOUSEFI, S.; ABBASSI-DALOII, T.; KRAAIJENBRINK, T.; VERMAAT, M.; MEI, H.; VAN'T HOF, P.; VAN ITERSON, M.; ZHERNAKOVA, D. v.; CLARINGBOULD, A.; FRANKE, L.; 'T HART, LM; SLIEKER, RC; VAN DER HEIJDEN, A.; DE KNIJFF, P.; 'T HOEN, PAC Um painel SNP para identificação de amostras de DNA e RNA. **BMC Genomics** , v. 19, n. 1, pág. 90, 25 dez. 2018. . Acesso em: 30 out. 2022.

ZHANG, F.; ZHANG, Z.-Y.; CAI, M.-D.; LI, X.-X.; LI, Y.-H.; LEI, Y.; YU, X.-L. Efeito da temperatura de vitrificação e das concentrações de crioprotetores no transcriptoma de mRNA de oócitos bovinos maduros após vitrificação em estágio imaturo. **Theriogenology** , v. 148, p. 225–235, 1 maio 2020. . Acesso em: 8 nov. 2022.