

UNIVERSIDADE ESTADUAL PAULISTA JÚLIO DE MESQUITA FILHO
FACULDADE DE CIÊNCIAS AGRÁRIAS E VETERINARIAS
CAMPUS DE JABOTICABAL
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E MELHORAMENTO
ANIMAL

USO DE SNP E HAPLÓTIPO EM ESTUDO DE ASSOCIAÇÃO GENÔMICA
AMPLA PARA CARACTERÍSTICAS PRODUTIVAS EM BÚFALOS
LEITEIROS DA RAÇA MURRAH

Diego Ortunio Rosa Gobo

Biólogo

2023

UNIVERSIDADE ESTADUAL PAULISTA JÚLIO DE MESQUITA FILHO
FACULDADE DE CIÊNCIAS AGRÁRIAS E VETERINARIAS

CAMPUS DE JABOTICABAL

**USO DE SNP E HAPLÓTIPO EM ESTUDO DE ASSOCIAÇÃO GENÔMICA
AMPLA PARA CARACTERÍSTICAS PRODUTIVAS EM BÚFALOS
LEITEIROS DA RAÇA MURRAH**

Pós-Graduando: Diego Ortunio Rosa Gobo

Orientador: Prof. Dr. Humberto Tonhati

Coorientadora: Sirlene Fernandes Lázaro

**Tese apresentada à Faculdade de
Ciência Agrárias e Veterinárias –
Unesp/FCAV, Campus de Jaboticabal,
como parte das exigências para a
obtenção do título de Doutor em
Genética e Melhoramento Animal.**

2023

G575u	Gobo, Diego Ortunio Rosa Uso de SNP e haplótipo em estudo de associação genômica ampla para características produtivas em búfalos leiteiros da raça Murrah / Diego Ortunio Rosa Gobo. -- Jaboticabal, 2023 94 p. Tese (doutorado) - Universidade Estadual Paulista (Unesp), Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal Orientador: Humberto Tonhati Coorientadora: Sirlene Fernandes Lázaro 1. Bubalinocultura. 2. Genômica. 3. ssGWAS. 4. Desequilíbrio de Ligação. 5. Bloco de Haplótipo. I. Título.
-------	--

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.



UNIVERSIDADE ESTADUAL PAULISTA

Câmpus de Jaboticabal



CERTIFICADO DE APROVAÇÃO

TÍTULO DA TESE: USO DE SNPs E HAPLÓTIPO EM ESTUDO DE ASSOCIAÇÃO GENÔMICA AMPLA PARA CARACTERÍSTICAS PRODUTIVAS EM BÚFALOS LEITEIROS DA RAÇA MURRAH

AUTOR: DIEGO ORTUNIO ROSA GOBO

ORIENTADOR: HUMBERTO TONHATI

COORDENADORA: SIRLENE FERNANDES LÁZARO

Aprovado como parte das exigências para obtenção do Título de Doutor em Genética e Melhoramento Animal, pela Comissão Examinadora:

Prof. Dr. HUMBERTO TONHATI (Participação Virtual)
Departamento de Zootecnia / FCAV UNESP Jaboticabal

Documento assinado digitalmente

gov.br FERNANDO SEBASTIAN BALDI REY
Data: 17/02/2023 12:03:26-0300
Verifique em <https://verificador.iti.br>

Prof.Dr. FERNANDO SEBASTIAN BALDI REY (Participação Virtual)
Departamento de Zootecnia / FCAV UNESP Jaboticabal

Documento assinado digitalmente

gov.br DANIELE BOTELHO DINIZ MARQUES
Data: 16/02/2023 10:50:20-0300
Verifique em <https://verificador.iti.br>

Profa.Dra. DANIELE BOTELHO DINIZ MARQUES (Participação Virtual)
Departamento de Zootecnia / Universidade Federal de Viçosa - Viçosa/MG

Documento assinado digitalmente

gov.br MARCOS VINÍCIUS GUALBERTO BARBOSA DA SILVA
Data: 15/02/2023 16:03:34-0300
Verifique em <https://verificador.iti.br>

Pesquisador Dr. MARCOS VINÍCIUS GUALBERTO BARBOSA DA SILVA (Participação Virtual)
Centro Nacional de Pesquisa de Gado de Leite Embrapa / Juiz de Fora/MG

Documento assinado digitalmente

gov.br LUCAS LIMA VERARDO
Data: 15/02/2023 10:47:19-0300
Verifique em <https://verificador.iti.br>

Prof. Dr. LUCAS LIMA VERARDO (Participação Virtual)
Departamento de Zootecnia / Universidade Federal dos Vales do Jequitinhonha e Mucuri - Diamantina/MG

Jaboticabal, 13 de fevereiro de 2023

DADOS CURRICULARES DO AUTOR

Diego Ortunio Rosa Gobo, filho de Vera Lúcia Mariano Gobo e Ortunio José Rosa Gobo, nascido em Cascavel-PR, em 03 de abril de 1993. Ingressou no curso de Biologia, na Universidade Federal do Paraná (UFPR), no ano de 2011 e concluiu-o em 2015. Em março de 2017 ingressou no Programa de Pós-Graduação em Zootecnia da Universidade Federal de Viçosa (UFV), em nível de mestrado sob a orientação da Prof. Dra. Simone Eliza Facioni Guimaraes obtendo o título de mestre em fevereiro de 2019. Em março de 2019, ingressou no Programa de Pós-Graduação em Genética e Melhoramento Animal da “Universidade Estadual Paulista “Júlio Mesquita Filho” (FCAV/UNESP) campos Jaboticabal, em nível de Doutorado sob a orientação do Prof. Dr. Humberto Tonhati.

O homem que não cultiva o hábito de pensar desperdiça um dos maiores prazeres da vida e não consegue aproveitar o máximo de si.

“Thomas Edison”

Dedicatória

Aos meus pais Vera Lúcia Mariano Gobo e Ortunio José Rosa Gobo pelo amor, incentivo, confiança e apoio incondicional

A minha irmã Danielly Rosa Gobo por todo o companheirismo, incentivo e apoio

A minha noiva Pamela Itajara Otto pelo companheirismo, contribuições, carinho e compreensão durante este percurso

AGRADECIMENTOS

Agradeço primeiramente a Deus, pelo maravilhoso dom da vida e por ter me dado força em toda a minha trajetória.

Aos meus pais, Ortunio José Rosa Gobo e Vera Lúcia Mariano Gobo, por sempre estarem ao meu lado em todos os momentos, pelos bons exemplos, carinho, atenção e aos bons conselhos sempre me orientando e me apoiando em todas as minhas escolhas. A minha irmã Danielly por sempre me ajudar e pelos bons momentos em família. A toda minha família, por sempre me acolher bem e aos bons momentos juntos.

A minha noiva Pamela Itajara Otto, por todo afeto, companheirismo, momentos de descontração, contribuições durante esta etapa e paciência.

Ao meu orientador, Prof. Dr. Humberto Tonhati e minha coorientadora, Dra. Sirlene Fernandes Lazaro, pela atenção, conhecimento e paciência durante todas as etapas do trabalho.

Aos professores do programa pelos conselhos valiosos e toda a contribuição durante todo o mestrado.

Aos meus amigos de Cascavel e Jaboticabal, por todos os momentos felizes e as bagunças, aos companheiros de laboratório e amigos do departamento, pela parceria, conhecimento e bons conselhos.

A Universidade Estadual Paulista “Júlio Mesquita Filho” e a Capes pelo financiamento e pela concessão de bolsas.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

E a todos que de alguma forma contribuíram para meu crescimento pessoal e profissional.

Muito Obrigado a Todos!

SUMÁRIO

USO DE SNP E HAPLÓTIPO EM ESTUDO DE ASSOCIAÇÃO GENÔMICA AMPLA PARA CARACTERÍSTICAS PRODUTIVAS EM BÚFALOS DE LEITE DA RAÇA MURRAH.....	ii
RESUMO	ii
USE OF SNP AND HAPLOTYPE IN A BROAD GENOMIC ASSOCIATION STUDY FOR PRODUCTIVE TRAITS IN MURRAH BREED MILK BUFFALOES	iii
ABSTRACT	iii
CAPÍTULO 1 - CONSIDERAÇÕES GERAIS	1
1.1. Histórico da bubalinocultura no mundo	1
1.2. Histórico da Bubalinocultura no Brasil.....	2
1.3. Associação Brasileira de Criadores de Búfalos (ABCB)	3
1.4. Bubalinocultura de Leite.....	3
1.5. Controle Zootécnico	5
1.6. Características de interesse econômico	6
1.7. Análises Estatísticas	7
1.8. Melhoramento genético na Bubalinocultura	8
1.9. Desequilíbrio de Ligação.....	9
1.10. Blocos de haplótipos	11
1.11. Estudo de Associação Ampla do Genoma (GWAS).....	12
1.12. REFERÊNCIAS.....	13
CAPÍTULO 2 - Estudo de Associação Genômica Ampla em passo único (ssGWAS) para características produtivas em bubalinos da raça Murrah	24
2.1. INTRODUÇÃO	26
2.2. MATERIAL E MÉTODOS.....	27
2.3. RESULTADOS E DISCUSSÃO	32
2.4. CONCLUSÃO	44
2.5. REFERÊNCIAS.....	45
CAPÍTULO 3 – Estudo de associação genômica ampla com base em haplótipos para características produtivas em búfalas da raça Murrah.....	49
3.1. INTRODUÇÃO	51
3.2. MATERIAL E MÉTODOS.....	52
3.3. RESULTADOS E DISCUSSÃO	58
3.4. CONCLUSÃO	69
3.5. REFERÊNCIAS.....	70
CAPÍTULO 4 – Considerações finais	73

REFERÊNCIAS.....	74
APÊNDICES.....	76

USO DE SNP E HAPLÓTIPO EM ESTUDO DE ASSOCIAÇÃO GENÔMICA AMPLA PARA CARACTERÍSTICAS PRODUTIVAS EM BÚFALOS DE LEITE DA RAÇA MURRAH

RESUMO: Este estudo teve como objetivo identificar regiões associadas às características de produção de leite (PL), produção de proteína (PP), produção de gordura (PG), percentual de proteína (PORP), percentual de gordura (PORG) e escore celular (ECS), acumulado até 305 dias, usando duas metodologias diferentes: janelas de 10 SNP consecutivos e estrutura de blocos de haplótipos e identificar genes candidatos por meio de análise de ontologia gênica. Os dados fenotípicos consistiram em 2.336 a 10.159 registros de produção, composição e qualidade do leite acumulados ao longo de 305 dias de búfalas leiteiras Murrah. A estimativa dos componentes de variância e parâmetros genéticos, bem como as análises do estudo de associação genômica ampla (GWAS), foram realizadas usando o procedimento BLUP genômico de etapa única (ssGBLUP). Os blocos de haplótipos foram estimados com base no desequilíbrio de ligação (LD) estimado por r^2 usando o programa Big-LD, através do pacote “gpart” no software R. O GWAS permitiu a identificação de 22 genes candidatos para as características de produção e composição do leite avaliadas na população de búfalas Murrah, enquanto seis genes foram encontrados usando a metodologia de blocos de haplótipos. O número de marcadores SNP por região genômica variou de 2 a 10 em blocos de haplótipos GWAS, em comparação com 10 SNP fixados em análises de janelas. Desta forma, o tamanho das janelas em pares de bases (pb) é maior que os blocos, o que permitiu a identificação de um maior número de genes. Apenas a família de genes SLC26 foi localizado em ambas as metodologias, necessitando estudos adicionais para comprovar sua influência na produção e qualidade do leite em búfalas. Ambas as metodologias se mostraram eficientes na identificação de genes relacionados às características avaliadas na população de búfalos Murrah, proporcionando melhor compreensão biológica de sua arquitetura genética.

Palavras-chave: Búfalos, genes, genômica, ssGBLUP, ssGWAS.

USE OF SNP AND HAPLOTYPE IN A BROAD GENOMIC ASSOCIATION STUDY FOR PRODUCTIVE TRAITS IN MURRAH BREED MILK BUFFALOES

ABSTRACT: This study aimed to identify regions associated with the of milk production (PL), protein production (PP), fat production (PG), percentage of protein (PORP), percentage of fat (PORG) and cell score (ECS) traits, accumulated in 305 days, using two different methodologies: windows of 10 consecutive SNPs and haplotype block's structure and to identify candidate genes through gene ontology analysis. The phenotypic data consisted of 2,336 to 10,159 milk production, composition and quality records accumulated over 305 days from Murrah dairy buffalo. The estimation of variance components and genetic parameters, as well as the genomic wide association study (GWAS) analyzes, were performed using the single-step genomic BLUP procedure (ssGBLUP). The haplotype blocks were estimated based on the linkage disequilibrium (LD) estimated by r^2 using the Big-LD program, through the "gpart" package in the R software. GWAS allowed the identification of 22 candidate genes for the milk production and composition traits evaluated in the Murrah buffalo population, while six genes were found using the haplotype blocks methodology. The SNP markers number by genomic region ranged from 2 to 10 in haplotypes blocks GWAS, comparing to 10 SNP fixed in on windows analyses. In this way, the windows size in base pairs (bp) is bigger than the blocks, which allowed the identification of a greater number of genes. Only the SLC26 gene family was located in both methodologies, requiring additional studies to prove its influence on the production and quality of milk in buffaloes. Both methodologies were shown to be efficient in identifying genes related to the evaluated traits in the Murrah buffalo population, providing better biological comprehension of its genetic architecture.

Keywords: Buffaloes, genes, genomics, ssGBLUP, ssGWAS.

CAPÍTULO 1 - CONSIDERAÇÕES GERAIS

1.1. Histórico da bubalinocultura no mundo

O início da domesticação e criação dos búfalos pelo homem é datada por volta de 5.000 anos, período recente quando comparada a criação dos bovinos, que ocorreu por volta de 10.000 anos (Damasceno et al., 2010). Durante a domesticação dos búfalos, sua difusão passou a ser consideravelmente rápida por serem animais de dupla ou até tripla aptidão, extremamente dóceis e de fácil adaptação aos diferentes climas e condições as quais são submetidos (Marcondes et al., 2010).

O búfalo (*Bubalus bubalis*) se originou no continente Asiático, região com climas abrangentes localizados nas zonas tropical e temperada. Os búfalos da água, também conhecidos como búfalos asiáticos, abrangem duas subespécies: os búfalos de rio (*B. bubalis bubalis*) e os búfalos do pântano (*B. bubalis kerebau*). Apesar dos búfalos da água serem animais domesticados, eles apresentam grande semelhança com as espécies selvagens, podendo ser facilmente confundidos com os búfalos selvagens, dificultando assim as estimativas das populações domesticadas e selvagens (Martins, 1991).

Diferente dos bovinos que apresentam 60 cromossomos, 58 autossomos e o par sexual, os búfalos apresentam variação entre as subespécies. Os búfalos de rio apresentam 50 cromossomos, 48 autossomos e dois sexuais (Iannuzzi, 1994) e os búfalos de pântano contém 46 cromossomos autossomos e dois sexuais (Albuquerque et al., 2006).

Segundo estimativas da Fao (2017), a população mundial de bubalinos domesticados chega a aproximadamente 208 milhões de cabeças, com um aumento estimado de aproximadamente 3% ao ano. Destes, aproximadamente 164 milhões de animais (79%) são búfalos de rio e aproximadamente 44 milhões (21%) do pântano. Dentre todos os animais domesticados contabilizados, cerca de 196 milhões de cabeças, ou seja, 97% de toda a produção mundial está localizada no continente asiático.

Estes animais são mundialmente conhecidos pela qualidade de carne e leite, responsável por fornecer o leite para a confecção de um dos queijos mais

nobres do mundo, a mozzarella de búfala. Em 2019, registrou-se uma produção de leite mundial de aproximadamente 134.791.772 toneladas, parte utilizada para a produção de *mozzarella* de búfala, atingindo uma produção de aproximadamente 244.3 toneladas. No mesmo ano foram produzidas 4.124.188 toneladas de carne de búfalo (Fao, 2019).

1.2. Histórico da Bubalinocultura no Brasil

Os primeiros relatos da introdução de bubalinos no Brasil se dá por volta de 1890. Estes primeiros animais foram importados da China, introduzidos por colonizadores ingleses ou holandeses. O primeiro relato de entrada de animais confirmada é na ilha de Marajó, feita por Vicente Chermont de Miranda para suas fazendas no ano de 1906 (Associação Brasileira De Criadores De Búfalos, 2019).

A partir de 1919, o rebanho de búfalos deixou de estar presente apenas no Pará e passou a ser criado em outras regiões do Brasil. Por apresentarem rusticidade, alta adaptabilidade e elevada fertilidade, a produção de bubalinos apresentou bom aceite entre os criadores. As criações brasileiras são caracterizadas por pequenas ou médias propriedades no geral, contudo, na região Norte podem ser encontrados grandes rebanhos (Bernardes, 2006).

Segundo a Associação Brasileira de Criadores de Búfalos (ABCB), atualmente o país apresenta aproximadamente 3 milhões de animais. De acordo com o Instituto Brasileiro de Geografia e Estatística (IBGE, 2020), o Brasil possui o maior rebanho da América Latina, com cerca de 1.502.482 cabeças, sendo que a região Norte do país destaca na produção, com 67,5% do rebanho nacional (Tabela 1). Ceará, Rio Grande do Norte e São Paulo apresentam um rebanho de 1.705, 1.848 e 114,523, respectivamente.

Tabela 1. Distribuição efetiva do rebanho de bubalinos em 2020 no Brasil e Grande Região.

Regiões	Rebanho	% Rebanho
Norte	1.013.630	67,50

Nordeste	131.493	8,75
Centro-Oeste	61.778	4,00
Sudeste	201.304	13,50
Sul	94.277	6,25
Total	1.502.482	100,00

Fonte: (IBGE, 2020)

Atualmente, segundo a ABCB, cerca de 30% de todo o rebanho nacional é focado na produção de leite. Entre as raças de bubalinos, o Brasil conta com quatro raças registradas e reconhecidas pela ABCB: Mediterrâneo, Jafarabadi e Murrah (búfalos de rio) e o Carabao (búfalo de pântano). Cada raça apresenta especialidades, porém, a docilidade do rebanho é uma característica universal das quatro raças.

1.3. Associação Brasileira de Criadores de Búfalos (ABCB)

Com o intuito de incentivar os produtores, a Associação Brasileira de Criadores de Búfalos foi fundada em 1960, com o objetivo de promover a união entre os criadores de búfalos no país e defender seus interesses, bem como incentivar o desenvolvimento dos rebanhos por meio do incentivo técnico-científico da espécie e otimizar o mercado interno e externo no país.

A ABCB conta com cerca de 80 associados em todo o país. O rebanho leiteiro tem apresentado crescimento contínuo, com aumento aproximado de 45%, representando o maior rebanho de búfalos do ocidente. Atualmente, a associação conta com diversos benefícios para seus associados, como a orientação técnica de produção, simpósios, feiras, palestras, diversos cursos, registros genealógicos, boletim informativo, entre outros.

1.4. Bubalinocultura de Leite

A produção leiteira conta com três raças de búfalas: Murrah, Mediterrâneo e Jafarabadi. Búfalos submetidos a boas condições de manejo e alimentação produzem entre 10 e 16 litros de leite por dia. Os rebanhos normalmente são criados a pasto, com o fornecimento de suplementação e em alguns casos em

sistemas intensivos a pasto e confinamento. A extração do leite é realizada manualmente ou de forma mecânica, com o uso da ordenhadeira (Dias; Luiz; Machado Pfeifer, 2020).

Dentre as raças, a Murrah é a mais utilizada para a produção de leite. De origem indiana, esta raça é representada por animais compactos, com cor característica negra dos olhos e pelos, sinal de maior produtividade pelos indianos. Manchas brancas só são permitidas no extremo da traseira e rabo. Estes animais apresentam chifres curtos e espiralados, de onde vem seu nome, Murrah do hindu significa espiralado. Possuem boa conversão alimentar e é considerada a melhor raça produtora de leite dentre as raças criadas no Brasil.

A raça Murrah é considerada a mais populosa do país, sendo a principal indicação para produtores que apresentam como objetivo a produção da *mozzarella* de búfala. Os machos apresentam até 800kg e as fêmeas cerca de 600kg. A raça apresenta uma produção média de 1.650 litros de leite em 305 dias de lactação.

A criação de búfalas leiteiras no Brasil vem ganhando espaço ao longo dos anos, com crescimento de aproximadamente 10% ao ano (ABCB). Muitos produtores de bovinos passaram a diversificar suas criações com pequenos lotes de animais como um possível investimento, e, por suas qualidades de manejo e do leite, muitos migraram gradativamente da bovinocultura para a criação de bubalinos. O leite de búfala está em segundo lugar em importância produtiva e econômica, perdendo apenas para a produção de leite de vaca.

O leite de búfala é mais nutritivo e apresenta características específicas que o tornam superior quando comparado ao leite de vaca (Tabela 2). O valor nutricional do leite, gordura, proteína, cálcio e fósforo são alguns dos destaques diferenciais que podem ser observados. Por esses motivos, a utilização do leite de búfalas para a produção de derivados desperta interesse nos produtores por seu maior rendimento.

Tabela 2. Comparação do valor médio das propriedades do leite de búfalas e de vacas.

Parâmetros	Espécie
------------	---------

	Búfalas	Vacas
Gordura (%)	8,16	3,68
Proteína (%)	4,50	3,70
Cinzas (%)	0,70	0,70
Extrato seco total (%)	17,00	12,00
Cálcio (mg)	180,00	107,00
Ferro (mg)	0,12	0,07
Fósforo (mg)	120,00	82,00
Vitamina A (mcg)	54,00	48,90
Calorias por 100 mL	104,29	62,83

Fonte: (Verruma; Salgado, 1994)

As búfalas apresentam capacidade de produção superior quando comparadas as vacas, em média 5 litros de leite são necessários para a produção de 1 kg de *mozzarella* de búfala e aproximadamente 13 litros para se produzir 1 kg de manteiga. Com o leite de vaca, por sua vez, são necessários 22 litros para produzir 1 kg de manteiga e aproximadamente 10 litros para a produção de 1 kg de queijo *mozzarella*. Pelo sabor característico, a *mozzarella* e a ricota de búfala são produtos com alta demanda no mercado mundial (Teixeira; Bastianetto; Oliveira, 2005).

1.5. Controle Zootécnico

O controle leiteiro é a principal ferramenta para o controle produtivo do rebanho, permitindo o acompanhamento e comparação do potencial produtivo dos animais. Este controle consiste na coleta de informações fenotípicas de cada animal do rebanho e armazenadas no registro da propriedade, permitindo avaliar a real situação do rebanho. Além disso, possibilita o controle da qualidade do leite, por meio das informações de proteína e gordura e saúde dos animais (Lopes; Cardoso; Demeu, 2009). É de extrema importância também a anotação da identificação do animal e de seus ascendentes, garantindo o controle do pedigree dos animais na propriedade.

Vários índices podem ser utilizados como base para compreender o atual cenário da população avaliada e contribuir para um melhor direcionamento do programa de melhoramento. Dados de produção, qualidade do leite e índices reprodutivos são comumente mais coletados entre os rebanhos leiteiros. Com

estas informações, juntamente com informações de pedigree, o profissional poderá direcionar o produtor a tomar decisões mais eficientes.

Embora o controle zootécnico seja tão importante para a implementação dos programas de melhoramento genético nos rebanhos, ainda existe resistência entre produtores de búfalos na adequada coleta de informações dos seus animais, dificultando assim a avaliação dos rebanhos.

1.6. Características de interesse econômico

A produção de leite de búfala apresenta grande demanda como matéria prima para a produção de produtos lácteos, como a *mozzarella* e queijo coalho, além de ser um excelente leite para consumo *in natura*, devido sua composição. No Brasil, o seu consumo ainda é pouco difundido e a preferência do consumidor ainda é maior pelo leite bovino, porém, cada vez mais temos um aumento no consumo de derivados do leite de búfalas (Fuquay; Fox; Mcsweeney, 2011).

Informações sobre a produção de leite, gordura e proteína são características que influenciam diretamente a lucratividade do rebanho. Vários fatores podem afetar a produção de leite, como o ambiente, qualidade nutricional dos alimentos fornecidos, conforto térmico, número de ordenhas, manejo de modo geral e a constituição genética dos animais (Tonhati; Vasconcellos; Albuquerque, 2000). A procura por leite de búfala vem ganhando cada vez mais destaque no país, aumentando assim a procura por animais altamente produtivos. Atualmente, a raça Murrah vem se destacando entre os produtores por apresentar bons resultados de produtividade quando submetida a volumoso de boa qualidade e suplementação.

Um importante componente do leite é a produção de gordura, pois é o principal componente energético do leite e responsável pelo aumento do rendimento dos produtos industriais derivados lácteos. A proteína do leite está diretamente relacionada com a quantidade de aminoácidos absorvidos pelo intestino delgado, logo, quanto maior a concentração absorvida, maior o substrato para a produção das proteínas no leite (Hanigan et al., 1998). O manejo nutricional pode afetar sutilmente a composição, onde o fornecimento de dietas

balanceadas contribuem no aumento dos carboidratos e proteínas essenciais para a produção de leite (Fredeen, 1996). O fornecimento de fontes de ureia também favorece o aumento de proteína no leite (Cherdthong; Wanapat, 2010).

Por fim, avalia-se a contagem de células somáticas (CCS), convertida como escore de células somáticas (ECS) conforme descrito por Feng et al. (2021):

$$ECS = [\text{Log}_2 (CCS/100) + 3].$$

A CCS no leite são células imunológicas presentes no organismo direcionadas através do sangue para o interior das glândulas mamárias, a fim de combater principalmente agentes causadores de mastite, que é a infecção da glândula mamária. Em um animal com mastite, cerca de 99% das células de defesa no leite são relacionadas a glândula mamária infectada (Cunha et al., 2008).

Segundo Aspilcueta-Borquis et al. (2010), em búfalas, a produção acumulada até 305 (produção/lactação) é em média de $1.814 \pm 641,08$ kg de leite, produção de gordura em média de $119,26 \pm 37,40$ kg, produção de proteína de $77,38 \pm 24,26$ kg, porcentagem de gordura de $6,84 \pm 1,35\%$, porcentagem de proteína de $4,27 \pm 0,54\%$ e um escore células somáticas de $5,02 \pm 1,26$. Segundo Rosati; Van Vleck (2002) ao estudar bubalinos das raças Murrah, Ravi, Kundi, Beheri, Saidi e Shumen com duração de lactação de 270 dias, encontraram uma produção e porcentagem de gordura de 196,9kg e 8,59 e de proteína de 104,7 e 4,55, respectivamente. Em um estudo com animais da raça Murrah em rebanhos localizados nos estados de São Paulo e Rio Grande do Norte, Barros et al. (2016) descreveram uma produção de leite de 1.872 kg durante a lactação.

1.7. Análises Estatísticas

A produção e qualidade do leite são características expressas ao longo da vida produtiva dos animais, podendo ser avaliadas diariamente, mensalmente ou a cada lactação. Estatisticamente, essas características são ditas como medidas repetidas da mesma característica no animal e podem ser inseridas no modelo estatístico de avaliação genética como efeito de ambiente permanente,

utilizando-se o modelo de repetibilidade. Este modelo vem sendo implementado em avaliações de produção de leite e tamanhos de ninhada por exemplo e tem como vantagem ser um modelo de fácil aplicabilidade, menos parametrizado e necessitar de menor requisito computacional (Mrode, 2014). Assim, pode-se adotar o modelo animal de repetibilidade para avaliação deste tipo de característica:

$$y = X\beta + Za + Wp + e$$

em que y é o vetor de observações; β é o vetor de efeitos fixos; a é o vetor de efeitos genéticos aditivos aleatórios; p é o vetor de efeito ambiente permanente aleatório; e é o vetor de efeitos residuais aleatórios; e X , Z e W são matrizes de incidência de β , a e p , respectivamente. Assume-se que $a \sim N(0; H\sigma_a^2)$, $p \sim N(0; I\sigma_p^2)$ e $e \sim N(0; I\sigma_e^2)$, onde σ_a^2 , σ_p^2 e σ_e^2 representam as variâncias genética aditiva, de ambiente permanente e residual, respectivamente. A matriz H a matriz que combina informações genômicas e de pedigree e I a matriz identidade.

O modelo de repetibilidade vem sendo utilizado com eficiência para analisar características de produção, qualidade e saúde do leite em várias espécies (Otto et al., 2020a; Da Costa Barros et al., 2018^a; Sesana et al., 2014; Weller; Ezra, 2004).

1.8. Melhoramento genético na Bubalinocultura

O atual sistema de produção segue basicamente três pilares de sustentação: genética, manejo e nutrição. O foco em otimizar a produção dos animais está diretamente ligado com melhorias nos pilares, onde o tempo para obtenção do resultado por meio de alterações variam de acordo com as medidas adotadas e o tempo de resposta. Em um ângulo geral, mudanças em nutrição e manejo para a produção tendem a apresentar respostas mais rápidas que o manejo genético, porém, este representa a única forma de melhorar permanentemente o rebanho.

Ao melhorar a genética de um rebanho, o melhorista deve se atentar com as condições ambientais em que os animais estão submetidos, tentando ajustar as condições para melhorar o rendimento dos animais, possibilitando a máxima

expressão dos genótipos. O sucesso na produção de progênes geneticamente superiores depende de alguns fatores, como a correta anotação genealógica dos animais, seleção dos melhores indivíduos de forma acurada, com base nos valores genéticos dos animais e o ideal direcionamento do acasalamento dos animais selecionados (Marcondes; Marques; Cunha, 2008).

O conhecimento da natureza e da magnitude dos efeitos genéticos é fundamental para o delineamento do melhoramento genético para bubalinos. A inclusão desses efeitos no modelo estatístico, juntamente com a utilização de informações genômicas, pode possibilitar a estimação dos parâmetros genéticos e predição do mérito genético de forma mais precisa e acurada. Isto permite a seleção dos melhores animais, além de auxiliar no planejamento e direcionamento do programa de melhoramento, de forma que as melhores combinações gênicas possam ser obtidas, aproveitando o ganho em aditividade e complementariedade.

Atualmente, o Brasil conta com dois programas de melhoramento, o programa de controle leiteiro, coordenado pelo Departamento de Zootecnia da Universidade Estadual Paulista, campus Jaboticabal, São Paulo e o Promebull. O programa de controle leiteiro surgiu com o intuito de contribuir para o progresso genético dos rebanhos de búfalas. Já o programa Promebull surgiu em 2018 com convênio entre Embrapa Amazônia Oriental e Secretaria de Estado de Desenvolvimento Agropecuário e da Pesca (Sedap), na tentativa de alavancar a bubalinocultura paraense. O programa atua em quatro linhas: manejo animal, manejo sanitário, manejo nutricional e alimentar e melhoramento genético. Dentre os direcionamentos do programa, a importação de sêmen de bubalinos da Índia e Itália, a fim de garantir material genético superior aos produtores, porém, restrita apenas para as regiões de Marajó, Baixo Amazonas, Carajás e Bragantina (Cabral, 2018).

1.9. Desequilíbrio de Ligação

Segundo Weiss e Clark (2002), o desequilíbrio de ligação (LD) por definição é a associação não-aleatória entre alelos de dois ou mais loci, onde não necessariamente devem estar presentes no mesmo cromossomo. De modo

estatístico, o desequilíbrio de ligação é a medida de associação entre alelos posicionados em diferentes loci.

Vários fatores podem alterar o LD, tanto positivamente como negativamente. O pequeno número de indivíduos na população, seleção, autofecundação e isolamento genético são alguns dos fatores responsáveis por aumentar o LD, por outro lado, mutações, recombinações e o acasalamento ao acaso são fatores que afetam negativamente sobre o LD (Hartl; Clark, 2010).

Dentre os diferentes métodos para estimar o LD, a medida r^2 (Hill; Robertson, 1968) é a mais robusta e utilizada, sendo obtido por meio de:

$$r^2 = \frac{D^2}{p_1 p_2 q_1 q_2}$$

em que $D_{ij} = P_{ij} - p_i q_j$ é a medida de LD, conforme descrito por Lewontin (1964); P_{ij} é a frequência do haplótipo que consiste nos alelos i e j; p_i é a frequência do alelo i; q_j é a frequência do alelo j; e $p_1 p_2 q_1 q_2$ é o produto das 4 frequências alélicas em ambos os loci.

Vale ressaltar o cuidado que se deve tomar com a influência das frequências alélicas sobre a mensuração do LD, ignorar esse fator pode acarretar conclusões errôneas. Todos os métodos são influenciados pelas frequências alélicas, porém, o r^2 apresenta a menor influência, sendo considerada a metodologia mais robusta e por isso a mais utilizada para estimação do LD e montagem de blocos de haplótipos.

Estudos prévios que avaliaram o LD em diferentes raças de búfalos encontraram um r^2 médio entre SNP adjacentes variando de 0,09 a 0,32 (Deng et al., 2019; Mokhber et al., 2019). Deng et al. (2019) analisando o decaimento do LD em búfalos puros Mediterrâneo e mestiços, observaram que os valores médios de $r^2 \geq 0,2$ estenderam-se até ~50 kb e 170 kb, enquanto $r^2 \geq 0,3$ foram observados em ~10 e 60 kb nas populações mestiças e puras, respectivamente. De acordo com estes autores, considerando o tamanho do genoma do búfalo de aproximadamente 2,83 Gb (Williams et al., 2017), seriam necessários 47,167 SNP para a captura da maioria das informações de LD para GWAS para avaliação de búfalos do Mediterrâneo.

1.10. Blocos de haplótipos

Por definição, haplótipos são regiões ou agrupamentos de alelos adjacentes distribuídos pelo genoma com alto desequilíbrio de ligação entre os marcadores e pouco ou nenhum relato de recombinação reconhecida (Cardon; Abecasis, 2003).

Segundo Wall; Pritchard, (2003), a maioria dos blocos formados ao longo do genoma não apresentam grande número de marcadores, sendo que a maioria dos blocos relatados na literatura apresentam tamanho médio de 5-20 kb. Blocos grandes são considerados improváveis em um modelo genético populacional, pois para a sua ocorrência seria necessárias grandes áreas com baixo nível de recombinação.

A montagem dos blocos de haplótipos pode ser baseada em dois métodos: 1) método direto, onde são aplicados experimentos especializados no DNA genômico do indivíduo; e 2) método inferencial, comumente mais utilizado com base na literatura, onde utiliza-se base estatística de genótipos não faseados de vários indivíduos relacionados (Snyder et al., 2015).

Atualmente a metodologia de montagem de blocos de haplótipo Big-LD, disponível no pacote do software R “gpart”, que utiliza genótipos faseados para montagem de seus blocos, vem sendo utilizada com sucesso, apresentando bons resultados. Segundo Kim et al. (2018), a metodologia apresenta boa concordância com as regiões recombinantes do genoma e apresenta um menor tempo de execução médio, quando comparado aos outros métodos disponíveis.

Basicamente, o Big-LD estima os desequilíbrios de ligação dos SNPs com base em modelagem gráfica de intervalos através do método r^2 , montando o agrupamento entre SNPs com forte LD. Com os agrupamentos (intervalos) formados, a abordagem determina quais intervalos sobrepostos podem ser classificados como um bloco através do número de marcadores e inferência bayesiana. Deste modo, teremos blocos com maior probabilidade de serem herdados por serem blocos com forte LD (Kim et al., 2018).

Relatos na literatura mostram a superioridade na implementação dos estudos de associação com haplótipos quando comparado ao uso de marcadores SNPs de forma individual. Segundo Villumsen; Janss; Lund, (2009), a implementação dos haplótipos nas análises genômicas aumenta a confiabilidade de predição em dados simulados em comparação a utilização de marcadores SNPs. Em uma população de animais holandeses utilizando marcadores de alta densidade, Cuyabano; Su; Lund (2014) avaliaram a confiabilidade da utilização de SNPs e blocos de haplótipos e constataram que a utilização dos haplótipos aumentou a confiabilidade da predição genômica em 3% para características de interesse comercial.

1.11. Estudo de Associação Ampla do Genoma (GWAS)

A extensão do desequilíbrio de ligação ao longo do genoma apresenta grande influência na acurácia da predição de valores genéticos genômicos, em estudos de GWAS, de seleção genômica (GS), seleção assistida por marcadores (MAP), mapeamento de QTL, testes de parentesco, testes de doenças baseadas em marcadores genômicos, entre outros. Estas metodologias se tornaram amplamente utilizadas e fazem parte das ferramentas de rotina aplicadas na criação de animais e em programas de genética e melhoramento animal (O'Brien et al., 2014). Desta forma, a análise prévia da extensão do LD no genoma é necessária para determinar o número de marcadores que serão suficientes para se obter bons resultados em GWAS (Khatkar et al., 2008; McKay et al., 2007).

O desenvolvimento de marcadores moleculares identificados por toda extensão do genoma bubalino, ainda que recente, além de seu uso em testes genômicos, também possibilitou a realização de GWAS, que usam tecnologias de genotipagem de alto rendimento para identificar SNPs e relacioná-los com características de interesse. Os GWAS dependem da força do LD entre marcadores genéticos e loci de características quantitativas (QTL) de uma população, além do número de marcadores utilizados e a frequência dos alelos (Marigorta; Navarro, 2013; Veroneze et al., 2013). Os marcadores devem estar em LD suficiente com um QTL para permitir a estimação de seus efeitos em toda a população e entre gerações (Neves et al., 2015). Desta forma, a extensão do

LD dentro de uma população ajuda a determinar o número de marcadores necessários para o mapeamento em análises de GWAS.

A análise de GWAS representa uma das principais estratégias para a investigação das mutações subjacentes aos marcadores em características complexas. Estes estudos identificam marcadores de DNA que estão associados às características de interesse, apontando regiões onde podem estar localizados os genes potencialmente envolvidos com o fenótipo de interesse (Plomin et al., 1994).

A identificação dos QTLs, SNPs e genes associados as características é apenas o primeiro passo em análises de GWAS. Estudos adicionais, de ontologia gênica, são necessários para identificar os genes candidatos, ou seja, genes que apresentam alguma ligação funcional com a característica estudada e que podem apresentar importância para sua expressão. Diversos estudos têm aplicado com eficiência o uso de análises de ontologia gênica, como a identificação dos processos biológicos, para a identificação dos genes candidatos nas mais variadas espécies avaliando-se diferentes características de interesse econômico (Marques et al., 2018; Otto et al., 2020; Soares et al., 2017; Verardo et al., 2016).

1.12. REFERÊNCIAS

- AGUILAR, I. et al. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. **Journal of Dairy Science**, v. 93, n. 2, p. 743–752, fev. 2010.
- AHLAWAT, S. et al. Comparative gene expression profiling of milk somatic cells of Sahiwal cattle and Murrah buffaloes. **Gene**, v. 764, 5 jan. 2021.
- ALBUQUERQUE, M. DO S. M. et al. **Conservação e caracterização de búfalos no brasil: uma revisão da literatura**. Brasília: [s.n.]. Disponível em: <<http://www.cenargen.embrapa.br>>.
- ANTON, I. et al. **Effect of DGAT1, leptin and TG gene polymorphisms on some milk production traits in different dairy cattle breeds in Hungary**. [s.l.: s.n.].
- ARAUJO, A. C. et al. Haplotype-based single-step gwas for yearling temperament in american angus cattle. **Genes**, v. 13, n. 1, 1 jan. 2022.

ASPILCUETA-BORQUIS, R. R. et al. Genetic parameters for buffalo milk yield and milk quality traits using Bayesian inference. **Journal of Dairy Science**, v. 93, n. 5, p. 2195–2201, maio 2010.

ASPILCUETA-BORQUIS, R. R. et al. Multiple-trait random regression models for the estimation of genetic parameters for milk, fat, and protein yield in buffaloes. **Journal of Dairy Science**, v. 96, n. 9, p. 5923–5932, set. 2013.

ASSOCIAÇÃO BRASILEIRA DE CRIADORES DE BÚFALOS. **Boletim do Búfalo**. Disponível em: <<https://www.bufalo.com.br/home/wp-content/uploads/2021/01/boletim-bufalo-2019-no1.pdf>>. Acesso em: 11 ago. 2021.

BAKHTIARIZADEH, M. R.; ALAMOUTI, A. A. RNA-Seq based genetic variant discovery provides new insights into controlling fat deposition in the tail of sheep. **Scientific Reports**, v. 10, n. 1, 1 dez. 2020.

BARROS, C. DA C. et al. Estimativas de parâmetros genéticos para características de produção e reprodução em búfalos leiteiros. **Revista Caatinga**, v. 29, n. 1, p. 216–221, 1 jan. 2016.

BEECHER, C. et al. Polymorphisms in bovine immune genes and their associations with somatic cell count and milk production in dairy cattle. **BMC Genetics**, v. 11, 5 nov. 2010.

BEGUM, F. et al. Regionally Smoothed Meta-Analysis Methods for GWAS Datasets. **Genetic Epidemiology**, v. 40, n. 2, p. 154–160, 1 fev. 2016.

BERNARDES, O. Os Búfalos no Brasil. In: **II Simpósio de Búfalo de Las Américas**, v. 2, p. 18–23, 2006.

BINDEA, G. et al. ClueGO: A Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. **Bioinformatics**, v. 25, n. 8, p. 1091–1093, 2009.

BINDEA, G.; MLECNIK, B. ClueGO Documentation. p. 1–33, 2009.

CABRAL, K. **Programa de melhoramento genético de búfalos é apresentado no Baixo Amazonas**.

CALUS, M. P. et al. Effects of the number of markers per haplotype and clustering of haplotypes on the accuracy of QTL mapping and prediction of genomic breeding values. **Genetics Selection Evolution**, v. 41, n. 11, 2009.

CARDON, L. R.; ABECASIS, G. R. **Using haplotype blocks to map human complex trait loci**. **Trends in Genetics** Elsevier Ltd, , 1 mar. 2003.

CHENG, F. et al. RNA-Binding Protein Dnd1 Promotes Breast Cancer Apoptosis by Stabilizing the Bim mRNA in a miR-221 Binding Site. **BioMed Research International**, v. 2017, 2017.

CHERDTHONG, A.; WANAPAT, M. Development of Urea Products as Rumen Slow-Release Feed for Ruminant Production: A Review. **Australian Journal of Basic and Applied Sciences**, v. 4, n. 8, p. 2232–2241, 2010.

CHHABRA, S. N.; BOOTH, B. W. **Asymmetric cell division of mammary stem cells**. **Cell Division**BioMed Central Ltd, , 1 dez. 2021.

CHOI, E. J. et al. Intratumoral Heterogeneity of Frameshift Mutations in MECOM Gene is Frequent in Colorectal Cancers with High Microsatellite Instability. **Pathology and Oncology Research**, v. 23, n. 1, p. 145–149, 1 jan. 2017.

COCHRAN, S. D. et al. Discovery of single nucleotide polymorphisms in candidate genes associated with fertility and production traits in Holstein cattle. **BMC Genetics**, v. 14, 7 jun. 2013.

COELHO, A. I.; BERRY, G. T.; RUBIO-GOZALBO, M. E. **Galactose metabolism and health**. **Current Opinion in Clinical Nutrition and Metabolic Care**Lippincott Williams and Wilkins, , 1 jul. 2015.

CONNOR, E. E. et al. Effects of increased milking frequency on gene expression in the bovine mammary gland. **BMC Genomics**, v. 9, 31 jul. 2008.

COUSSENS, L. M.; WERB, Z. **Inflammation and cancer**. [s.l: s.n.]. Disponível em: <www.nature.com/nature>.

CRISÀ, A. et al. RNA-Sequencing for profiling goat milk transcriptome in colostrum and mature milk. **BMC Veterinary Research**, v. 12, n. 1, 25 nov. 2016.

CRONIN, J. G. et al. Toll-like receptor 4 and MYD88-dependent signaling mechanisms of the innate immune system are essential for the response to lipopolysaccharide by epithelial and stromal cells of the bovine endometrium. **Biology of Reproduction**, v. 86, n. 2, 1 fev. 2012.

CUNHA, R. P. L. et al. Mastite subclínica e relação da contagem de células somáticas com número de lactações, produção e composição química do leite em vacas da raça Holandesa. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, v. 60, p. 19–24, 2008.

CUYABANO, B. C.; SU, G.; LUND, M. S. **Genomic prediction of genetic merit using LD-based haplotypes in the Nordic Holstein population****BMC Genomics**. [s.l: s.n.]. Disponível em: <<http://www.biomedcentral.com/1471-2164/15/1171>>.

CUYABANO, B. C.; SU, G.; LUND, M. S. **Genomic prediction of genetic merit using LD-based haplotypes in the Nordic Holstein population****BMC Genomics**. [s.l: s.n.]. Disponível em: <<http://www.biomedcentral.com/1471-2164/15/1171>>.

DA COSTA BARROS, C. et al. Use of single-step genome-wide association studies for prospecting genomic regions related to milk production and milk quality of buffalo. **Journal of Dairy Research**, v. 85, n. 4, p. 402–406, 1 nov. 2018a.

DA COSTA BARROS, C. et al. Use of single-step genome-wide association studies for prospecting genomic regions related to milk production and milk quality of buffalo. **Journal of Dairy Research**, v. 85, n. 4, p. 402–406, 1 nov. 2018b.

DAMASCENO, F. A. et al. **ADAPTAÇÃO DE BUBALINOS AO AMBIENTE TROPICAL** **TROPICAL Artigo**. [s.l.] Setembro/Outubro, 2010.

DE ARAUJO NETO, F. R. et al. Genome-wide association studies for growth traits in buffaloes using the single step genomic BLUP. **Journal of Applied Genetics**, v. 61, n. 1, p. 113–115, 1 fev. 2020.

DE ARAUJO NETO, F. R. et al. Genomic study of the resilience of buffalo cows to a negative energy balance. **Journal of Applied Genetics**, v. 63, n. 2, p. 379–388, 1 maio 2022.

DENG, T. et al. 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. 10, n. JAN, 2019.

DIAS, A. K.; LUIZ, S.; MACHADO PFEIFER, F. **PECUÁRIA LEITEIRA NA AMAZÔNIA**. [s.l: s.n.].

DIMRI, G.; BAND, H.; BAND, V. **Mammary epithelial cell transformation: Insights from cell culture and mouse models**. **Breast Cancer Research**, jul. 2005.

FAGIOLO, A.; LAI, O. Mastitis in buffalo. **Italian Journal of Animal Science**, v. 6, n. SUPPL. 2, p. 200–206, 2007.

FENG, L. et al. The effect of exogenous melatonin on milk somatic cell count in buffalo. **Pakistan Veterinary Journal**, v. 41, n. 1, p. 152–155, 2021.

FREDEEN, A. H. **Considerations in the nutritional modification of milk composition** **Animal Feed Science Technology**. [s.l: s.n.].

FUQUAY, JHON. W.; FOX, PATRICK. F.; MCSWEENEY, PAUL. L. H. **Encyclopedia of dairy sciences second edition**. AMSTERDAM: [s.n.].

GAOWA, N. et al. The effects of artificially dosed adult rumen contents on abomasum transcriptome and associated microbial community structure in calves. **Genes**, v. 12, n. 3, 1 mar. 2021.

GARCIA-CORTÉS, L. A. et al. **variance component estimation of yield traits between different lactations using an animal model** **Livestock Production Science**. [s.l.] Co, 1995.

GONÇALVES, J. L. et al. Bovine subclinical mastitis reduces milk yield and economic return. **Livestock Science**, v. 210, p. 25–32, 1 abr. 2018.

GONZALEZ GUZMAN, J. L. et al. Genome-wide association study applied to type traits related to milk yield in water buffaloes (*Bubalus bubalis*). **Journal of Dairy Science**, v. 103, n. 2, p. 1642–1650, 1 fev. 2020.

GUERRERO-ZOTANO, A. L. et al. ERp Breast cancers resistant to prolonged neoadjuvant letrozole exhibit an e2f4 transcriptional program sensitive to cdk4/6 inhibitors. **Clinical Cancer Research**, v. 24, n. 11, p. 2517–2529, 1 jun. 2018.

GUIMARÃES, J. L. B. et al. Estimate of the economic impact of mastitis: A case study in a Holstein dairy herd under tropical conditions. **Preventive Veterinary Medicine**, v. 142, p. 46–50, 1 jul. 2017.

HABIER, D. et al. Extension of the bayesian alphabet for genomic selection. **BMC Bioinformatics**, v. 12, 23 maio 2011.

HANIGAN, M. D. et al. An Evaluation of Postabsorptive Protein and Amino Acid Metabolism in the Lactating Dairy Cow. **Journal of Dairy Science**, v. 81, n. 12, p. 3385–3401, 1998.

HARTL, D. L.; CLARK, A. G. **Princípios de Genética de Populações**. 4. ed. [s.l.] USA: Jones and Bartlett Publishers, 2010.

HATSELL, S.; FROST, A. R. Hedgehog signaling in mammary gland development and breast cancer. **Journal of Mammary Gland Biology and Neoplasia**, v. 12, n. 2–3, p. 163–173, set. 2007.

HILL, W. G.; ROBERTSON, A. Linkage disequilibrium in finite populations. **TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik**, v. 38, n. 6, p. 226–31, 1968.

HILL, W. G.; ROBERTSON, A. **Linkage Disequilibrium in Finite Populations**. [s.l.: s.n.].

HOWARD, J. T. et al. Genome-wide association study on legendre random regression coefficients for the growth and feed intake trajectory on Duroc Boars. **BMC Genetics**, v. 16, n. 1, 30 maio 2015a.

HOWARD, J. T. et al. Genome-wide association study on legendre random regression coefficients for the growth and feed intake trajectory on Duroc Boars. **BMC Genetics**, v. 16, n. 1, 30 maio 2015b.

IANNUZZI, L. Standard karyotype of the river buffalo (*Bubalus bubalis* L., 2n = 50). **Cytogenetics and Cell Genetics**, p. 102–113, 4 jan. 1994.

IBGE. **PPM - Pesquisa da Pecuária Municipal**.

JIANG, L. Y. et al. Formation of Blood Neutrophil Extracellular Traps Increases the Mastitis Risk of Dairy Cows During the Transition Period. **Frontiers in Immunology**, v. 13, 27 abr. 2022.

JOHNSON, G. C. et al. Haplotype tagging for the identification of common disease genes. **Nature Genetics**, v. 29, n. october, p. 233–237, 2001.

KENDRICK, H. et al. Transcriptome analysis of mammary epithelial subpopulations identifies novel determinants of lineage commitment and cell fate. **BMC genomics**, v. 9, p. 591, 2008.

KHAN, M. Z. et al. **Role of the jak-stat pathway in bovine mastitis and milk production.** *AnimalsMDPI AG*, , 1 nov. 2020.

KHATIB, H. et al. **Evaluation of association between polymorphism within the thyroglobulin gene and milk production traits in dairy cattle.** [s.l: s.n.]. Disponível em: <<http://ep.espacenet.com>>.

KHATKAR, M. S. et al. Extent of genome-wide linkage disequilibrium in Australian Holstein-Friesian cattle based on a high-density SNP panel. **BMC Genomics**, v. 9, p. 1–18, 2008.

KHIALEEVA, E.; LANE, T. F.; CARPENTER, E. M. Disruption of reelin signaling alters mammary gland morphogenesis. **Development**, v. 138, n. 4, p. 767–776, 15 fev. 2011.

KIKU, Y. et al. Decrease in bovine CD14 positive cells in colostrum is associated with the incidence of mastitis after calving. **Veterinary Research Communications**, v. 34, n. 2, p. 197–203, fev. 2010.

KIM, S. A. et al. A new haplotype block detection method for dense genome sequencing data based on interval graph modeling of clusters of highly correlated SNPs. **Bioinformatics**, v. 34, n. 3, p. 388–397, 1 fev. 2018.

KIM, S. A. et al. Gpart: Human genome partitioning and visualization of high-density SNP data by identifying haplotype blocks. **Bioinformatics**, v. 35, n. 21, p. 4419–4421, 1 nov. 2019.

KORTE, A.; FARLOW, A. **The advantages and limitations of trait analysis with GWAS: A review.** *Plant Methods*, 22 jul. 2013.

KROPYVKO, S. et al. Mammalian verprolin CR16 acts as a modulator of ITSN scaffold proteins association with actin. **Biochemical and Biophysical Research Communications**, v. 484, n. 4, p. 813–819, 18 mar. 2017.

LÁZARO, S. F. et al. Genomic studies of milk-related traits in water buffalo (*Bubalus bubalis*) based on single-step genomic best linear unbiased prediction and random regression models. **Journal of Dairy Science**, v. 104, n. 5, p. 5768–5793, 1 maio 2021.

LEE, G. Y. et al. Comparative oncogenomics identifies psmb4 and shmt2 as potential cancer driver genes. **Cancer Research**, v. 74, n. 11, p. 3114–3126, 1 jun. 2014.

LEWONTIN, R. C. The Interaction of Selection and Linkage. I. General Considerations; Heterotic Models. **Genetics**, v. 49, n. 1, p. 49–67, 1964.

LI, D.; LIU, Q.; SCHNABLE, P. S. Twas results are complementary to and less affected by linkage disequilibrium than gwas. **Plant Physiology**, v. 186, n. 4, p. 1800–1811, 1 ago. 2021.

LI, Y.; SUNG, W. K.; LIU, J. J. Association mapping via regularized regression analysis of single-nucleotide-polymorphism haplotypes in variable-sized sliding windows. **American Journal of Human Genetics**, v. 80, n. 4, p. 705–715, 2007.

LI, Z.; TZENG, C. M. Integrated analysis of miRNA and mRNA expression profiles to identify miRNA targets. Em: **Methods in Molecular Biology**. [s.l.] Humana Press Inc., 2018. v. 1720p. 141–148.

LOPES, M. A.; CARDOSO, M. G.; DEMEU, F. A. **INFLUÊNCIA DE DIFERENTES ÍNDICES ZOOTÉCNICOS NA COMPOSIÇÃO E EVOLUÇÃO DE REBANHOS BOVINOS LEITEIROS** *Ciência Animal Brasileira*. [s.l: s.n.].

LOW, W. Y. et al. Chromosome-level assembly of the water buffalo genome surpasses human and goat genomes in sequence contiguity. **Nature Communications**, v. 10, n. 1, 1 dez. 2019.

LU, L. M. et al. Proteomic and functional analyses reveal MAPK1 regulates milk protein synthesis. **Molecules**, v. 18, n. 1, p. 263–275, 2013.

MA, Y. et al. **Genetic markers associated with milk production traits in dairy cattle**. **Agriculture (Switzerland)** MDPI, , 1 out. 2021.

MARCONDES, C. R. et al. **Variabilidade genética de búfalos em rebanho-núcleo com base na análise de pedigree** *Arq. Bras. Med. Vet. Zootec.* [s.l: s.n.].

MARCONDES, C. R.; MARQUES, J. R. F.; CUNHA, B. R. DO N. **Melhoramento Genético em Bubalinos**. AMAZONPEC. **Anais...** Belém: out. 2008.

MARIGORTA, U. M.; NAVARRO, A. High Trans-ethnic Replicability of GWAS Results Implies Common Causal Variants. **PLoS Genetics**, v. 9, n. 6, jun. 2013.

MARÍN-GARZÓN, N. A. et al. Genome-wide association study identified genomic regions and putative candidate genes affecting meat color traits in Nellore cattle. **Meat Science**, v. 171, 1 jan. 2021.

MARQUES, D. B. D. et al. Weighted single-step GWAS and gene network analysis reveal new candidate genes for semen traits in pigs. **Genetics Selection Evolution**, v. 50, n. 1, 6 ago. 2018.

MARTINS, J. M. **Estudo da qualidade da carne de bovinos da raça Cachena / Study of the meat quality from Cachena cattle View project TREASURE-Diversity of local pig breeds and production systems for high quality traditional products and sustainable pork chains View project**. [s.l: s.n.]. Disponível em: <<https://www.researchgate.net/publication/331634460>>.

MCKAY, S. D. et al. Whole genome linkage disequilibrium maps in cattle. **BMC Genetics**, v. 8, p. 1–12, 2007.

MISZTAL, I. et al. **BLUPF90 AND RELATED PROGRAMS (BGF90)**. [s.l: s.n.]. Disponível em: <<http://www.ozemail.com.au/~milleraj>>.

MISZTAL, I. et al. **Manual for BLUPF90 family of programs**. [s.l: s.n.].

MIZERANSCHI, A. E. et al. GENOME ASSOCIATION STUDY IN COW'S RESISTANCE TO MASTITIS. **LUCRĂRI ȘTIINȚIFICE MEDICINĂ VETERINARĂ**, v. 54, n. 4, p. 112–126, 2021.

MOKHBER, M. et al. Study of whole genome linkage disequilibrium patterns of Iranian water buffalo breeds using the Axiom Buffalo Genotyping 90K Array. **PLoS ONE**, v. 14, n. 5, 1 maio 2019.

MRODE, R. A. **Linear Models for the Prediction of Animal Breeding Values**. 3rd Edition ed. [s.l.] Cabi; 3rd UK ed. edição, 2014.

NASCIMENTO, A. V. et al. An updated Axiom buffalo genotyping array map and mapping of cattle quantitative trait loci to the new water buffalo reference genome assembly. **Animal Genetics**, v. 52, n. 4, p. 505–508, 1 ago. 2021.

NEVES, H. H. R. et al. Preliminary study to determine extent of linkage disequilibrium and estimates of autozygosity in Brazilian Gyr dairy cattle. 2015.

OGOLA, H.; SHITANDI, A.; NANUA, J. Effect of mastitis on raw milk compositional quality. **Jornal of Veterinary Science**, v. 8, n. 3, p. 237–242, 2007.

OTTO, P. I. et al. 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. 2020a.

OTTO, P. I. et al. 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. 2020b.

OTTO, P. I. et al. 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. 2020c.

PAL, A. et al. Mutations in CD14 gene causes mastitis in different breeds of buffalo as confirmed by in silico studies and experimental validation. **ResearchSquare**, 2019.

PEI, X. H. et al. CDK Inhibitor p18INK4c Is a Downstream Target of GATA3 and Restrains Mammary Luminal Progenitor Cell Proliferation and Tumorigenesis. **Cancer Cell**, v. 15, n. 5, p. 389–401, 5 maio 2009.

PÉREZ O'BRIEN, A. M. et al. Linkage disequilibrium levels in *Bos indicus* and *Bos taurus* cattle using medium and high density SNP chip data and different minor allele frequency distributions. **Livestock Science**, v. 166, n. 1, p. 121–132, 2014.

PLOMIN, R. et al. **DNA Markers Associated with High Versus Low IQ: The IQ Quantitative Trait Loci (QTL) Project** *Behavior Genetics*. [s.l.: s.n.].

ROSATI, A.; VAN VLECK, L. D. **Estimation of genetic parameters for milk, fat, protein and mozzarella cheese production for the Italian river buffalo**

Bubalus bubalis population *Livestock Production Science*. [s.l: s.n.]. Disponível em: <www.elsevier.com/locate/livprodsci>.

SAKIHAMA, S.; KARUBE, K. **Genetic Alterations in Adult T-Cell Leukemia/Lymphoma: Novel Discoveries with Clinical and Biological Significance**. *Cancers* MDPI, , 1 maio 2022.

SARGOLZAEI, M.; CHESNAIS, J. P.; SCHENKEL, F. S. **A new approach for efficient genotype imputation using information from relatives**. [s.l: s.n.]. Disponível em: <<http://www.biomedcentral.com/1471-2164/15/478>>.

SESANA, R. C. et al. Estimates of genetic parameters for total milk yield over multiple ages in Brazilian Murrah buffaloes using different models. **Genetics and Molecular Research**, v. 13, n. 2, p. 2784–2795, 14 abr. 2014.

SINGH, A. et al. Genome Wide Scan to Identify Potential Genomic Regions Associated With Milk Protein and Minerals in Vrindavani Cattle. **Frontiers in Veterinary Science**, v. 9, 10 mar. 2022.

SLATTERY, M. L.; LUNDGREEN, A.; WOLFF, R. K. MAP kinase genes and colon and rectal cancer. **Carcinogenesis**, v. 33, n. 12, p. 2398–2408, dez. 2012.

SNYDER, M. W. et al. **Haplotype-resolved genome sequencing: Experimental methods and applications**. *Nature Reviews Genetics* Nature Publishing Group, , 18 maio 2015.

SOARES, A. C. C. et al. Multiple-trait genomewide mapping and gene network analysis for scrotal circumference growth curves in Brahman cattle. **Journal of Animal Science**, v. 95, n. 8, p. 3331, 2017.

SUMMER, A. et al. Impact of heat stress on milk and meat production. **Animal Frontiers**, v. 9, n. 1, p. 39–46, 3 jan. 2019.

TAYE, M. et al. Exploring evidence of positive selection signatures in cattle breeds selected for different traits. **Mammalian Genome**, v. 28, n. 11–12, p. 528–541, 1 dez. 2017.

TEIXEIRA, V. L.; BASTIANETTO, E.; OLIVEIRA, D. A. A. **Leite de búfala na indústria de produtos lácteos**. [s.l: s.n.]. Disponível em: <www.cbpa.org.br>.

TONHATI, H.; MENDOZA-SÁNCHEZ, G.; SESANA, R. C. **Programa de mejoramiento genético de búfalos * Buffaloes breeding program** *Revista Argentina de Producción Animal*. [s.l: s.n.].

TONHATI, H.; VASCONCELLOS, F. B.; ALBUQUERQUE, L. G. Genetic aspects of productive and reproductive traits in a Murrah buffalo herd in São Paulo, Brazil. **J. Anim. Breed. Genet**, v. 117, p. 331–336, 2000.

VANRADEN, P. M. Efficient methods to compute genomic predictions. **Journal of Dairy Science**, v. 91, n. 11, p. 4414–4423, 2008.

VERARDO, L. L. et al. Revealing new candidate genes for reproductive traits in pigs: Combining Bayesian GWAS and functional pathways. **Genetics Selection Evolution**, v. 48, n. 1, 1 fev. 2016.

VERONEZE, R. et al. Linkage disequilibrium and haplotype block structure in six commercial pig lines. v. 91, p. 3493–3501, 2013.

VERRUMA, M. R.; SALGADO, J. M. **Análise química do leite de búfala em comparação ao leite de vaca**. [s.l: s.n.].

VILLUMSEN, T. M.; JANSS, L.; LUND, M. S. The importance of haplotype length and heritability using genomic selection in dairy cattle. **Journal of Animal Breeding and Genetics**, v. 126, n. 1, p. 3–13, fev. 2009.

VISSCHER, P. M. et al. **Five years of GWAS discovery**. **American Journal of Human Genetics**, 13 jan. 2012.

WALL, J. D.; PRITCHARD, J. K. **Haplotype blocks and linkage disequilibrium in the human genome**. **Nature Reviews Genetics**, 1 ago. 2003.

WANG, H. et al. Genome-wide association mapping including phenotypes from relatives without genotypes in a single-step (ssGWAS) for 6-week body weight in broiler chickens. **Frontiers in Genetics**, v. 5, n. MAY, 2014.

WEISS, K. M.; CLARK, A. G. **Linkage disequilibrium and the mapping of complex human traits** **TRENDS in Genetics**. [s.l: s.n.]. Disponível em: <[http://tig.trends.com/0168-9525/02/\\$-see-frontmatter](http://tig.trends.com/0168-9525/02/$-see-frontmatter)>.

WELLER, J. I.; EZRA, E. Genetic analysis of the Israeli Holstein dairy cattle population for production and nonproduction traits with a multitrait animal model. **Journal of Dairy Science**, v. 87, n. 5, p. 1519–1527, 2004.

WIGGANS, G. R. et al. Genomic Selection in Dairy Cattle: The USDA Experience*. **Annual Review of Animal Biosciences**, v. 5, p. 309–327, 8 fev. 2017.

WILLIAMS, J. L. et al. **Genome assembly and transcriptome resource for river buffalo, *Bubalus bubalis* (2n = 50)**. **GigaScience** Oxford University Press, , 1 out. 2017.

XU, F. et al. Bioinformatic mining and validation of the effects of ferroptosis regulators on the prognosis and progression of pancreatic adenocarcinoma. **Gene**, v. 795, 30 ago. 2021.

ZAMORANO-ALGANDAR, R. et al. Molecular marker prediction for days open and pregnancy rate in Holstein cows managed in a warm climate. **Livestock Science**, v. 250, 1 ago. 2021.

ZHANG, Q.; SAKAMOTO, K.; WAGNER, K. U. D-type cyclins are important downstream effectors of cytokine signaling that regulate the proliferation of normal and neoplastic mammary epithelial cells. **Molecular and Cellular Endocrinology**, v. 382, n. 1, p. 583–592, 25 jan. 2014.

ZHANG, Z. et al. Best linear unbiased prediction of genomic breeding values using a trait-specific marker-derived relationship matrix. **PLoS ONE**, v. 5, n. 9, p. 1–8, 2010.

CAPÍTULO 2 - Estudo de Associação Genômica Ampla em passo único (ssGWAS) para características produtivas em bubalinos da raça Murrah

Resumo

O método de associação genômica ampla de passo único (ssGWAS) tem sido utilizado com sucesso em diversos estudos avaliando-se a produção e/ou qualidade do leite em diferentes espécies, incluindo bubalinos, contribuindo para a identificação de regiões e genes candidatos para estas características. Desta forma, o objetivo com este trabalho foi realizar estudos ssGWAS, utilizando modelos de repetibilidade e análises pós-GWAS, para identificar genes candidatos associados a características de produção e qualidade do leite em búfalos Murrah. Os dados fenotípicos utilizados são provenientes de seis fazendas de búfalos leiteiros Murrah localizadas nos estados brasileiros do Ceará, Rio Grande do Norte e São Paulo. Um total de 403.398 registros baseados em testes diários de produção de leite foram utilizados para estimar as produções acumuladas até 305 dias para as características de produção de leite (PL), produção de proteína (PP), porcentagem de proteína (PORP), produção de gordura (PG), porcentagem de gordura (PORG) e escore de células somáticas (ECS). O arquivo de genótipos original é composto por 978 animais genotipados com o chip 90K Axiom Buffalo Genotyping Array (Affymetrix/Thermo Fisher Scientific) e o modelo animal de repetibilidade unicaracterístico foi adotado para avaliação. Após análises de associação, foram listados 60 genes no total, identificados nas 10 janelas com maior porcentagem de variância para cada característica avaliada. As análises da ontologia gênica forneceram evidências das funções dos genes candidatos e suas interações com a biologia do leite, o que contribuiu para a identificação e descrição de 22 genes relacionados as características analisadas. Este estudo permitiu maior conhecimento sobre a arquitetura genética da raça avaliada e aponta genes candidatos que podem ser investigados em pesquisas adicionais com o objetivo de validar seus efeitos sobre as características de produção e qualidade de leite em búfalos Murrah.

Palavras-chave: búfalos, genes, janelas, ssGWAS

Abstract

The single-step genomic wide association method (ssGWAS) has been successfully used to identify genomic regions and candidate genes associated with milk production and/or quality traits in different species, including water buffalo. This research aimed to carry out ssGWAS studies, using repeatability models and post-GWAS analyses, to identify candidate genes associated with milk production, composition and quality traits in Murrah buffaloes. The phenotypic data used come from six Murrah dairy buffalo farms located in the Brazilian states of Ceará, Rio Grande do Norte and São Paulo. A total of 403,398 records based on daily milk production tests were used to estimate the cumulative productions over 305 days for the milk production (PL), protein production (PP), protein percentage (PORP), milk production fat (PG), percentage of fat (PORG) and somatic cell score (ECS) traits. A total of 978 animals were genotyped with the 90K Axiom Buffalo Genotyping Array chip (Affymetrix/Thermo Fisher Scientific) and the single-trait repeatability animal model was adopted for evaluation. After association analyses, a total of 60 genes were listed, identified in the 10 windows with the highest percentage of variance for each evaluated trait. Gene ontology analyzes provided evidence of candidate gene functions and their interactions with milk biology, which contributed to the identification and description of 22 genes related to the analyzed traits. This study allowed greater knowledge about the genetic architecture of the evaluated breed and points out candidate genes that can be investigated in additional research with the objective of validating their effects on the milk production, composition and quality traits in the Murrah buffaloes population.

Keywords: buffaloes, genes, windows, ssGWAS

2.1. INTRODUÇÃO

Com o aumento da demanda por produtos lácteos de búfalos, responsáveis pela produção de aproximadamente 134.791.772 toneladas de leite no mundo em 2019 (Fao, 2019), a busca por animais mais produtivos se tornou uma realidade no setor. Entretanto, a resistência dos produtores a adequar a nutrição e o manejo dos animais, paralelamente com o inadequado controle zootécnico de alguns rebanhos leiteiros no país, representam grandes desafios para o incremento desta cultura.

Características zootécnicas de interesse, como a produção e qualidade do leite, são características que se expressam várias vezes ao longo da vida dos animais e seus registros, podem ser coletados no mesmo indivíduo e utilizados para avaliar diferentes lactações, contribuem para uma estimativa mais acuradas dos parâmetros genéticos e valor genético dos animais. Essas medidas repetidas podem ser incluídas no modelo animal por meio do efeito de ambiente permanente e, quanto maior o número de registros do animal, mais precisa será sua avaliação (Mrode, 2014).

A inclusão de informações genômicas nos modelos estatísticos para avaliação genética, também apresenta grande importância na predição do mérito genético dos animais de forma mais acurada e tem contribuído com uma seleção mais criteriosa dos melhores indivíduos geneticamente (Wiggans et al., 2017). O desenvolvimento de marcadores moleculares localizados em todo o genoma do búfalo, embora recente, tem possibilitado a realização de diversos estudos genômicos, entre eles, os estudos de associação genômica ampla (GWAS) (Nascimento et al., 2021). As análises de GWAS nos permitem integrar informações de fenótipo, genótipo e pedigree para a localização de regiões genômicas e genes relacionados com as características de interesse.

Atualmente, o método de associação genômica ampla de passo único (ssGWAS) passou a ser mais implementado nas análises genômicas por apresentar maior capacidade em analisar modelos complexos, não exigir a utilização de pseudo fenótipos e por integrar informações de fenótipo, genótipo e pedigree em um único passo/análise. Esta metodologia tem sido utilizada com sucesso em diversos estudos avaliando-se a produção e/ou qualidade do leite

em diferentes espécies, incluindo bubalinos, contribuindo para a identificação de regiões e genes candidatos para estas características (Costa Barros et al., 2018b; Araujo Neto et al., 2020; Guzman et al., 2020; Lázaro et al., 2021).

Desta forma, o objetivo com este trabalho foi realizar um estudo de associação genômica ampla em único passo (ssGWAS), utilizando modelos de repetibilidade e análises pós-GWAS, para identificar genes candidatos associados a características de produção e qualidade do leite em búfalos Murrah.

2.2. MATERIAL E MÉTODOS

Dados fenotípicos

Os dados fenotípicos utilizados são provenientes de seis fazendas de búfalos leiteiros Murrah localizadas nos estados brasileiros do Ceará, Rio Grande do Norte e São Paulo, participantes do programa de controle leiteiro, coordenado pelo Departamento de Zootecnia da Universidade Estadual Paulista, campus Jaboticabal, São Paulo. Um total de 403.398 registros baseados em testes diários (TD) de produção de leite de 5077 animais foram utilizados para estimar as produções acumuladas até 305 dias. O controle de qualidade dos dados fenotípicos constituiu-se da eliminação dos animais sem data de nascimento ou parto, além de animais com menos de três informações de TD e com ausência de registros nos 70 primeiros dias de lactação, informação de grande importância para moldar a curva de lactação (Lázaro et al., 2021; Aspilcueta-Borquis et al., 2013; Tonhati et al., 2008). Todas as características avaliadas apresentam informações de até três lactações.

As características avaliadas foram produção de leite acumulada até 305 dias (PL), produção de proteína acumulada até 305 dias (PP), porcentagem de proteína acumulada até 305 dias (PORP), produção de gordura acumulada até 305 dias (PG), porcentagem de gordura acumulada até 305 dias (PORG) e escore de células somáticas acumulada até 305 dias (ECS), conforme descrito na Tabela 1. A contagem de células somáticas (CCS) foi convertido para escore

de células somáticas (ECS) pelo: $ECS = [\text{Log2} (CCS/100) + 3]$ (FENG et al., 2021).

Tabela 1. Estatística descritiva para características de produção e qualidade do leite para búfalas da raça Murrah.

Característica	Animais	N	Média	DP	CV%	GC
PL	4.544	10.159	2.019,75	742,44	36,76	218
PP	1.419	2.572	88,68	29,16	32,89	40
PORP	1.409	2.567	4,27	0,31	7,37	40
PG	1.418	2.573	128,62	46,20	35,92	40
PORG	1.305	2.567	5,97	0,81	13,57	39
ECS	1.428	2.336	4,17	1,71	40,91	44

N: número de registros; PL: Produção de leite; PP: Produção de proteína; PORO: Porcentagem de proteína; PG: Produção de gordura; PORG: Porcentagem de gordura; ECS: escore de células somáticas; DP: Desvio padrão; CV%: Coeficiente de variação em porcentagem; GC: Grupo contemporâneo.

A produção de leite foi a característica com maior número de informação, número de animais, e grupos contemporâneos (GC) entre as características avaliadas. Para as demais características (PL, PP, PORP, PG, PORG e ECS) o número de informações média FOI de 2.523, em média 1.396 animais com aproximadamente 41 grupos contemporâneos formados. Dentre as características avaliadas, os dados de ECS apresentaram maior heterogeneidade e PORP maior homogeneidade.

Para a formação dos GC, foram utilizadas as informações de fazenda (n=6), ano de parto e estação de parto (n=2), em um único valor codificado. A estação de parto foi classificada como: 1) para animais que pariram de outubro a março e 2) para animais que pariram de abril a setembro. Grupos contemporâneos com menos de três informações foram desconsiderados e dados com 3,5 desvios padrões abaixo ou acima da média também foram descartados. O arquivo de pedigree conta com 93.673 informações, destas 66.253 fêmeas (19.822 vacas com progênie) e 27.388 machos (1.385 touros com progênie).

Dados genômicos

O arquivo de genótipos original é composto por 978 animais genotipados com o chip 90K Axiom Buffalo Genotyping Array (Affymetrix/Thermo Fisher

Scientific), o qual foi submetido a análise de controle de qualidade por meio do software Plink v.1.9 (Chang et al., 2015). O material genético para genotipagem foi obtido através da coleta de pelos da cauda dos animais. A extração e purificação do DNA foi realizada com base no protocolo do Kit comercial Macherey-Nagel NucleoSpin® Tissue (Düren, Germany). Para o controle de qualidade, marcadores autossômicos, com posição conhecida e não duplicada, com call rate maior que 95%, frequência do menor alelo (MAF) maior que 3% e sem desvio extremo do equilíbrio de Hardy-Weinberg ($p > 10^{-6}$) foram selecionados. Ainda, amostras com call rate maior que 90% foram selecionados. Após o controle de qualidade, 45.690 marcadores e 960 amostras permaneceram para as análises posteriores.

Análise de associação genômica ampla em passo único (ssGWAS)

Primeiramente, foram estimados os componentes de variância para as características PL, PP, PORP, PG, PORG e ECS por meio do REMLF90 da família de programas do BLUPF90 (Misztal et al., 2022), utilizando-se modelos animal de repetibilidade unicaracterístico. Na sequência, os valores genéticos genômicos foram preditos utilizando-se a abordagem de melhor predição linear não viesada genômica de passo único (ssGBLUP). Assim, o modelo animal de repetibilidade unicaracterístico adotado foi:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{a} + \mathbf{W}\mathbf{p} + \mathbf{e}$$

em que $\mathbf{y}$ é o vetor de observações de PL, PP, PG, PORP, PORG e ECS; $\boldsymbol{\beta}$ é o vetor de efeitos fixos (grupo contemporâneo, número de ordenha e a covariável idade da búfala no momento do parto, efeito linear e/ou quadrático); $\mathbf{a}$ é o vetor de efeitos genéticos aditivos aleatórios, assumindo $\mathbf{a} \sim N(0; \mathbf{H}\sigma_a^2)$, em que σ_a^2 representa a variância genética aditiva e $\mathbf{H}$ a matriz que combina informações genômicas e de pedigree, conforme proposto por Aguilar et al. (2010); $\mathbf{p}$ é o vetor de efeito ambiente permanente aleatório, assumindo $\mathbf{p} \sim N(0; \mathbf{I}\sigma_p^2)$, em que σ_p^2 é a variância de ambiente permanente e $\mathbf{I}$ a matriz identidade; $\mathbf{e}$ é o vetor de efeitos residuais aleatórios, assumindo $\mathbf{e} \sim N(0; \mathbf{I}\sigma_e^2)$, com σ_e^2 variância residual; $\mathbf{X}$, $\mathbf{Z}$ e $\mathbf{W}$ são matrizes de incidência de $\boldsymbol{\beta}$, $\mathbf{a}$ e $\mathbf{p}$, respectivamente. A matriz $\mathbf{H}$ a matriz

que combina informações genômicas e de pedigree e sua inversa foi estimada por:

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} \end{bmatrix}$$

em que $\mathbf{A}$ é a matriz de relacionamento do numerador baseada no pedigree para todos os animais; $\mathbf{A}_{22}$ é uma matriz de relacionamento do numerador para animais genotipados; e $\mathbf{G}$ é a matriz de relacionamento genômico, calculada com base no primeiro método de Vanraden, (2008), ou seja:

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

onde $\mathbf{M}$ é uma matriz $n \times m$ (número de indivíduos por número de marcadores, respectivamente); $\mathbf{p}$ é um vetor com as frequências alélicas p_i para os marcadores e $\mathbf{D}$ corresponde a uma matriz identidade ($\mathbf{I}$) $m \times m$ quando $\mathbf{G}$ é aplicado com o mesmo peso para os marcadores.

Para mapear os *loci* que melhor explicam as diferenças genéticas para todas as características avaliadas, os efeitos do SNP foram calculados usando o software postGSf90 (<http://nce.ads.uga.edu/>; Aguilar et al., 2010) e os resultados do ssGWAS foram relatados como a porcentagem de variância genética explicada por janelas sobrepostas de 10 SNPs. A porcentagem de variância genética aditiva explicada por janela foi calculada, conforme descrito por Wang et al. (2014a). Segundo Habier et al. (2011), a utilização de janelas para a avaliação genômica tem apresentado melhores resultados quando comparada à avaliação por SNP individuais, devido ao fato de os QTLs apresentarem SNPs em desequilíbrio de ligação, capturando melhor o efeito do QTL do que um único SNP, além disso, ao utilizar janelas com 5 a 10 marcadores, temos uma diminuição na taxa de detecção de falsos positivos (Howard et al., 2015a). A porcentagem de variância genética aditiva explicada pelo *i*-ésimo conjunto de SNPs consecutivos (*i*-ésima janela SNP) foi calculada conforme descrito por WANG et al. (2014a):

$$\frac{var(a_i)}{\sigma_a^2} \times 100\% = \frac{var(\sum_{j=1}^n Z_j \hat{u}_j)}{\sigma_a^2} \times 100\%$$

em que a_i é o valor genético da i -ésima janela SNP que consiste em uma região de 10 SNPs consecutivos; $var(a_i)$ é a variância da i -ésima janela SNP; σ_a^2 é a variância genética aditiva total; Z_j é um vetor de conteúdo gênico do j -ésimo SNP para todos os indivíduos; e $\hat{u}_j$ é o efeito do j -ésimo SNP dentro da i -ésima janela.

Com o software R (R versão 4.2.1) foram construídos os gráficos “Manhattan Plot” com base no pacote “mhtplot” para demonstrar a distribuição das janelas e suas variâncias explicadas.

Análises Post-GWAS

Foram selecionadas as 10 janelas com maior porcentagem de variância explicada (TOP 10) dentro de cada característica avaliada (PL, PP, PORP, PG, PORG e ECS), com o objetivo de localizar e identificar genes posicionais com base nas coordenadas inicial e final de cada janela (Howard et al., 2015b). Para isso, foi utilizado o conjunto do genoma UOA_WB_1 (LOW et al., 2019) disponível no National Center for Biotechnology Information (NCBI) (<http://www.ncbi.nlm.nih.gov>). As TOP 10 janelas foram selecionadas com base na metodologia utilizada em estudos prévios de GWAS (Araujo et al., 2022; Otto et al., 2020c).

A lista de genes posicionais, contendo somente genes codificantes, ou seja, excluindo-se os genes sem identificação ou ncRNA (RNA não codificante) foi utilizada para a análise de ontologia gênica. Utilizando os aplicativos “ClueGO” e “CluePedia”, disponíveis para acesso dentro da plataforma Cytoscape (Bindea; Mlecnik, 2009), as redes gênicas foram construídas a partir dos processos biológicos dos genes envolvidos. Como o software ainda não contém informações sobre *Bubalus bubalis*, utilizou-se informações sobre *Bos taurus*. Redes de processos biológicos foram configuradas com base em um teste hipergeométrico unilateral e correção de Benjamini-Hochberg (Li; Tzeng, 2018), realizando a busca de genes ou conjuntos de genes de forma síncrona e procurando por um termo de *Functional Gene Ontology* (GO) relatado ou caminhos relacionados aos genes selecionados.

2.3. RESULTADOS E DISCUSSÃO

Com os resultados obtidos através da metodologia de ssGWAS com densidade de 10 SNPs por janela, foi encontrado um total médio de 3.565 janelas para as características avaliadas, janelas estas utilizadas para formação dos gráficos de Manhattan plot (Figura 1), com base no percentual da variância genética aditiva explicada para cada janela.

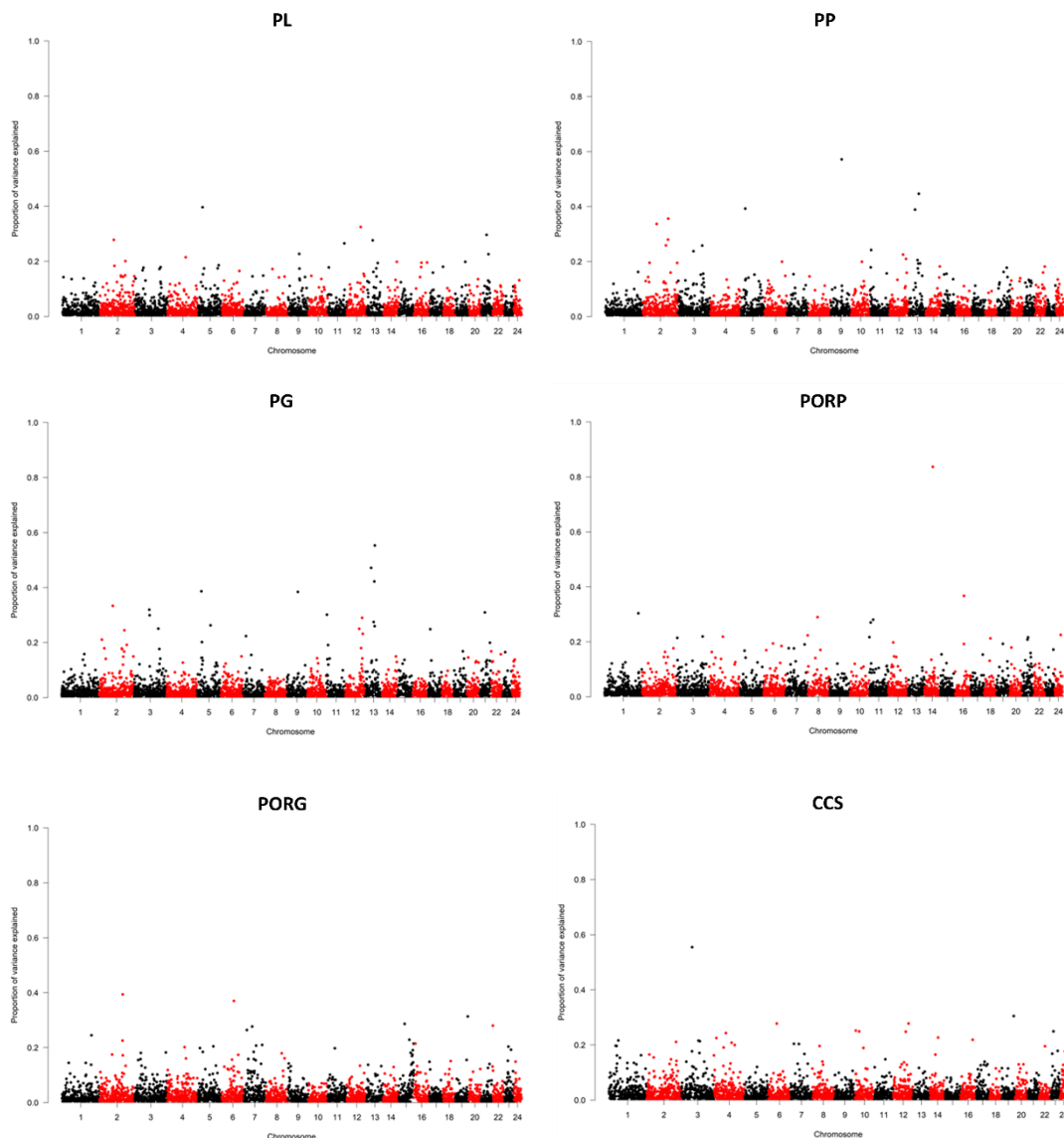


Figura 1. Manhattan Plot da variância genética aditiva explicada por janelas contendo 10 SNP para: produção de leite (PL); produção de proteína (PP); produção de gordura (PG); porcentagem de proteína (PORP); porcentagem de gordura (PORG) e escore de células somáticas (ECS).

Para melhor visualização e discussão dos resultados, 10 janelas com maior porcentagem da variância genética explicada para cada característica avaliada foram selecionadas para busca de genes posicionais, análise ontológica e discussão, podendo ser observados a seguir (Tabela 2). Apenas genes relacionados com as características estudadas foram selecionados para a presente discussão.

Tabela 2. Informações relacionadas as 10 janelas com maior porcentagem da variância explicada para as análises pós-GWAS para as características avaliadas em uma população de búfalos da raça Murrah.

Produção de Leite						
Jan	Chr	Início	Final	Tamanho	Var_e	Genes
1	5	22.882.327	23.367.722	485.395	0,40	<i>ASTN1</i>
2	12	74.171.537	74.580.524	408.987	0,32	<i>NCOA1</i>
3	21	28.197.211	28.493.766	296.555	0,29	-
4	2	74.082.787	74.417.499	334.712	0,28	-
5	13	32.055.911	32.632.965	577.054	0,28	-
6	11	87.734.368	88.114.496	380.128	0,26	<i>PIAS1</i>
7	9	59.203.043	59.664.067	461.024	0,23	<i>CD14,</i> <i>DND1</i>
8	21	38.244.251	38.710.893	466.642	0,23	-
9	4	97.720.215	97.943.648	223.433	0,21	-
10	2	136.451.314	136.826.115	374.801	0,20	-
Produção de Proteína						
Jan	Chr	Início	Final	Tamanho	Var_e	Genes
1	9	59.203.043	59.664.067	461.024	0,57	<i>CD14,</i> <i>DND1</i>
2	13	51.905.104	52.865.113	960.009	0,45	-
3	5	22.882.327	23.367.722	485.395	0,39	-
4	13	32.055.911	32.632.965	577.054	0,39	-
5	2	136.451.314	136.826.115	374.801	0,35	-
6	2	74.040.263	74.385.057	344.794	0,34	-
7	2	135.144.109	135.762.880	618.771	0,28	-
8	2	124.233.692	124.713.447	479.755	0,26	<i>CFAP221</i>
9	3	130.877.405	131.230.304	352.899	0,26	-
10	11	4.242.175	4.970.016	727.841	0,24	-
Porcentagem de Proteína						
Jan	Chr	Início	Final	Tamanho	Var_e	Genes
1	14	44.796.559	45.216.106	419.547	0,84	-
2	16	48.010.995	48.399.590	388.595	0,37	<i>SOX6</i>
3	1	182.181.195	182.495.987	314.792	0,30	-
4	8	54.583.552	55.025.566	442.014	0,29	-
5	11	20.812.433	21.360.230	547.797	0,28	<i>MAP3K9</i>
6	11	7.499.217	7.924.058	424.841	0,27	-
7	24	33.059.598	33.523.710	464.112	0,22	-

8	8	1.390.411	1.917.371	526.960	0,22	-
9	3	138.689.052	139.153.333	464.281	0,22	-
10	4	73.183.227	73.812.097	628.870	0,22	-
Produção de Gordura						
Jan	Chr	Início	Final	Tamanho	Var_e	Genes
1	13	51.905.104	52.865.113	960.009	0,55	-
2	13	32.055.911	32.632.965	577.054	0,47	-
3	13	49.134.889	50.040.735	905.846	0,42	-
4	5	22.882.327	23.367.722	485.395	0,39	<i>ASTN1</i>
5	9	59.203.043	59.664.067	461.024	0,38	-
6	2	74.040.263	74.385.057	344.794	0,33	-
7	3	81.840.631	82.380.929	540.298	0,32	-
8	21	28.197.211	28.493.766	296.555	0,31	-
9	11	1.999.363	2.471.913	472.550	0,30	<i>CD14,</i> <i>DND1</i>
10	3	83.902.255	84.250.027	347.772	0,30	-
Porcentagem de Gordura						
Jan	Chr	Início	Final	Tamanho	Var_e	Genes
1	2	124.548.843	125.016.355	467.512	0,39	<i>INHBB</i>
2	6	66.245.168	66.946.549	701.381	0,37	-
3	19	67.465.425	67.817.734	352.309	0,31	-
4	15	32.167.001	33.032.940	865.939	0,29	-
5	22	1.915.805	2.229.915	314.110	0,28	-
6	7	44.474.034	44.799.967	325.933	0,28	-
7	7	16.021.230	16.287.888	266.658	0,26	-
8	1	158.670.966	159.169.276	498.310	0,24	-
9	15	57.525.732	57.892.793	367.061	0,23	<i>STEAP3</i>
10	2	123.755.686	124.233.692	478.006	0,22	-
Escore de Células Somáticas						
Jan	Chr	Início	Final	Tamanho	Var_e	Genes
1	3	57.055.515	57.723.147	667.632	0,55	-
2	19	64.950.232	65.319.092	368.860	0,30	-
3	12	82.278.725	82.550.881	272.156	0,28	-
4	6	45.725.219	46.165.125	439.906	0,28	-
5	10	4.876.936	5.216.651	339.715	0,25	-
6	23	15.901.734	16.583.571	681.837	0,25	-
7	10	22.683.031	22.962.597	279.566	0,25	-
8	12	67.224.166	67.517.096	292.930	0,25	<i>ANTXR1</i> <i>CDK4,</i> <i>GLI1,</i>
9	4	63.963.147	64.628.584	665.437	0,24	<i>LRP1,</i> <i>SHMT2,</i> <i>SLC26A10</i>
10	14	44.827.731	45.248.556	420.825	0,23	-

Chr: Cromossomo; Início: Posição inicial da janela em pares de base; Final: Posição final da janela em pares de bases; Var_e : Proporção da variância explicada.

Neto et al, (2022), ao avaliarem uma população de búfalos da raça Murrah, encontraram regiões genômicas associadas as características de produção de leite localizadas nos cromossomos 2, 4, 9, 12 e 21. Já Lázaro et al, (2021) com base na abordagem ssGBLUP e em modelos de regressão aleatória encontraram regiões localizadas nos cromossomos 2, 4, 5 e 21 para PL, 2, 3, 5, 9 e 13 para PP, 2, 3, 5, 9 13 e 21 para PG e 4, 6, 14, 19 e 23 para ECS, corroborando com algumas posições identificadas neste trabalho.

A partir das janelas selecionadas, foram identificados 43 genes para PL, 52 genes para PP, 40 genes para PORP, 50 genes para PG, 45 genes para PORG e 93 genes para ECS (Tabela 1A a 6A). Destes, cinco genes estão envolvidos com processos biológicos ligados a PL, três genes para PP, quatro para PORP, seis para PG, dois para PORG e quatro genes estão ligados a ECS.

Produção de leite (PL)

O gene *ASTN1*, *Astrotactin 1* (Figura 2), foi localizado no cromossomo 5 e foi associado ao processo de resposta celular ao lipopeptídeo bacteriano diacil e pouco descrito na literatura sobre sua relação com a produção de leite. Apesar deste gene ser amplamente detalhado na literatura com principal atuação em funções do sistema nervoso, segundo Taye et al. (2017) ao buscar genes e regiões genômicas em animais das raças Holandesas, Hanwoo e N'Dama, o gene *ASTN1* foi detectado em regiões de QTL bovino holandês para característica de produção de leite.

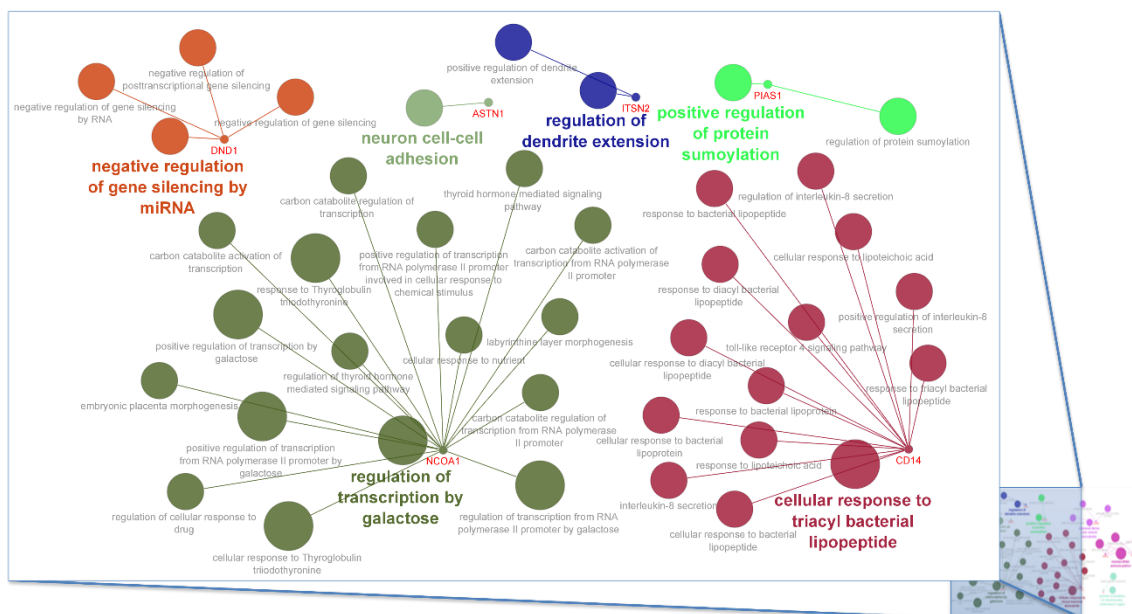


Figura 2. Rede funcional e suas interações para o PL. Círculos menores com escrita em vermelho são genes candidatos, os processos biológicos enriquecidos estão em negrito de suas respectivas cores e seus subprocessos descritos em letra preta.

O gene *CD14*, ou *Myeloid Cell-Specific Leucine-Rich Glycoprotein*, identificado ao processo biológico enriquecido de neurônio adesão célula-célula, foi localizado no cromossomo 9. Uma de suas funções é atuar como receptor transmembranar, sendo um gene alvo de seleção para produção e qualidade do leite em bovinos (Cochran et al., 2013). Segundo Pal et al. (2019), mutações deste gene em búfalos ocasionam maior susceptibilidade à mastite.

O estado sanitário do animal, mesmo não afetando diretamente a glândula mamária, pode afetar negativamente a produção e qualidade do leite. O gene DND MicroRNA-Mediated Repression Inhibitor 1 (*DND1*) foi localizado no cromossomo 9 e associado a saúde da glândula mamária. Este gene atuou como regulador da apoptose de células cancerígenas, reduzindo a apoptose de células relacionadas ao câncer (Cheng et al., 2017).

Outro gene ligado a diversos processos biológicos associados a produção de leite foi o *NCOA1* (*Nuclear Receptor Coactivator 1*), representado pelo processo de resposta à tiroglobulina triiodotironina e localizado no cromossomo 12. Este gene foi relacionado a processos de regulação da galactose, açúcar

presente no leite e quando integrado a uma molécula de açúcar forma-se a lactose (Coelho; Berry; Rubio-Gozalbo, 2015).

O gene PIAS1, localizado no cromossomo 11 esteve envolvido na fertilidade das fêmeas da raça holandesa, principalmente em vacas expostas frequentemente ao estresse térmico, inibindo a síntese de *IGF1* (Zamorano-Algandar et al., 2021). Este gene também tem participação na regulação da via JAK-STAT e a inibição de NF-kB, funções reguladoras da citosina. Também apresenta papel nas respostas imunes e apoptose nuclear (KHAN et al., 2020). Este gene é responsável pela regulação da sumoilação de proteínas (GO:0033233). Segundo Summer et al., 2019) animais submetidos a condições climáticas desfavoráveis, principalmente climas quentes e húmidos, resultam na diminuição da produção e qualidade do leite dos animais, além dos efeitos negativos relacionados ao bem-estar animal.

Produção de Proteína (PP)

Entre os genes que apresentaram processos biológicos para PP (Figura 3), os genes *CD14* e *DND1* foram novamente relatados contendo os mesmos processos biológicos enriquecidos descritos anteriormente para produção de leite. Em um estudo realizado na China com vacas holandesas, a precipitação imunológica do CD14 do leite e as digestões in vitro exibiram que o CD14 é capaz de modelar complexos com outras proteínas do leite, incluindo a α -lactalbumina, o que o protege da degradação. Além disso, o CD14 foi incluído na via de ativação de LXR / RXR associada à proteína do leite.

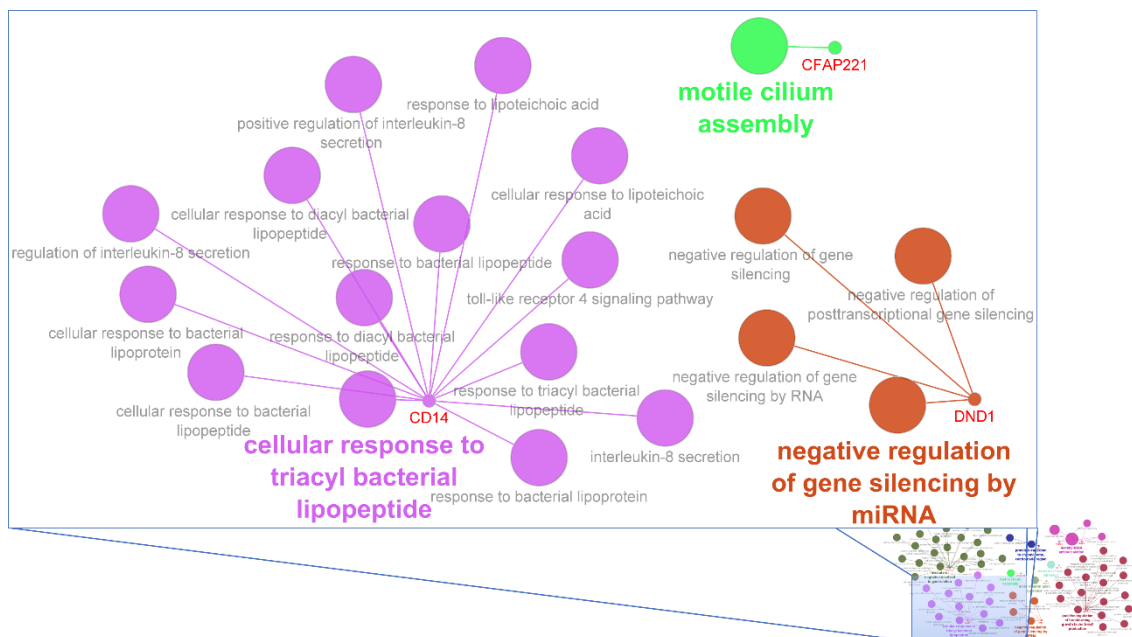


Figura 3. Rede funcional e suas interações para o PP. Círculos menores com escrita em vermelho são genes candidatos, os processos biológicos enriquecidos estão em negrito de suas respectivas cores e seus subprocessos descritos em letra preta.

O gene *CFAP221* (*Cilia And Flagella Associated Protein 221*), foi lincado ao processo de regulação positiva da produção do fator de crescimento transformador beta2 e localizado no cromossomo 2. Poucos relatos ligam este gene aos componentes do leite, segundo Singh et al. (2022) este gene apresenta grande associação com minerais presentes no leite em bovinos.

Porcentagem de Proteína

Entre os genes observados na figura 4, os genes *MAP3K9* e *SOX6* apresentaram relações com a característica avaliada. O gene *Mitogen-Activated Protein Kinase Kinase Kinase 9* (*MAP3K9*) foi associado ao processo de ativação da atividade da JUN quinase. Localizado no cromossomo 11, segundo Lu et al., (2013) foi diretamente ligado a síntese do leite, além da proliferação celular, apoptose celular e uma série de funções relacionadas a fisiologia celular. As quinases são proteínas responsáveis por modular as cascatas de transdução de sinal (Slattery; Lundgreen; Wolff, 2012).

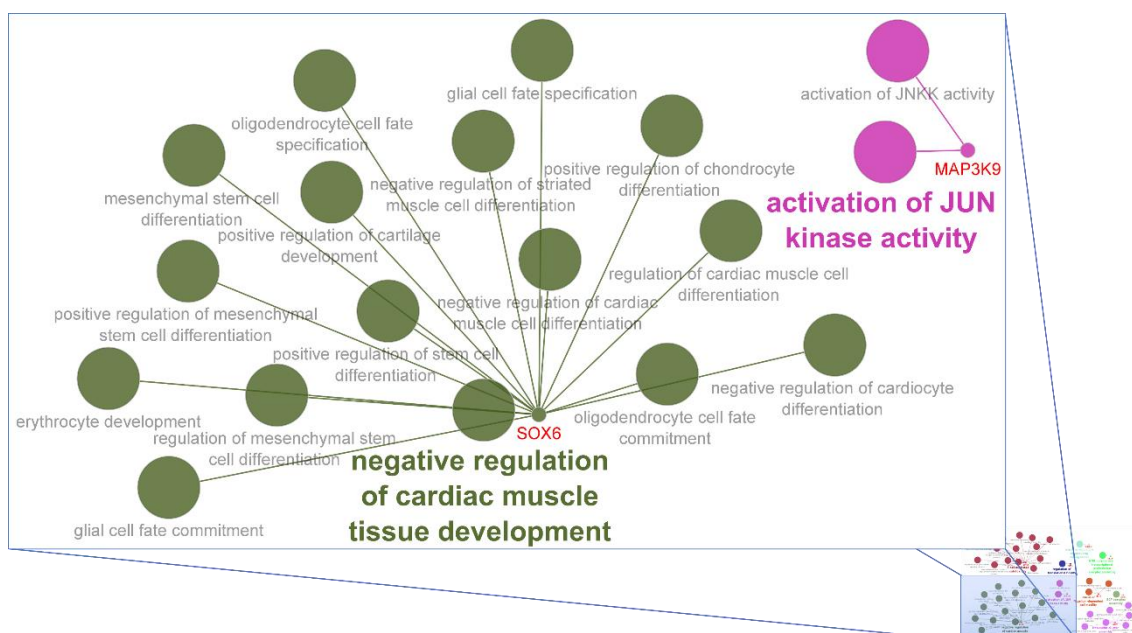


Figura 4. Rede funcional e suas interações para porcentagem de proteína. Círculos menores com escrita em vermelho são genes candidatos, os processos biológicos enriquecidos estão em negrito de suas respectivas cores e seus subprocessos descritos em letra preta.

O gene SOX6 (SRY-Box Transcription Factor 6) foi responsável pela regulação negativa do desenvolvimento do tecido muscular cardíaco e localizado no cromossomo 16. Segundo Kendrick et al., (2008) em um estudo com camundongos, o gene SOX6 expressa negativamente no epitélio mamário para receptores de hidrogênio, contribuindo com o destino celular no epitélio mamário, responsável pelo desenvolvimento e formação da glândula mamária.

Produção de Gordura

Os genes encontrados para PG ligados as características avaliadas foram *ASTN1*, *CD14* e *DND1* (Figura 5), identificados previamente na avaliação das características PL e PP. Estes genes já foram descritos anteriormente e apresentaram os mesmos processos biológicos enriquecidos anteriormente informados. Segundo Mizeranschi et al. (2021) ao avaliar genes candidatos a mastite em bovinos de duas raças leiteiras da Romênia, o gene *ASTN1* foi significativamente associado a característica, podendo influenciar a produção e qualidade do leite.

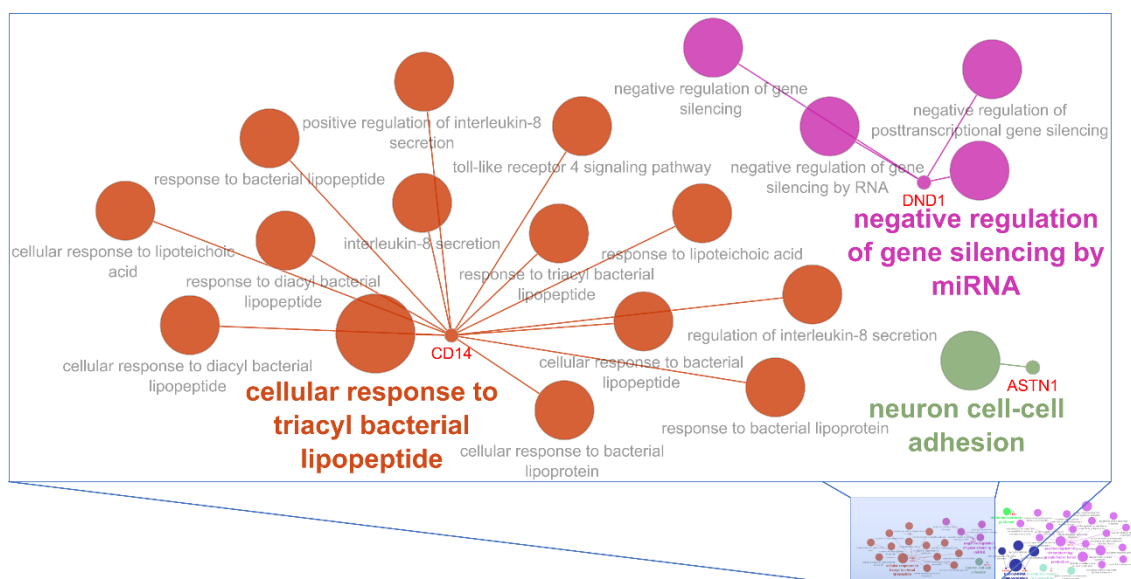


Figura 5. Rede funcional e suas interações para o PG. Círculos menores com escrita em vermelho são genes candidatos, os processos biológicos enriquecidos estão em negrito de suas respectivas cores e seus subprocessos descritos em letra preta. As redes superiores de A a C foram os genes descritos para esta característica.

Segundo Beecher et al., (2010) ao avaliar 246 animais de diferentes raças produtoras de leite na Irlanda, buscando genes associados ao escore de células somáticas apresentaram o primeiro relato de que o gene CD14 foi associado à diminuição da produção de gordura e proteína do leite.

O gene DND1, como descrito anteriormente, foi associado a saúde da glândula mamária. Ao analisar os efeitos da mastite no leite, Ogola; Shitandi; Nanua, (2007) avaliaram vacas da raça Holstein Friesian e Zebu (*Bos indicus*) localizados na província de Rift Valley, região com produção de leite do Quênia, comprovaram que animais com mastite apresentaram um aumento significativo ($p < 0,05$) no teor de ácidos graxos livres no leite juntamente com o aumento da CCS.

Porcentagem de Gordura

O *Inhibin Subunit Beta B (INHBB)*, gene localizado no cromossomo 2 e responsável por regulação da transcrição do promotor de RNA polimerase III, realiza funções de regulação hormonal relacionadas a lactação. Segundo Ma et al. (2021) o gene faz parte de um grupamento de genes distribuídos nas vias de

sinalização de liberação da gonadotrofina, hormônio responsável pelo crescimento mamário e a via de sinalização da endotelina, via atuante na produção hormonal.

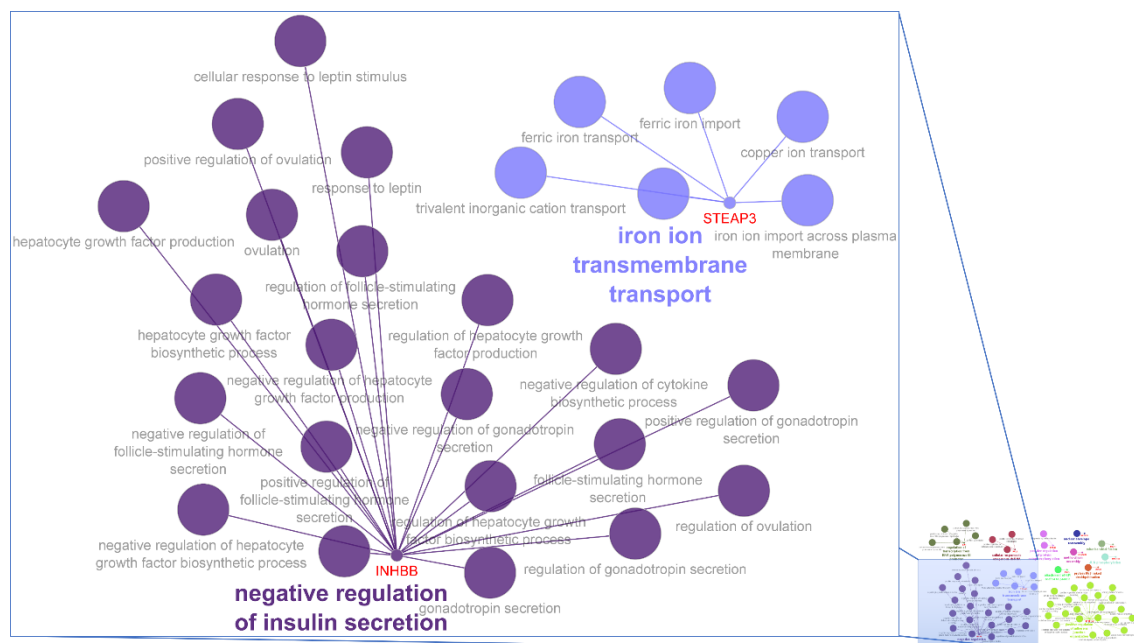


Figura 6. Rede funcional e suas interações para a porcentagem de gordura. Círculos menores com escrita em vermelho são genes candidatos, os processos biológicos enriquecidos estão em negrito de suas respectivas cores e seus subprocessos descritos em letra preta.

O gene *STEAP3 Metalloreductase*, responsável pelo transporte transmembranal de íons de ferro e localizado no cromossomo 2, é um dos possíveis genes candidatos para a regulação dos processos de armazenamento de lipídios (Bakhtiarizadeh; Alamouti, 2020). Segundo (Xu et al., (2021) o gene atua no fornecimento de elétrons para espécies reativas de oxigênio (ROS), onde os acúmulos de ROS resultam na oxidação lipídica.

Escore de Células Somáticas

O gene *ANTXR Cell Adhesion Molecule 1 (ANTXR1)* (Figura 7), localizado no cromossomo 12, foi associado a regulação negativa da montagem da matriz extracelular e segundo Connor et al. (2008), o atual gene foi mais expresso em animais submetidos a uma maior frequência de ordenha, atuando negativamente

na angiogênese em vacas multíparas. Segundo Coussens e Herb. (2002) a angiogênese contribui em processos de resposta inflamatória e inflamatória crônica.

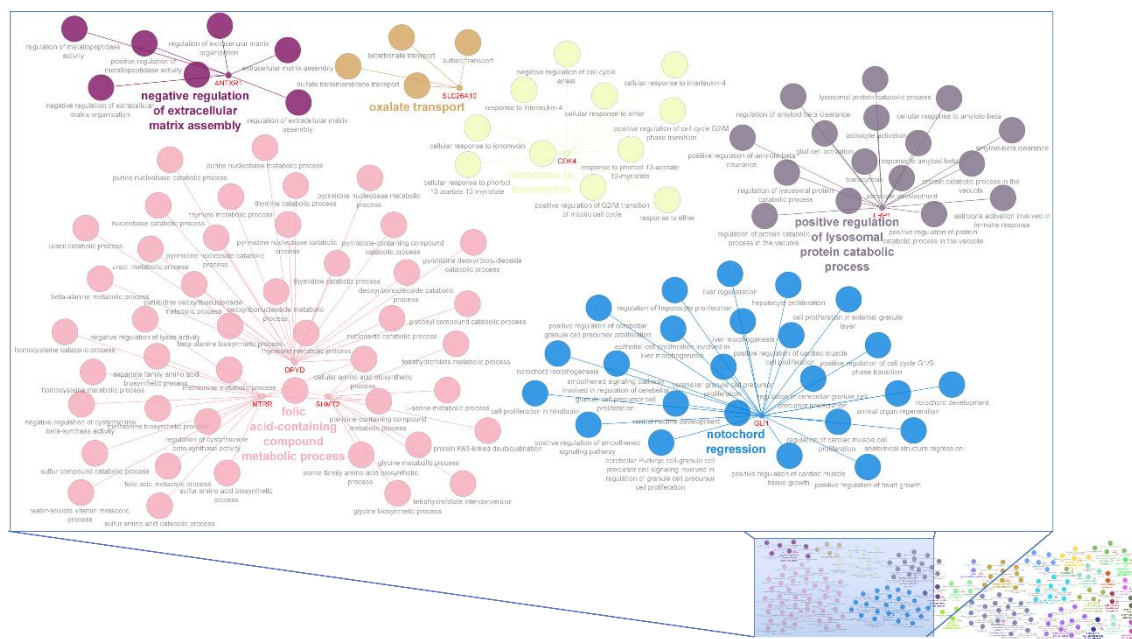


Figura 7. Rede funcional e suas interações para o ECS. Círculos menores com escrita em vermelho são genes candidatos, os processos biológicos enriquecidos estão em negrito de suas respectivas cores e seus subprocessos descritos em letra preta.

Segundo Jiang et al., (2022) o gene *CDK4*, localizado no cromossomo 4, foi expresso em vacas leiteiras com altos níveis de CCS durante o início do pós-parto através de amostras de neutrófilos no sangue e armadilha extracelular de neutrófilos (NET), onde a expressão do gene esteve diretamente ligada ao aumento da CCS no leite. O gene *CDK4*, responsável pela resposta à ionomicina, foi associado a células precursoras do câncer de mama em um estudo sobre células epiteliais mamárias (Dimri; Band; Band, 2005; Pei et al., 2009; Zhang; Sakamoto; Wagner, 2014).

O gene *GLI1* (*GLI Family Zinc Finger 1*), segundo (Hatsell; Frost, (2007), auxilia na supressão de células tronco mamárias adultas, contribuindo para o desenvolvimento da mama e foi localizado no cromossomo 4. Além disso, a super expressão deste gene pode representar problemas em gestações subsequentes (Chhabra; Booth, 2021). Segundo Ahlawat et al. (2021) o gene *LDL Receptor*

Related Protein 1 (LRP1) foi relacionado a processos do sistema imunológicos em búfalas Murrah como resposta imune inata e adaptativa, diferenciação e adesão de células imunes. Este gene foi relacionado no atual trabalho com a regulação positiva do processo catabólico da proteína lisossomal e foi localizado no cromossomo 4.

O *Serine Hydroxymethyltransferase 2 (SHMT2)* expressou o processo metabólico composto contendo ácido fólico e foi localizado no cromossomo 4. Além de ser superexpresso em células cancerígenas e alavancar o desenvolvimento de câncer de mama e pulmões (LEE et al., 2014). Já o gene *Solute Carrier Family 2 Member 10 (SLC26A10)*, localizado no cromossomo 14 foi expresso durante um desafio inflamatório, é um gene que atuante na glândula mamária e foi responsável pelo transporte de açúcar (CRISÀ et al., 2016).

Dentre os resultados obtidos e discutidos acima, relatamos vários genes identificados no ssGWAS que apresentaram importantes funções associadas as características estudadas, já consolidados em outras espécies, podendo, portando, serem classificados como genes candidatos alvo de estudos de validação, comprovando ou não sua importância com características produtivas e de qualidade de leite em búfalos da raça Murrah.

De modo geral, os 22 genes identificados no presente estudo estão diretamente ou indiretamente ligados as características avaliadas, porém todos afetando sua produção. Dentre os genes candidatos, conseguimos ressaltar os genes *CD14* e *DND1* associados as características de produção de leite, proteína e gordura, além do gene *ASTN1* associado a PL e PG. O gene *CD14* foi bem relatado em bovinos, humanos e camundongos em células ligadas ao tecido mamário, principalmente relacionado a saúde das glândulas mamarias (Cronin et al., 2012; Kiku et al., 2010) . O gene *DND1* foi relatado em outros trabalhos em populações de búfalos, sendo associado a produção de proteína (Lázaro et al., 2021) e para características morfológicas ligadas a produção de leite (Gonzalez Guzman et al., 2020). Por fim, o gene *ASTN1* esteve associado a características ligadas a produção de leite em bovinos leiteiros (Taye et al., 2017).

2.4. CONCLUSÃO

Em resumo, as análises de ssGWAS e pós-GWAS do presente estudo possibilitaram a identificação de 22 genes candidatos para as características de produção e composição do leite avaliadas na população de búfalos da raça Murrah. A análise da ontologia gênica forneceu evidências das funções dos genes candidatos e suas interações com a biologia do leite, o que contribuiu para a identificação daqueles relacionados as características analisadas. Este estudo permitiu maior conhecimento sobre a arquitetura genética da raça avaliada e aponta genes candidatos que podem ser investigados em pesquisas adicionais com o objetivo de validar seus efeitos sobre as características de produção e qualidade de leite em búfalos Murrah.

2.5. REFERÊNCIAS

- AHLAWAT, S. et al. Comparative gene expression profiling of milk somatic cells of Sahiwal cattle and Murrah buffaloes. **Gene**, v. 764, 5 jan. 2021.
- ARAUJO, A. C. et al. Haplotype-based single-step gwas for yearling temperament in american angus cattle. **Genes**, v. 13, n. 1, 1 jan. 2022.
- ASPILCUETA-BORQUIS, R. R. et al. Multiple-trait random regression models for the estimation of genetic parameters for milk, fat, and protein yield in buffaloes. **Journal of Dairy Science**, v. 96, n. 9, p. 5923–5932, set. 2013.
- BAKHTIARIZADEH, M. R.; ALAMOUTI, A. A. RNA-Seq based genetic variant discovery provides new insights into controlling fat deposition in the tail of sheep. **Scientific Reports**, v. 10, n. 1, 1 dez. 2020.
- BEECHER, C. et al. Polymorphisms in bovine immune genes and their associations with somatic cell count and milk production in dairy cattle. **BMC Genetics**, v. 11, 5 nov. 2010.
- BINDEA, G.; MLECNIK, B. ClueGO Documentation. p. 1–33, 2009.
- CHENG, F. et al. RNA-Binding Protein Dnd1 Promotes Breast Cancer Apoptosis by Stabilizing the Bim mRNA in a miR-221 Binding Site. **BioMed Research International**, v. 2017, 2017.
- CHHABRA, S. N.; BOOTH, B. W. **Asymmetric cell division of mammary stem cells**. **Cell Division** BioMed Central Ltd, , 1 dez. 2021.
- COCHRAN, S. D. et al. Discovery of single nucleotide polymorphisms in candidate genes associated with fertility and production traits in Holstein cattle. **BMC Genetics**, v. 14, 7 jun. 2013.
- COELHO, A. I.; BERRY, G. T.; RUBIO-GOZALBO, M. E. **Galactose metabolism and health**. **Current Opinion in Clinical Nutrition and Metabolic Care** Lippincott Williams and Wilkins, , 1 jul. 2015.
- CONNOR, E. E. et al. Effects of increased milking frequency on gene expression in the bovine mammary gland. **BMC Genomics**, v. 9, 31 jul. 2008.
- COUSSENS, L. M.; WERB, Z. **Inflammation and cancer**. [s.l: s.n.]. Disponível em: <www.nature.com/nature>.
- CRISÀ, A. et al. RNA-Sequencing for profiling goat milk transcriptome in colostrum and mature milk. **BMC Veterinary Research**, v. 12, n. 1, 25 nov. 2016.
- CRONIN, J. G. et al. Toll-like receptor 4 and MYD88-dependent signaling mechanisms of the innate immune system are essential for the response to lipopolysaccharide by epithelial and stromal cells of the bovine endometrium. **Biology of Reproduction**, v. 86, n. 2, 1 fev. 2012.
- DA COSTA BARROS, C. et al. Use of single-step genome-wide association studies for prospecting genomic regions related to milk production and milk

quality of buffalo. **Journal of Dairy Research**, v. 85, n. 4, p. 402–406, 1 nov. 2018.

DE ARAUJO NETO, F. R. et al. Genome-wide association studies for growth traits in buffaloes using the single step genomic BLUP. **Journal of Applied Genetics**, v. 61, n. 1, p. 113–115, 1 fev. 2020.

DE ARAUJO NETO, F. R. et al. Genomic study of the resilience of buffalo cows to a negative energy balance. **Journal of Applied Genetics**, v. 63, n. 2, p. 379–388, 1 maio 2022.

DIMRI, G.; BAND, H.; BAND, V. **Mammary epithelial cell transformation: Insights from cell culture and mouse models. Breast Cancer Research**, jul. 2005.

FENG, L. et al. The effect of exogenous melatonin on milk somatic cell count in buffalo. **Pakistan Veterinary Journal**, v. 41, n. 1, p. 152–155, 2021.

GONZALEZ GUZMAN, J. L. et al. Genome-wide association study applied to type traits related to milk yield in water buffaloes (*Bubalus bubalis*). **Journal of Dairy Science**, v. 103, n. 2, p. 1642–1650, 1 fev. 2020.

HABIER, D. et al. Extension of the bayesian alphabet for genomic selection. **BMC Bioinformatics**, v. 12, 23 maio 2011.

HATSELL, S.; FROST, A. R. Hedgehog signaling in mammary gland development and breast cancer. **Journal of Mammary Gland Biology and Neoplasia**, v. 12, n. 2–3, p. 163–173, set. 2007.

HOWARD, J. T. et al. Genome-wide association study on legendre random regression coefficients for the growth and feed intake trajectory on Duroc Boars. **BMC Genetics**, v. 16, n. 1, 30 maio 2015a.

HOWARD, J. T. et al. Genome-wide association study on legendre random regression coefficients for the growth and feed intake trajectory on Duroc Boars. **BMC Genetics**, v. 16, n. 1, 30 maio 2015b.

JIANG, L. Y. et al. Formation of Blood Neutrophil Extracellular Traps Increases the Mastitis Risk of Dairy Cows During the Transition Period. **Frontiers in Immunology**, v. 13, 27 abr. 2022.

KENDRICK, H. et al. Transcriptome analysis of mammary epithelial subpopulations identifies novel determinants of lineage commitment and cell fate. **BMC genomics**, v. 9, p. 591, 2008.

KHAN, M. Z. et al. **Role of the jak-stat pathway in bovine mastitis and milk production. Animals**MDPI AG, , 1 nov. 2020.

KIKU, Y. et al. Decrease in bovine CD14 positive cells in colostrum is associated with the incidence of mastitis after calving. **Veterinary Research Communications**, v. 34, n. 2, p. 197–203, fev. 2010.

LÁZARO, S. F. et al. Genomic studies of milk-related traits in water buffalo (*Bubalus bubalis*) based on single-step genomic best linear unbiased prediction and random regression models. **Journal of Dairy Science**, v. 104, n. 5, p. 5768–5793, 1 maio 2021.

LEE, G. Y. et al. Comparative oncogenomics identifies psmb4 and shmt2 as potential cancer driver genes. **Cancer Research**, v. 74, n. 11, p. 3114–3126, 1 jun. 2014.

LI, Z.; TZENG, C. M. Integrated analysis of miRNA and mRNA expression profiles to identify miRNA targets. In: **Methods in Molecular Biology**. [s.l.] Humana Press Inc., 2018. v. 1720p. 141–148.

LOW, W. Y. et al. Chromosome-level assembly of the water buffalo genome surpasses human and goat genomes in sequence contiguity. **Nature Communications**, v. 10, n. 1, 1 dez. 2019.

LU, L. M. et al. Proteomic and functional analyses reveal MAPK1 regulates milk protein synthesis. **Molecules**, v. 18, n. 1, p. 263–275, 2013.

MA, Y. et al. **Genetic markers associated with milk production traits in dairy cattle**. **Agriculture (Switzerland)**MDPI, , 1 out. 2021.

MISZTAL, I. et al. **BLUPF90 AND RELATED PROGRAMS (BGF90)**. [s.l.: s.n.]. Disponível em: <<http://www.ozemail.com.au/~milleraj>>.

MISZTAL, I. et al. **Manual for BLUPF90 family of programs**. [s.l.: s.n.].

MIZERANSCHI, A. E. et al. GENOME ASSOCIATION STUDY IN COW'S RESISTANCE TO MASTITIS. **LUCRĂRI ȘTIINȚIFICE MEDICINĂ VETERINARĂ**, v. 54, n. 4, p. 112–126, 2021.

MRODE, R. A. **Linear Models for the Prediction of Animal Breeding Values**. 3rd Edition ed. [s.l.] Cabi; 3rd UK ed. edição, 2014.

NASCIMENTO, A. V. et al. An updated Axiom buffalo genotyping array map and mapping of cattle quantitative trait loci to the new water buffalo reference genome assembly. **Animal Genetics**, v. 52, n. 4, p. 505–508, 1 ago. 2021.

OGOLA, H.; SHITANDI, A.; NANUA, J. Effect of mastitis on raw milk compositional quality. **Jornal of Veterinary Science**, v. 8, n. 3, p. 237–242, 2007.

OTTO, P. I. et al. 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.

PAL, A. et al. Mutations in CD14 gene causes mastitis in different breeds of buffalo as confirmed by in silico studies and experimental validation. **ResearchSquare**, 2019.

PEI, X. H. et al. CDK Inhibitor p18INK4c Is a Downstream Target of GATA3 and Restrains Mammary Luminal Progenitor Cell Proliferation and Tumorigenesis. **Cancer Cell**, v. 15, n. 5, p. 389–401, 5 maio 2009.

SINGH, A. et al. Genome Wide Scan to Identify Potential Genomic Regions Associated With Milk Protein and Minerals in Vrindavani Cattle. **Frontiers in Veterinary Science**, v. 9, 10 mar. 2022.

SLATTERY, M. L.; LUNDGREEN, A.; WOLFF, R. K. MAP kinase genes and colon and rectal cancer. **Carcinogenesis**, v. 33, n. 12, p. 2398–2408, dez. 2012.

SUMMER, A. et al. Impact of heat stress on milk and meat production. **Animal Frontiers**, v. 9, n. 1, p. 39–46, 3 jan. 2019.

TAYE, M. et al. Exploring evidence of positive selection signatures in cattle breeds selected for different traits. **Mammalian Genome**, v. 28, n. 11–12, p. 528–541, 1 dez. 2017.

TONHATI, H.; MENDOZA-SÁNCHEZ, G.; SESANA, R. C. **Programa de mejoramiento genético de búfalos * Buffaloes breeding program** Revista Argentina de Producción Animal. [s.l: s.n.].

VANRADEN, P. M. Efficient methods to compute genomic predictions. **Journal of Dairy Science**, v. 91, n. 11, p. 4414–4423, 2008.

WANG, H. et al. Genome-wide association mapping including phenotypes from relatives without genotypes in a single-step (ssGWAS) for 6-week body weight in broiler chickens. **Frontiers in Genetics**, v. 5, n. MAY, 2014.

WIGGANS, G. R. et al. Genomic Selection in Dairy Cattle: The USDA Experience*. **Annual Review of Animal Biosciences**, v. 5, p. 309–327, 8 fev. 2017.

XU, F. et al. Bioinformatic mining and validation of the effects of ferroptosis regulators on the prognosis and progression of pancreatic adenocarcinoma. **Gene**, v. 795, 30 ago. 2021.

ZAMORANO-ALGANDAR, R. et al. Molecular marker prediction for days open and pregnancy rate in Holstein cows managed in a warm climate. **Livestock Science**, v. 250, 1 ago. 2021.

ZHANG, Q.; SAKAMOTO, K.; WAGNER, K. U. D-type cyclins are important downstream effectors of cytokine signaling that regulate the proliferation of normal and neoplastic mammary epithelial cells. **Molecular and Cellular Endocrinology**, v. 382, n. 1, p. 583–592, 25 jan. 2014.

CAPÍTULO 3 – Estudo de associação genômica ampla com base em haplótipos para características produtivas em búfalos da raça Murrah

Resumo

O objetivo deste trabalho foi realizar estudo de ssGWAS com blocos de haplótipos para identificar genes associados a produção de leite, gordura e proteína, porcentagem de gordura e de proteína e escore de células somáticas, acumuladas em 305 dias em búfalos Murrah e identificar genes candidatos por meio de estudos pós-GWAS, utilizando-se enriquecimento de termos de ontologia gênica. O banco de dados engloba 403.398 registros baseados em testes diários de produção de leite, utilizados para estimar as produções acumuladas até 305 dias. Para a montagem dos blocos foi utilizado o método Big-LD implementado ao pacote “gpart” no software R, os blocos foram montados com base no LD, estimados por meio da medida r^2 . O modelo animal de repetibilidade foi usado para este estudo. Objetivando a identificação dos genes candidatos associados as características avaliadas, foram selecionados os 10 blocos (TOP 10) com maior proporção da variância genética explicada para cada característica. A metodologia de ssGWAS com blocos de haplótipos foi eficiente para a identificação de regiões onde estão localizados genes importantes para as características avaliadas, podendo essa metodologia ser utilizada em outros estudos para avaliar diferentes características em búfalos. Este foi o primeiro estudo a utilizar a metodologia de ssGWAS com blocos de haplótipos na avaliação de características produtivas em búfalos Murrah. Esta metodologia foi eficiente na identificação de genes ligados as características avaliadas, podendo ser aplicada em estudos adicionais em búfalos. As redes gênicas permitiram a identificação dos genes *ITSN2* associado a produção de leite, proteína e de gordura, *SLC26A7* para porcentagem de proteína, *DAB1* e *TG* para porcentagem de gordura e *MECOM* e *TBL1XR1* para escore de células somáticas. Nossos resultados podem ser mais explorados para validar os resultados de GWAS para as características leiteiras em búfalos da raça Murrah.

Palavras-chave: Bubalinos, gene, genômica, Murrah, ssGWAS.

Abstract

The objectives of this study were to perform a ssGWAS using haplotype block structures to identify genes associated with milk, fat and protein production, percentage of fat and protein and somatic cell score traits, accumulated in 305 days in Murrah buffaloes and to identify the most likely candidate genes through post-GWAS studies. The phenotypic data consisted of 403,398 records based on daily milk production tests. The haplotype blocks were estimated based on the linkage disequilibrium (LD) estimated by r^2 using the Big-LD program, through the “gpart” package in the R software. The single-trait repeatability animal model was used in the current study. The 10 blocks (TOP 10) with the highest proportion of genetic variance explained for each trait were selected to identify the candidate genes linked with the evaluated traits. The ssGWAS methodology with haplotype blocks allowed the identification of genomic regions, where genes with important functions associated with the analyzed traits are located. Genetic networks allowed the identification of genes ITSN2 associated with milk, protein and fat production, SLC26A7 for protein percentage, DAB1 and TG for fat percentage and MECOM and TBL1XR1 for somatic cell score. This was the first study to run ssGWAS studies using haplotype blocks structures in the evaluation of productive traits in Murrah buffaloes, showing that this is a methodology to be considered in additional studies to evaluate other interest traits in this species. Our results can be further explored to validate the GWAS results for dairy traits in Murrah buffaloes.

Keywords: Buffaloes, gene, genomics, Murrah, ssGWAS.

3.1. INTRODUÇÃO

A partir da evolução das técnicas de sequenciamento do genoma, juntamente com o aumento dos bancos de dados e poder computacional, foi possível a inclusão de informações de marcadores moleculares em estudos genômicos, tornando acessível a genotipagem de milhares de marcadores genéticos distribuídos em todo o genoma. Com estas informações, regiões genômicas podem ser identificadas e relacionadas com as características desejadas (Khatkar; Randhaw; Raadsma, 2017).

Os estudos de GWAS utilizam como base as informações de marcadores SNP pré-determinados ao longo do genoma como ferramenta para a associação entre regiões pontuais do genoma com as características fenotípicas (Korte; Farlow, 2013; Visscher et al., 2012). O sucesso da análise de GWAS é parcialmente devido ao desequilíbrio de ligação (LD) entre marcadores e *loci* de características quantitativas (QTL), permitindo a estimativa de seus efeitos (Li; Liu; Schnable, 2021).

Deste modo, a utilização de haplótipos, grupo de nucleotídeos vizinhos em que seus alelos possivelmente serão transmitidos conjuntamente para a próxima geração, passa a se tornar uma metodologia promissora para os estudos de associação, com o intuito de diminuir as variantes sem perder informações através do LD presente dentro dos blocos (CUYABANO; SU; LUND, 2014b). Segundo JOHNSON et al. (2001b) análises de associação genômica com base em blocos de haplótipos permitem captar grande parte da variância genética a partir de um pequeno número de SNPs. Apesar da importância da utilização de haplótipos, poucos estudos utilizam desta metodologia pela complexidade da sua implementação para análise dos dados. Até o momento não há relatos na literatura sobre estudos de GWAS utilizando a metodologia de blocos de haplótipos para avaliação da produção e qualidade do leite em búfalos.

Para garantir uma boa estimativa dos valores genéticos do animal com base em várias medidas de uma mesma característica na população, a incorporação de medidas repetidas no modelo passa a ser uma alternativa a ser escolhida para a análise estatística, além da simplicidade de aplicação, a necessidade de menores requisitos computacionais e o menor número de

parâmetros a serem estimados e comparados. O modelo de repetibilidade assume que as medidas repetidas são expressões da mesma característica ao longo do tempo, logo, quanto maior o número de informações, mais acurado será a avaliação (MRODE, 2014).

Os objetivos deste trabalho foram realizar estudos de ssGWAS com blocos de haplótipos para identificar genes associados a produção de leite, gordura e proteína, porcentagem de gordura e de proteína e escore de células somáticas, acumuladas em 305 dias em búfalos Murrah e identificar genes candidatos por meio de estudos pós-GWAS, incluindo enriquecimento de termos ontologia gênica.

3.2. MATERIAL E MÉTODOS

Banco de Dados

O banco de dados contendo as informações fenotípicas de produção de leite, gordura, proteína e contagem de células somáticas foi fornecido pelo Departamento de Zootecnia da Universidade Estadual Paulista, campus Jaboticabal, São Paulo. Os dados são provenientes de seis fazendas localizadas nos estados brasileiros do Ceará, Rio Grande do Norte e São Paulo, participantes do programa de controle leiteiro. O banco de dados engloba 403.398 registros baseados em testes diários de produção de leite formados por 5077 animais foram utilizados neste estudo. As informações genealógicas apresentam 93.673 informações, 66.253 fêmeas e 27.388 machos, contendo 19.822 vacas com progênie e 1.385 touros com progênie.

O controle de qualidade dos dados fenotípicos foi estabelecido através da exclusão de animais com ausência de informações de data de nascimento e parto, acompanhado da eliminação de animais com registros de testes diários (TD) iniciados a partir dos 70 dias de lactação e animais com menos de três informações de TD. Informações obtidas no período inicial da lactação apresentam grande contribuição para delinear corretamente a curva de lactação (LÁZARO et al., 2021; ASPILCUETA-BORQUIS et al., 2013; TONHATI et al., 2008). Os dados foram convertidos em valores acumulados aos 305 dias, sendo

estes produção de leite (PL), produção de gordura (PG), produção de proteína (PP), porcentagem de gordura (PORG), porcentagem de proteína (PORP) e escore de células somáticas (ECS) podendo ser observados na Tabela 1. A contagem de células somáticas (CCS) foi convertido para ECS pelo $ECS = [\text{Log2} (CCS/100) + 3]$ (FENG et al., 2021). Todas as características avaliadas possuem mensurações de até três lactações.

Tabela 1. Estatística descritiva para características de produção e qualidade do leite para búfalas da raça Murrah.

Característica	Animais	N	Média	DP	CV%	GC
PL	4.544	10.159	2.019,75	742,44	36,76	218
PP	1.419	2.572	88,68	29,16	32,89	40
PORP	1.409	2.567	4,27	0,31	7,37	40
PG	1.418	2.573	128,62	46,20	35,92	40
PORG	1.305	2.567	5,97	0,81	13,57	39
ECS	1.428	2.336	4,17	1,71	40,91	44

PL: Produção de leite, PP: Produção de proteína; PORO: Porcentagem de proteína; PG: Produção de gordura; PORG: Porcentagem de gordura; ECS: escore de células somáticas; DP: Desvio padrão; CV%: Coeficiente de variação em porcentagem; GC: Grupo contemporâneo.

Dentre as características avaliadas a produção de leite contém o maior número de informações. Entre as características PP, PORP, PG, PORG e ECS temos valores aproximados entre si, o número de animais utilizados foi de aproximadamente 1.396, cerca de 2.523 informações e a formação média de 41 grupos contemporâneos. A característica PORP apresentou informações mais homogêneas, já ECS maior heterogeneidade entre as informações.

Para a formação dos grupos contemporâneos (GC), foram utilizadas as informações de fazenda (n=6), ano de parto e estação de parto (n=2), em um único valor codificado. A estação de parto foi classificada como: 1) para animais que pariram de outubro a março e 2) para animais que pariram de abril a setembro. GC com menos de três informações foram desconsiderados e dados com 3,5 desvios padrões abaixo ou acima da média também foram descartados.

Dados genômicos

O arquivo de genótipos original é composto por 978 animais genotipados com o chip 90K Axiom Buffalo Genotyping Array (Affymetrix/Thermo Fisher Scientific), o qual foi submetido a análise de controle de qualidade por meio do software Plink v.1.9 (Chang et al., 2015). O material utilizado para genotipagem foi obtido através da coleta de pelos da cauda dos animais. A extração e purificação foram conduzidos com base no protocolo do Kit comercial Macherey-Nagel NucleoSpin® Tissue (Düren, Germany). No controle de qualidade marcadores autossômicos, com posição conhecida e não duplicada, com call rate maior que 95%, frequência do menor alelo (MAF) maior que 3% e sem desvio extremo do equilíbrio de Hardy-Weinberg ($p > 10^{-6}$) foram selecionados. Ainda, amostras com call rate maior que 90% foram selecionados. Após o controle de qualidade, 45.690 marcadores e 960 amostras permaneceram para as análises posteriores.

Dentre os animais genotipados, 749 búfalas são fenotipadas, 735 búfalas estão presentes na avaliação de produção de leite, 555 búfalas para produção de proteína, 553 búfalas para porcentagem de proteína, 556 búfalas para produção de gordura, 520 búfalas para porcentagem de gordura e 539 búfalas para escore de células somáticas, considerando informações de búfalas com até três lactações.

Montagem dos Blocos de Haplótipos

Para iniciar o processo de montagem dos blocos, foi realizado o faseamento dos SNPs dos genótipos através do software FImpute v3.0 (SARGOLZAEI; CHESNAIS; SCHENKEL, 2014). Com os genótipos já faseados, através do método Big-LD (KIM et al., 2018) implementado ao pacote “gpart” (KIM et al., 2019) no software R (R versão 4.2.1), os blocos foram montados com base no LD, estimados por meio da medida r^2 (HILL; ROBERTSON, 1968). O Big-LD foi utilizado por ser uma metodologia precisa na estimação dos hotspots de recombinação e eficiente computacionalmente (ARAUJO et al., 2022). A construção dos blocos foi baseada nos valores médios de LD estimados por r^2 , sendo obtido por meio de:

$$r^2 = \frac{D_{ij}^2}{p_1 p_2 q_1 q_2}$$

em que $D_{ij} = P_{ij} - p_i q_j$ é a medida de LD, conforme descrito por LEWONTIN (1964); P_{ij} é a frequência do haplótipo que consiste nos alelos i e j ; p_i é a frequência do alelo i ; q_j é a frequência do alelo j ; e $p_1 p_2 q_1 q_2$ é o produto das 4 frequências alélicas em ambos os *loci*. Para computar blocos com nível de recombinação, utilizou-se um limite de LD (0,5), pois os QTL's podem apresentar diferentes estruturas genéticas (ARAUJO et al., 2022).

Análise de associação genômica ampla em passo único (ssGWAS)

Inicialmente foram estimados os componentes de variância e predição dos valores genéticos para as características PL, PP, PORP, PG, PORG e ECS por meio do REMLF90 e BLUPF90 da família de programas do BLUPF90 (MISZTAL et al., 2022), respectivamente. Para realizar o estudo de ssGWAS com haplótipos, o arquivo de entrada com os SNPs foi recodificado para pseudo-SNPs (haplótipos convertidos para o formato SNP bialélico), descrito anteriormente. O seguinte modelo animal de repetibilidade foi usado:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{a} + \mathbf{W}\mathbf{p} + \mathbf{e}$$

em que $\mathbf{y}$ é o vetor de observações de PL, PP, PG, PORP, PORG e ECS; $\boldsymbol{\beta}$ é o vetor de efeitos fixos (grupo contemporâneo, número de ordenha e a covariável idade da búfala no momento do parto, efeito linear e/ou quadrático); $\mathbf{a}$ é o vetor de efeitos genéticos aditivos aleatórios; $\mathbf{p}$ é o vetor de efeito ambiente permanente aleatório; $\mathbf{e}$ é o vetor de efeitos residuais aleatórios; $\mathbf{X}$, $\mathbf{Z}$ e $\mathbf{W}$ são matrizes de incidência de $\boldsymbol{\beta}$, $\mathbf{a}$ e $\mathbf{p}$, respectivamente. Foi assumindo $\mathbf{a} \sim N(0; \mathbf{H}\sigma_a^2)$, $\mathbf{p} \sim N(0; \mathbf{I}\sigma_p^2)$ e $\mathbf{e} \sim N(0; \mathbf{I}\sigma_e^2)$, em que σ_a^2 , σ_p^2 e σ_e^2 representam a variância genética aditiva, variância genética aditiva de ambiente permanente e variância residual, respectivamente. $\mathbf{I}$ é a matriz identidade e $\mathbf{H}$ é a matriz que combina informações genômicas e de pedigree, conforme proposto por Aguilar et al. (2010), onde sua inversa é estimada por:

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} \end{bmatrix}$$

onde **A** é a matriz de relacionamento do numerador baseada no pedigree para todos os animais; **A₂₂** é uma matriz de relacionamento do numerador para animais genotipados; e **G** é a matriz de relacionamento genômico, calculada com base no primeiro método de Vanraden, (2008), ou seja:

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

em que **M** é uma matriz $n \times m$ (número de indivíduos por número de marcadores, respectivamente); **p** é um vetor com as frequências alélicas p_i para os marcadores e **D** corresponde a uma matriz identidade (**I**) $m \times m$ quando **G** é aplicado com o mesmo peso para os marcadores.

Posteriormente a predição do GEBV, os efeitos dos pseudo-SNPs foram calculados utilizando o software postGSf90 (<http://nce.ads.uga.edu/>; Aguilar et al., 2010), ou seja:

$$\mathbf{g} = \mathbf{D}(\mathbf{M} - 2\mathbf{p})'\mathbf{G}^{-1}\hat{\mathbf{u}}$$

onde **g** é o vetor de efeito do marcador, $\mathbf{G}^{-1}$ é a inversa da matriz **G**, **u** é o vetor de GEBV previsto e os outros vetores e matrizes já foram descritas anteriormente. O software postGSf90 também foi utilizado para estimar a porcentagem de variância genética aditiva explicada para cada pseudo-SNP, estimados por:

$$VEM\%_i = \frac{V(g_i)}{\sigma_g^2} \times 100 = \frac{2p_i(1-p_i)\hat{\sigma}_i^2}{\sigma_g^2} \times 100$$

onde $VEM\%_i$ é a porcentagem da variância genética aditiva explicada pelo i-ésimo pseudo-SNP, $V(g_i)$ é a variância genética aditiva explicada pelo i-ésimo pseudo-SNP, $\hat{\sigma}_i^2$ é o quadrado do efeito de substituição alélica estimado, e os demais componente foram anteriormente descritos. Como os blocos de haplótipos são formados por pseudo-SNPs, a porcentagem da variância explicada para cada bloco de haplótipo foi calculada por:

$$VEM\%_j = \sum_{i=1}^{n_j} VEM\%_{ij}$$

onde $VEM\%_j$ é a porcentagem da variância genética aditiva explicada pelo j-ésimo bloco de haplótipos, n_j é o número de alelos do haplótipo (pseudo-SNP)

presentes no j -ésimo bloco de haplótipo e $VEM\%_{ij}$ são as variâncias explicadas pelo i -ésimo pseudo-SNP presentes no j -ésimo bloco de haplótipo.

Para demonstrar a distribuição dos blocos e suas variâncias explicadas, gráficos “Manhattan Plot” foram construídos com o software R (R versão 4.2.1) com base no pacote “mhtplot”.

Análises Post-GWAS

Objetivando a identificação dos genes candidatos associados as características avaliadas, foram selecionados os 10 blocos (TOP 10) com maior proporção da variância genética explicada para cada característica (PL, PP, PG, PORP, PORG e ECS). O método de seleção dos blocos foi definido com base em estudo anteriormente realizado, que obteve sucesso na identificação de genes associados para características de temperamento (Araujo et al., 2022). O banco de dados contendo o genoma UOA_WB_1 (Low et al., 2019) disponível no National Center for Biotechnology Information (NCBI) (<http://www.ncbi.nlm.nih.gov>) foi utilizado para identificar os genes nas regiões selecionadas.

Somente genes codificantes foram selecionados para as análises de ontologia gênica. Através da plataforma Cytoscape (Bindea et al., 2009), por meio dos aplicativos “ClueGO” e “CluePedia”, foram construídas as redes de processos biológicos com base na análise de ontologia dos genes identificados. Como o aplicativo ainda não possui informações sobre a espécie *Bubalus bubalis*, o banco de dados contendo informações de *Bos taurus* foi utilizado para as análises de ontologia. A estatística aplicada para correção foi o teste hipergeométrico unilateral Benjamini-Hochberg (Li; Tzeng, 2018), buscando as ontologias gênicas (GO) realizando buscas de um ou mais genes sincronicamente e caminhos relacionados aos genes selecionados.

3.3. RESULTADOS E DISCUSSÃO

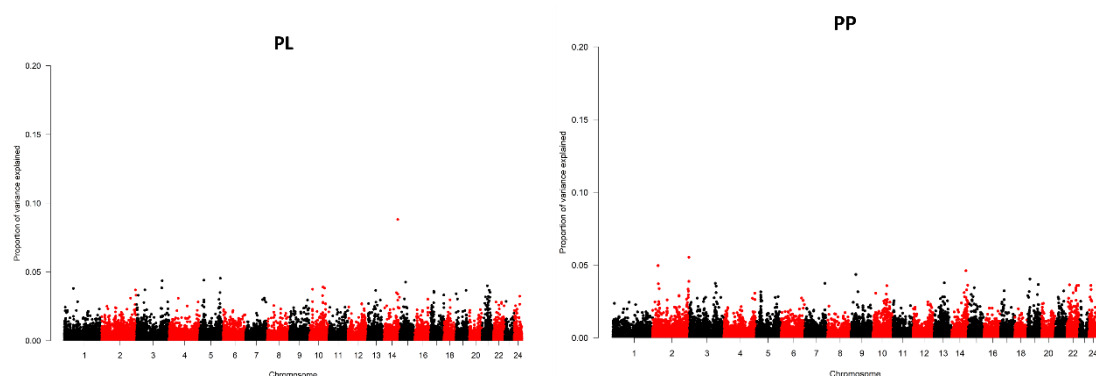
De acordo com as soluções encontradas com a abordagem ssGWAS com base em haplótipos, obtivemos 12.082 blocos de haplótipos e 36.096 SNPs não clusterizados (ncSNP) (Tabela 2), ou seja, SNPs não atribuídos a nenhum bloco de haplótipo (Li; Sung; Liu, 2007). Tanto os blocos quanto os ncSNP foram selecionados com base nos valores de variância genética aditiva explicada, obtidos através do postGSF90 para gerar os Manhattan Plot através dos pacotes “grDevices” e “qqman”, presentes para download no software R (R version 4.1.2).

Tabela 2. Tabela descritiva dos blocos de haplótipos com desequilíbrio de ligação de 0.5 após controle de qualidade.

Informações	LD 0.5
Número de SNPs não clusterizados	36.096
Número de SNPs clusterizados	12.082
Número mínimo de SNPs por bloco	2
Número máximo de SNPs por bloco	8
Tamanho mínimo dos blocos (Pb)	27
Tamanho máximo dos blocos (Mb)	0,14
DP médio do tamanho dos blocos (Mb)	0,01

LD: Desequilíbrio de Ligação, SNP: Marcadores genéticos DP: Desvio Padrão, Pb: Par de bases, Mb: Mega pares de base.

Após a montagem dos gráficos Manhattan Plot (Figura 1), conforme descrito em estudos anteriores (Begum et al., 2016; Marín-garzón et al., 2021; Otto et al., 2020c), foram selecionadas os top 10 blocos de haplótipos com maior percentual de variância genética aditiva explicada para todas as características.



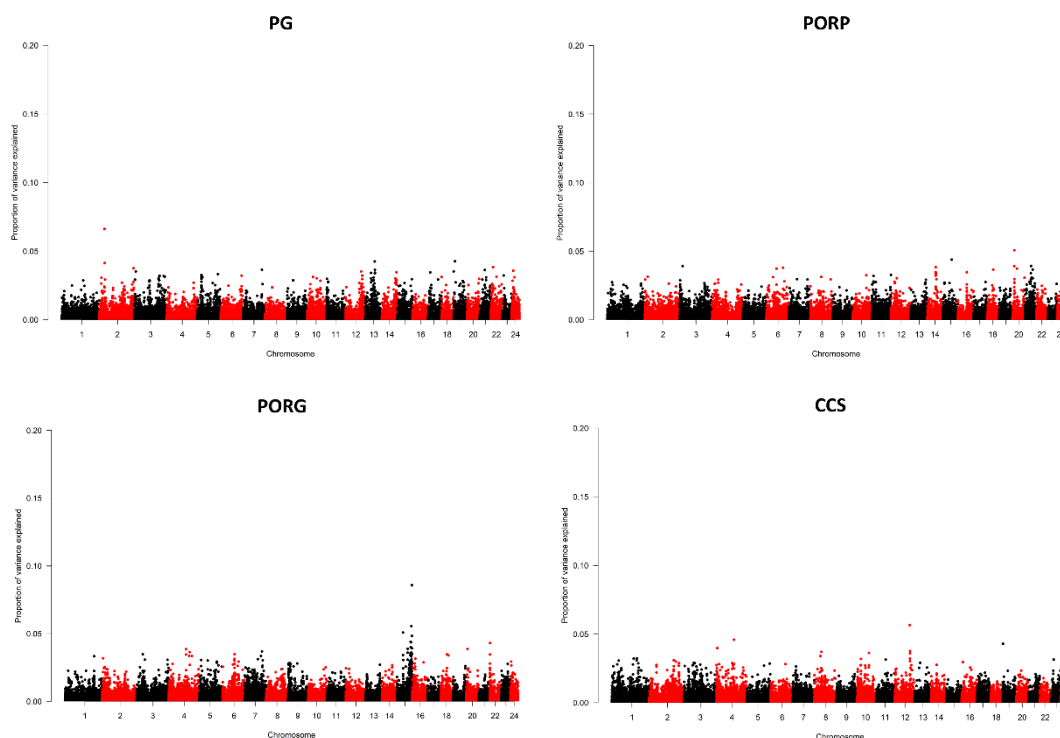


Figura 1. Manhattan Plot da variância genética aditiva explicada por haplótipos para: produção de leite (PL); produção de proteína (PP); produção de gordura (PG); porcentagem de proteína (PORP); porcentagem de gordura (PORG) e escore de células somáticas (ECS).

Para melhor visualização e discussão dos resultados, apenas os 10 blocos (enumerados de 1 a 10, em ordem decrescente com base na variância genética explicada por blocos) com maior variância explicada para cada característica avaliada foram selecionados para busca de genes, análise ontológica e discussão, podendo ser observados a seguir (Tabela 3). Apenas os genes relacionados com as características estudadas foram selecionados e discussão no presente trabalho.

Tabela 3. Informações relacionadas aos blocos com maior porcentagem da variância explicada para as análises pós-GWAS para todas as características

Produção de Leite							
Hap	CH R	SNP	Pos Inicial	Pos Final	Tamanho	Var_e	Genes
bloco 1 ¹	17	2	17.883.840	17.911.336	27.496	0,07	-
bloco 2 ²	10	2	14.870.334	14.905.564	35.230	0,07	-
bloco 3	17	2	71.927.676	71.951.418	23.742	0,06	-
bloco 4 ³	12	2	74.453.460	74.482.613	29.153	0,05	<i>ITSN2</i>
bloco 5 ⁴	21	2	28.220.032	28.240.175	20.143	0,05	-

bloco 6 ⁵	5	2	23.339.586	23.367.722	28.136	0,04	-
bloco 7 ⁶	21	2	28.325.698	28.351.030	25.332	0,04	-
bloco 8	9	2	87.901.882	87.911.410	9.528	0,04	-
bloco 9	14	3	65.222.836	65.261.082	38.246	0,04	-
bloco 10 ⁷	13	2	32.221.326	32.251.064	29.738	0,04	-

Produção de Proteína

Hap	CHR	SNP	Pos Inicial	Pos Final	Tamanho	Var_e	Genes
bloco 1	19	2	11.077.957	11.104.688	26.731	0,06	-
bloco 2 ¹	17	2	17.883.840	17.911.336	27.496	0,06	-
bloco 3 ⁸	13	2	52.712.877	52.747.153	34.276	0,05	-
bloco 4 ⁹	22	2	7.255.648	7.283.121	27.473	0,05	-
bloco 5 ³	12	2	74.453.460	74.482.613	29.153	0,05	<i>ITSN2</i>
bloco 6 ²	10	2	14.870.334	14.905.564	35.230	0,05	-
bloco 7 ⁷	13	2	32.221.326	32.251.064	29.738	0,04	-
bloco 8 ⁵	5	2	23.339.586	23.367.722	28.136	0,04	-
bloco 9	20	2	2.947.635	2.968.271	20.636	0,04	-
bloco 10	19	2	53.640.371	53.667.741	27.370	0,04	-

Porcentagem de Proteína

Hap	CHR	SNP	Pos Inicial	Pos Final	Tamanho	Var_e	Genes
bloco 1	20	3	10.243.689	10.269.523	25.834	0,05	-
bloco 2 ⁴	21	2	28.220.032	28.240.175	20.143	0,05	-
bloco 3	15	2	8.980.779	8.994.091	13.312	0,05	<i>SLC26A7</i>
bloco 4	11	2	101.755.349	101.789.795	34.446	0,04	-
bloco 5	24	2	33.378.098	33.401.488	23.390	0,04	-
bloco 6	12	2	20.349.856	20.378.216	28.360	0,04	-
bloco 7	7	2	101.951.592	101.977.604	26.012	0,04	-
bloco 8	4	2	27.490.121	27.526.234	36.113	0,04	-
bloco 9	22	2	31.824.785	31.859.201	34.416	0,04	-
bloco 10	12	2	27.013.203	27.041.472	28.269	0,04	-

Produção de Gordura

Hap	CHR	SNP	Pos Inicial	Pos Final	Tamanho	Var_e	Genes
bloco 1 ⁸	13	2	52.712.877	52.747.153	34.276	0,07	-
bloco 2 ⁷	13	2	32.221.326	32.251.064	29.738	0,06	-
bloco 3	3	2	82.001.158	82.031.525	30.367	0,06	-
bloco 4 ⁶	21	2	28.325.698	28.351.030	25.332	0,05	-
bloco 5 ⁵	5	2	23.339.586	23.367.722	28.136	0,05	-
bloco 6 ⁴	21	2	28.220.032	28.240.175	20.143	0,05	-
bloco 7	11	2	2.140.106	2.173.979	33.873	0,05	-
bloco 8 ³	12	2	74.453.460	74.482.613	29.153	0,05	<i>ITSN2</i>
bloco 9 ¹	17	2	17.883.840	17.911.336	27.496	0,05	-
bloco 10 ⁹	22	2	7.255.648	7.283.121	27.473	0,05	-

Porcentagem de Gordura							
Hap	CHR	SNP	Pos Inicial	Pos Final	Tamanho	Var_e	Genes
bloco 1	4	2	93.949.244	93.983.912	34.668	0,07	-
bloco 2	6	2	66.482.572	66.503.767	21.195	0,07	-
bloco 3	4	2	114.708.327	114.733.026	24.699	0,06	-
bloco 4	15	2	74.043.693	74.061.982	18.289	0,06	TG
bloco 5	20	2	7.943.940	7.973.300	29.360	0,05	-
bloco 6	5	2	47.370.665	47.399.072	28.407	0,05	-
bloco 7	5	3	11.277.275	11.334.109	56.834	0,05	-
bloco 8	16	2	62.932.274	62.953.872	21.598	0,04	-
bloco 9	6	2	87.417.070	87.439.439	22.369	0,04	DAB1
bloco 10	5	2	105.145.310	105.165.625	20.315	0,04	-
Escore de Células Somáticas							
Hap	CHR	SNP	Pos Inicial	Pos Final	Tamanho	Var_e	Genes
bloco 1	12	2	82.278.725	82.284.980	6.255	0,06	-
bloco 2	4	2	4.791.027	4.821.686	30.659	0,06	-
bloco 3	1	2	142.558.386	142.595.355	36.969	0,05	MECOM
bloco 4	1	3	134.669.557	134.744.473	74.916	0,05	TBL1XR1
bloco 5	10	2	45.887.118	45.910.218	23.100	0,05	-
bloco 6	3	3	66.131.195	66.174.010	42.815	0,04	-
bloco 7	7	2	17.719.204	17.743.720	24.516	0,04	-
bloco 8	14	2	30.744.325	30.779.491	35.166	0,04	-
bloco 9	24	2	17.788.040	17.817.079	29.039	0,04	-
bloco 10	3	3	57.099.594	57.135.931	36.337	0,04	-

Hap: Haplótipo; CHR: Cromossomo; Pos Inicial: Posição inicial do bloco em pares de base; Pos Final: Posição final do bloco em pares de bases; Var_e : Proporção da variância explicada; ¹ a ⁹: Blocos que se repetiram entre as características.

Alguns haplótipos foram identificados em diferentes características (Tabela 3 números sobrescritos de 1 a 9), os blocos presentes entre as regiões 17.883.840 a 17.911.336, 74.453.460 a 74.482.613, 23.339.586 a 23.367.722 e 32.221.326 a 32.251.064 pb foram encontrados na avaliação das características PL, PP e PG; os blocos localizados nas regiões 14.870.334 a 14.905.564 e 28.325.698 a 28.351.030 pb repetiram-se para PL e PP; o bloco presente entre 28.220.032 e 28.240.175 pb foi identificado para as características PL, PORP e PG; e, por fim, o bloco da região 52.712.877 a 52.747.153 pb foi localizado para as características PP e PG.

Para buscar os genes, utilizou-se o banco de dados de búfalos da água “UOA_WB_1”, disponível no National Center for Biotechnology Information (NCBI) (<http://www.ncbi.nlm.nih.gov>, acessado em 25 de julho de 2022). A partir dos blocos selecionadas, foram encontrados sete genes para PL, sete para PP, nove genes para PORP, sete para PG, 10 para PORG e seis genes para ECS, totalizando 46 genes (Tabela 7A a 12A).

Para melhor visualização dos resultados, a discussão foi separada por característica e, ao final, os resultados foram avaliados de modo geral.

Produção de leite

Como observado na Figura 2, foi identificado apenas o gene *ITSN2* para PL, localizado no cromossomo 14. O gene *Intersectin 2 (ITSN2)* foi associado a saúde da glândula mamária e foi associado ao processo da regulação positiva da extensão dendrítica. Em estudo com dados humanos, este gene foi associado a invasão celular e remodelação da actina, além de contribuir negativamente com a saúde mamária, estando ligado a tumores na mama (Kropyvko et al., 2017).

correlações genéticas entre as características. Segundo Garcia-Cortés et al. (1995) ao avaliar bovinos da rala Holandês na Espanha, a correlação entre PL e PP entre as três lactações variou de 0,70 e 0,91.

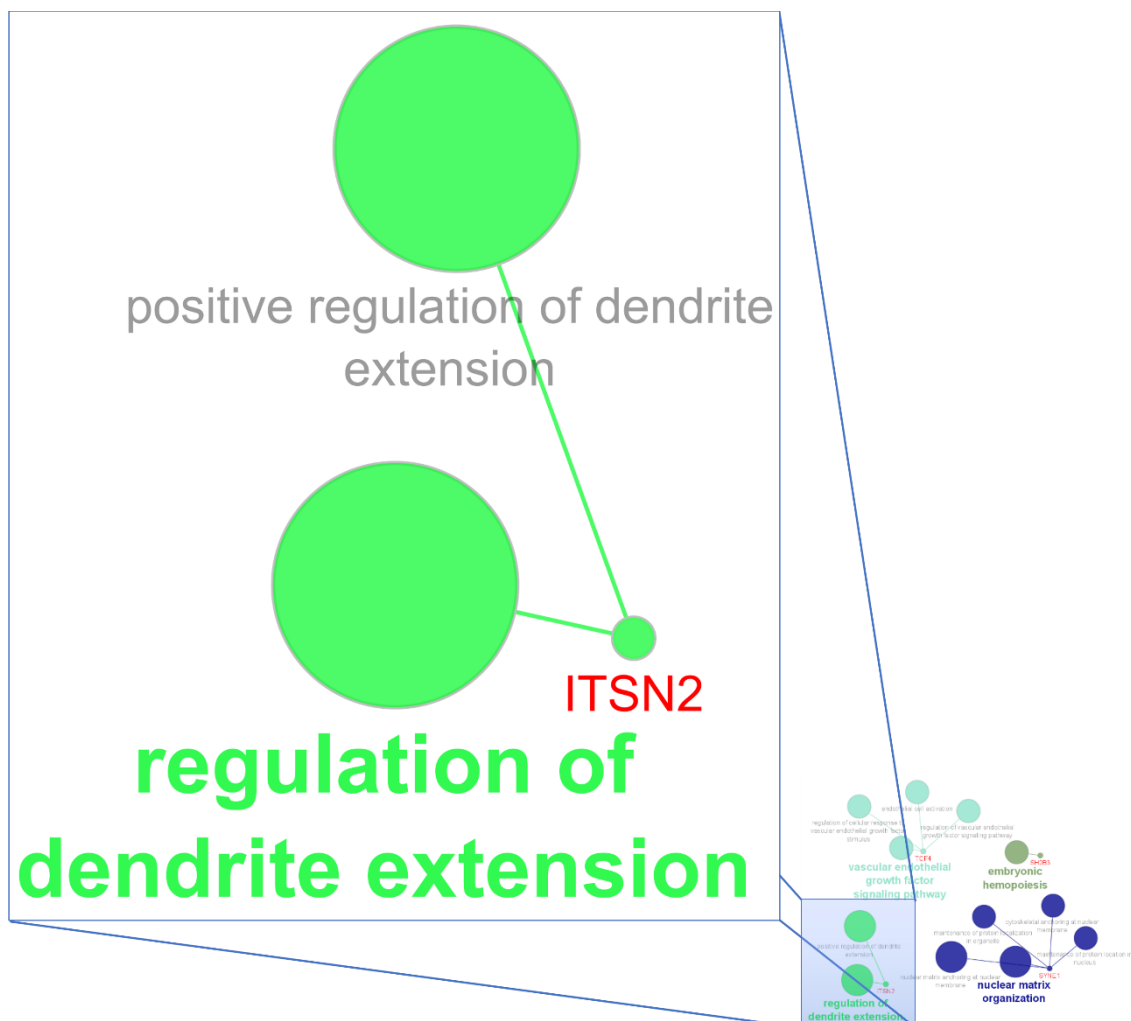


Figura 3. Rede funcional e suas interações para a produção de proteína. Círculos menores com escrita em vermelho são genes candidatos, os processos biológicos enriquecidos estão em negrito de suas respectivas cores e seus subprocessos descritos em letra preta.

O gene ITSN2, localizado no cromossomo 14 e associado ao processo enriquecido de regulação positiva da extensão dendrítica, foi descrito na literatura associado a infecções mamárias (Kropyvko et al., 2017). Segundo Gonçalves et al. (2018) ao avaliar vacas holandesas provenientes de rebanhos localizados no Centro-Oeste do estado de São Paulo, animais com infecção intramamária apresentaram aumento na concentração de proteína no leite quando comparado a animais do grupo controle (animais sem infecção),

podendo estar relacionado a permeabilidade da barreira sangue-leite, além de escapes de potássio e lactose no sangue.

Porcentagem de Proteína

O gene *SLC26A7* (Figura 4), localizado no cromossomo 15, apresentou relações com a característica avaliada. Segundo (Gaowa et al., (2021) , o gene *Solute Carrier Family 26 Member 7* (*SLC26A7*) atua diretamente nas comunidades microbianas presentes nas câmaras estomacais dos animais, afetando sua saúde e nutrição ocasionando variações na produção e qualidade do leite.

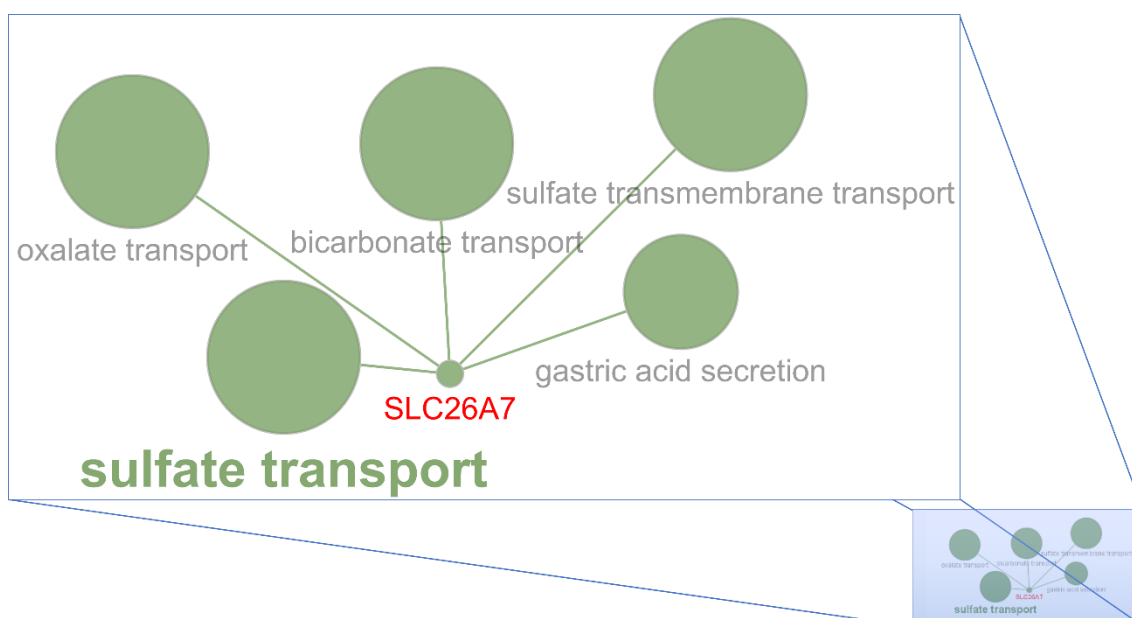


Figura 4. Rede funcional e suas interações para a porcentagem de proteína no leite. Círculos menores com escrita em vermelho são genes candidatos, os processos biológicos enriquecidos estão em negrito de suas respectivas cores e seus subprocessos descritos em letra preta.

Produção de Gordura

Novamente o gene *ITSN2* (Figura 5), descrito anteriormente para PL e PP, foi associado a outra característica avaliada. Esta característica também apresenta boas correlações genéticas entre PL, variando de 0,43 e 0,86 e PP, 0,68 e 0,89 (Garcia-cortés et al., 1995). O gene também foi relacionado ao

processo de regulação positiva da extensão dendrítica e foi localizado no cromossomo 14.

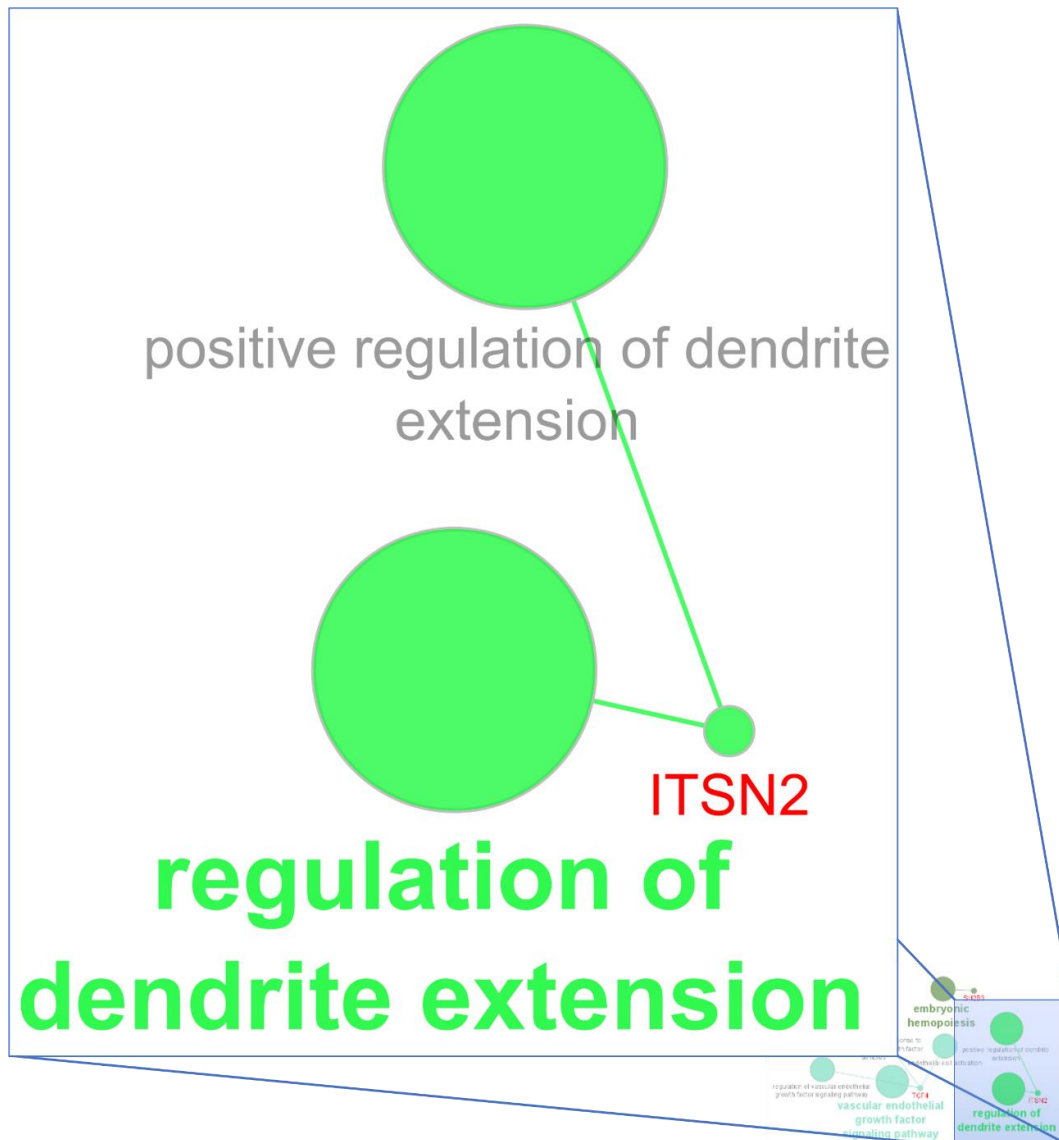


Figura 5. Rede funcional e suas interações para a produção de gordura. Círculos menores com escrita em vermelho são genes candidatos, os processos biológicos enriquecidos estão em negrito de suas respectivas cores e seus subprocessos descritos em letra preta.

Segundo (Kropyvko et al., 2017) o gene ITSN2 foi relacionado a saúde mamária em humanos, atuando negativamente no bem-estar da mama. Segundo Gonçalves et al. (2018) os eventos de mastite clínica ocasionaram a diminuição no rendimento de gordura, variando entre 3 e 22kg durante a lactação dependendo das raças de bovinos leiteiros. Em bubalinos, a mastite ocasionou

(2007) descreveram que em alguns casos o gene apresenta baixo efeito em características de qualidade do leite.

Escore de Células Somáticas

O gene *MDS1 And EVI1 Complex Locus (MECOM)* (Figura 7), localizado no cromossomo 1, foi relacionado com o processo de regulação negativa da cascata JNK e segundo (Guerrero-zotano et al., 2018) ele contribui para a regulação da transcrição do câncer de mama. Este gene também foi associado a expressão de diversos outros tipos de câncer, além de sua atuação como regulador transcricional, desenvolvimento, apoptose e proliferação (Choi et al., 2017).

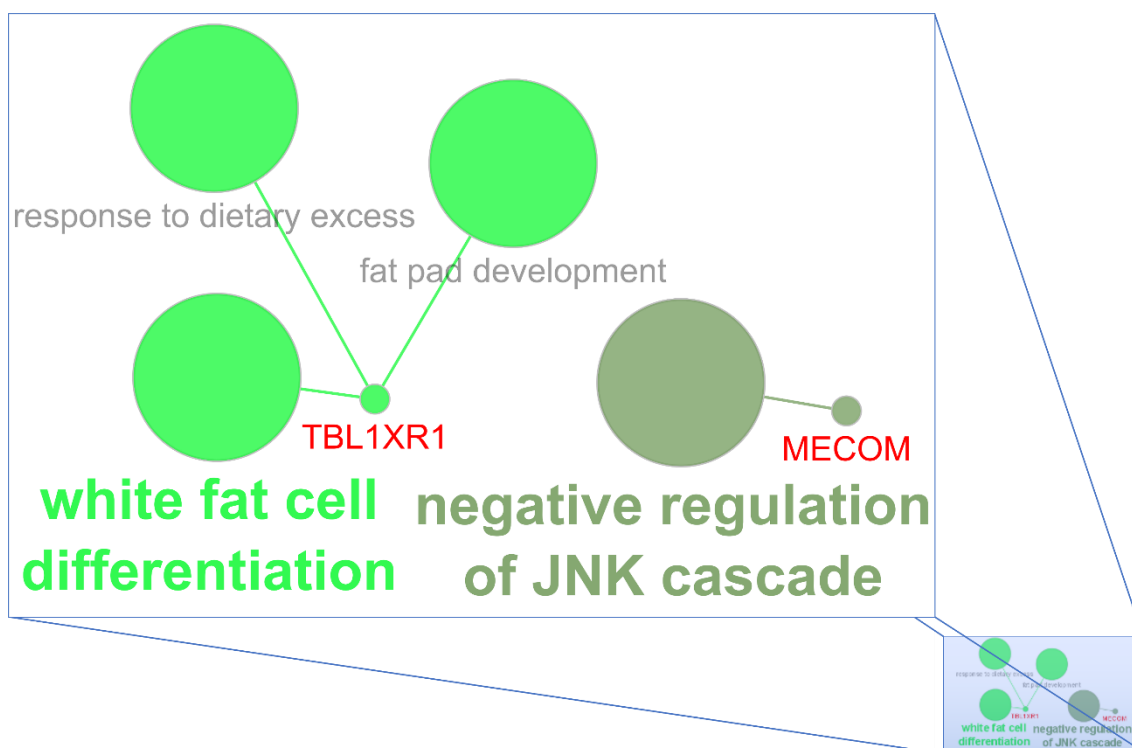


Figura 7. Rede funcional e suas interações para o escore de células somáticas no leite. Círculos menores com escrita em vermelho são genes candidatos, os processos biológicos enriquecidos estão em negrito de suas respectivas cores e seus subprocessos descritos em letra preta.

Por fim, o *TBL1X/Y Related 1 (TBL1XR1)*, localizado no cromossomo 1, foi ligado ao processo biológico enriquecido pelo desenvolvimento da almofada de gordura foi relato por Sakihama; Karube. (2022) como responsável por processos relacionados ao sistema imunológico, atuando principalmente a processos voltados a células T, células responsáveis pela resposta contra patógenos.

3.4. CONCLUSÃO

Este foi o primeiro estudo a utilizar a metodologia de ssGWAS com blocos de haplótipos na avaliação de características produtivas em búfalos Murrah. Esta metodologia foi eficiente na identificação de genes ligados as características avaliadas, podendo ser aplicada em estudos adicionais em búfalos. As redes gênicas permitiram a identificação dos genes *ITSN2* associado a produção de leite, proteína e de gordura, *SLC26A7* para porcentagem de proteína, *DAB1* e *TG* para porcentagem de gordura e *MECOM* e *TBL1XR1* para escore de células somáticas. Nossos resultados podem ser mais explorados para validar os resultados de GWAS para as características leiteiras em búfalos da raça Murrah.

3.5. REFERÊNCIAS

- AGUILAR, I. et al. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. **Journal of Dairy Science**, v. 93, n. 2, p. 743–752, fev. 2010.
- ANTON, I. et al. **Effect of DGAT1, leptin and TG gene polymorphisms on some milk production traits in different dairy cattle breeds in Hungary**. [s.l.: s.n.].
- ARAUJO, A. C. et al. Haplotype-based single-step gwas for yearling temperament in american angus cattle. **Genes**, v. 13, n. 1, 1 jan. 2022.
- ASPILCUETA-BORQUIS, R. R. et al. Multiple-trait random regression models for the estimation of genetic parameters for milk, fat, and protein yield in buffaloes. **Journal of Dairy Science**, v. 96, n. 9, p. 5923–5932, set. 2013.
- BEGUM, F. et al. Regionally Smoothed Meta-Analysis Methods for GWAS Datasets. **Genetic Epidemiology**, v. 40, n. 2, p. 154–160, 1 fev. 2016.
- BINDEA, G. et al. ClueGO: A Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. **Bioinformatics**, v. 25, n. 8, p. 1091–1093, 2009.
- CHOI, E. J. et al. Intratumoral Heterogeneity of Frameshift Mutations in MECOM Gene is Frequent in Colorectal Cancers with High Microsatellite Instability. **Pathology and Oncology Research**, v. 23, n. 1, p. 145–149, 1 jan. 2017.
- CUYABANO, B. C.; SU, G.; LUND, M. S. **Genomic prediction of genetic merit using LD-based haplotypes in the Nordic Holstein population** **BMC Genomics**. [s.l.: s.n.]. Disponível em: <<http://www.biomedcentral.com/1471-2164/15/1171>>.
- FAGIOLO, A.; LAI, O. Mastitis in buffalo. **Italian Journal of Animal Science**, v. 6, n. SUPPL. 2, p. 200–206, 2007.
- FENG, L. et al. The effect of exogenous melatonin on milk somatic cell count in buffalo. **Pakistan Veterinary Journal**, v. 41, n. 1, p. 152–155, 2021.
- GAOWA, N. et al. The effects of artificially dosed adult rumen contents on abomasum transcriptome and associated microbial community structure in calves. **Genes**, v. 12, n. 3, 1 mar. 2021.
- GARCIA-CORTÉS, L. A. et al. **variance component estimation of yield traits between different lactations using an animal model** **Livestock Production Science**. [s.l.] Co, 1995.
- GONÇALVES, J. L. et al. Bovine subclinical mastitis reduces milk yield and economic return. **Livestock Science**, v. 210, p. 25–32, 1 abr. 2018.

GUERRERO-ZOTANO, A. L. et al. ERp Breast cancers resistant to prolonged neoadjuvant letrozole exhibit an e2f4 transcriptional program sensitive to cdk4/6 inhibitors. **Clinical Cancer Research**, v. 24, n. 11, p. 2517–2529, 1 jun. 2018.

GUIMARÃES, J. L. B. et al. Estimate of the economic impact of mastitis: A case study in a Holstein dairy herd under tropical conditions. **Preventive Veterinary Medicine**, v. 142, p. 46–50, 1 jul. 2017.

HILL, W. G.; ROBERTSON, A. **Linkage Disequilibrium in Finite Populations**. [s.l: s.n.].

JOHNSON, G. C. et al. Haplotype tagging for the identification of common disease genes. **Nature Genetics**, v. 29, n. october, p. 233–237, 2001.

KHATIB, H. et al. **Evaluation of association between polymorphism within the thyroglobulin gene and milk production traits in dairy cattle**. [s.l: s.n.]. Disponível em: <<http://ep.espacenet.com>>.

KHIALEEVA, E.; LANE, T. F.; CARPENTER, E. M. Disruption of reelin signaling alters mammary gland morphogenesis. **Development**, v. 138, n. 4, p. 767–776, 15 fev. 2011.

KIM, S. A. et al. A new haplotype block detection method for dense genome sequencing data based on interval graph modeling of clusters of highly correlated SNPs. **Bioinformatics**, v. 34, n. 3, p. 388–397, 1 fev. 2018.

KIM, S. A. et al. Gpart: Human genome partitioning and visualization of high-density SNP data by identifying haplotype blocks. **Bioinformatics**, v. 35, n. 21, p. 4419–4421, 1 nov. 2019.

KORTE, A.; FARLOW, A. **The advantages and limitations of trait analysis with GWAS: A review**. **Plant Methods**, 22 jul. 2013.

KROPYVKO, S. et al. Mammalian verprolin CR16 acts as a modulator of ITSN scaffold proteins association with actin. **Biochemical and Biophysical Research Communications**, v. 484, n. 4, p. 813–819, 18 mar. 2017.

LÁZARO, S. F. et al. Genomic studies of milk-related traits in water buffalo (*Bubalus bubalis*) based on single-step genomic best linear unbiased prediction and random regression models. **Journal of Dairy Science**, v. 104, n. 5, p. 5768–5793, 1 maio 2021.

LI, D.; LIU, Q.; SCHNABLE, P. S. Twas results are complementary to and less affected by linkage disequilibrium than gwas. **Plant Physiology**, v. 186, n. 4, p. 1800–1811, 1 ago. 2021.

LI, Y.; SUNG, W. K.; LIU, J. J. Association mapping via regularized regression analysis of single-nucleotide-polymorphism haplotypes in variable-sized sliding windows. **American Journal of Human Genetics**, v. 80, n. 4, p. 705–715, 2007.

LI, Z.; TZENG, C. M. Integrated analysis of miRNA and mRNA expression profiles to identify miRNA targets. In: **Methods in Molecular Biology**. [s.l.] Humana Press Inc., 2018. v. 1720p. 141–148.

MARÍN-GARZÓN, N. A. et al. Genome-wide association study identified genomic regions and putative candidate genes affecting meat color traits in Nellore cattle. **Meat Science**, v. 171, 1 jan. 2021.

MISZTAL, I. et al. **BLUPF90 AND RELATED PROGRAMS (BGF90)**. [s.l: s.n.]. Disponível em: <<http://www.ozemail.com.au/~milleraj.>>.

MISZTAL, I. et al. **Manual for BLUPF90 family of programs**. [s.l: s.n.].

MRODE, R. A. **Linear Models for the Prediction of Animal Breeding Values**. 3rd Edition ed. [s.l.] Cabi; 3rd UK ed. edição, 2014.

OTTO, P. I. et al. 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.

SAKIHAMA, S.; KARUBE, K. **Genetic Alterations in Adult T-Cell Leukemia/Lymphoma: Novel Discoveries with Clinical and Biological Significance**. **CancersMDPI**, , 1 maio 2022.

SARGOLZAEI, M.; CHESNAIS, J. P.; SCHENKEL, F. S. **A new approach for efficient genotype imputation using information from relatives**. [s.l: s.n.]. Disponível em: <<http://www.biomedcentral.com/1471-2164/15/478>>.

TONHATI, H.; MENDOZA-SÁNCHEZ, G.; SESANA, R. C. **Programa de mejoramiento genético de búfalos * Buffaloes breeding program** *Revista Argentina de Producción Animal*. [s.l: s.n.].

VANRADEN, P. M. Efficient methods to compute genomic predictions. **Journal of Dairy Science**, v. 91, n. 11, p. 4414–4423, 2008.

VISSCHER, P. M. et al. **Five years of GWAS discovery**. **American Journal of Human Genetics**, 13 jan. 2012.

CAPÍTULO 4 – Considerações finais

De forma geral, as metodologias de associação genômica ampla com base em janelas e blocos de haplótipos foram eficazes na identificação de regiões genômicas e genes possíveis candidatos a seleção. Quando avaliamos a metodologia com base em SNP, observa-se maior número de informações de genes, com janelas maiores, quando comparado a metodologia de haplótipos. Isto pode ser explicado pelo número fixo de SNP, pré-definidos (10 marcadores), utilizados na formação das janelas, superior ao número utilizado na análise de blocos, que variou de 2 a 10 marcadores. O tamanho médio das janelas encontrado foi aproximadamente 16 vezes maior que o tamanho médio dos blocos, onde a extensão média estimadas foi de 474.504 e 28.780 pares de base (pb), respectivamente. Esta variação do tamanho dos blocos está de acordo com estudo realizado por Calus et al. (2009), onde descreve que o tamanho dos blocos construídos pode ser influenciado pela densidade do painel de genotipagem e metodologia utilizado. Esta diferença no comprimento das regiões genômicas identificadas também justifica o número de genes localizados e descritos entre as metodologias, em que maior número de genes localizados por região foram encontrados nas janelas.

Com base nos resultados encontrados nos presentes trabalhos, a utilização conjunta dos resultados obtidos por meio das metodologias com base em SNP e haplótipos, passa a ser uma eficaz alternativa de avaliação, maximizando o número de informações avaliadas por meio do aumento da dimensão de cobertura de todo o genoma, buscando associar e validar um maior número de genes candidatos para seleção.

Dentre os genes selecionados para a discussão, apenas uma família de genes pode ser associada as características avaliadas em ambos os trabalhos, a família de genes SLC26. Em mamíferos, esta família é constituída por 11 genes designados SLC26A1 a A11, os quais desempenham diferentes funções (KERE, 2006). Os genes identificados nos presentes estudos estão relacionados ao escore de células somáticas e porcentagem de proteína, necessitando estudos

adicionais para comprovar sua influência na produção e qualidade do leite em búfalas.

Os genes descritos nos trabalhos são genes possíveis candidatos a seleção, sendo necessário estudos adicionais para validar sua importância no desempenho das características avaliadas. Após confirmada, marcadores moleculares localizados nos genes, próximos e em desequilíbrio de ligação com estes ou, caso encontrado, mutações causais, podem ser utilizadas para seleção assistida por marcadores (MAS). Isto permitiria uma seleção de forma mais acurada dos animais com maior valor genético, que apresentam marcadores específicos ligados aos genes de interesse, associados as características de produção e qualidade do leite estudadas.

Outra abordagem buscando o aproveitamento dos resultados de estudos de GWAS em programas de melhoramento animal, seria a implementação das informações geradas nestes estudos em análises de predição genômica, buscando aumento na acurácia de predição. Estudos têm demonstrado que marcadores significativos encontrados em GWAS podem ser inseridos no modelo estatísticos como um efeito fixo ou utilizados na construção de uma matriz G específica, onde são atribuídos pesos para cada SNP (Brøndum et al., 2015; Lopes et al., 2017; Zhang et al., 2010). Desta forma, marcadores com maior contribuição genética para a característica serão favorecidos, contribuindo para uma estimação dos valores genéticos mais acurados quando comparado ao uso da matriz genômica tradicional (matriz G), podendo ser utilizados como uma importante ferramenta no programa de melhoramento genético de Búfalos, auxiliando no desenvolvimento e progresso da raça.

REFERÊNCIAS

- BRØNDUM, R. F. et al. Quantitative trait loci markers derived from whole genome sequence data increases the reliability of genomic prediction. **Journal of Dairy Science**, v. 98, n. 6, p. 4107–4116, 1 jun. 2015.
- CALUS, M. P. et al. Effects of the number of markers per haplotype and clustering of haplotypes on the accuracy of QTL mapping and prediction of genomic breeding values. **Genetics Selection Evolution**, v. 41, n. 11, 2009.

KERE, J. **Overview of the SLC26 family and associated diseases**. [s.l.] John Wiley, 2006.

LOPES, M. S. et al. Using markers with large effect in genetic and genomic predictions. **Journal of Animal Science**, v. 95, n. 1, p. 59, 2017.

ZHANG, Z. et al. Best linear unbiased prediction of genomic breeding values using a trait-specific marker-derived relationship matrix. **PLoS ONE**, v. 5, n. 9, p. 1–8, 2010.

APÊNDICES

Tabela 1A. Informações relacionadas as janelas com maior porcentagem da variância explicada estruturada em ordem decrescente para as análises pós-GWAS para Produção de Leite com todos os genes encontrados em suas respectivas regiões.

Produção de Leite				
Chr	Início	Final	Var_e	Genes
5	22.882.327	23.367.722	0,40	<i>ASTN1</i> , <i>LOC112585067</i> , <i>PAPPA2</i>
12	74.171.537	74.580.524	0,32	<i>NCOA1</i> , <i>ITSN2</i>
21	28.197.211	28.493.766	0,29	<i>PDZRN3</i> , <i>LOC112581277</i> , <i>LOC112581279</i>
2	74.082.787	74.417.499	0,28	<i>LOC102411690</i> , <i>LOC112583050</i> , <i>ATP5MC3</i> , <i>ATF2</i> , <i>CHN1</i> , <i>LOC112583094</i>
13	32.055.911	32.632.965	0,28	
11	87.734.368	88.114.496	0,26	<i>PIAS1</i> , <i>LOC112587918</i> , <i>LOC112577867</i> , <i>LOC112577895</i> , <i>SKOR1</i> , <i>MAP2K5</i> , <i>LOC112587919</i>
9	59.203.043	59.664.067	0,23	<i>LOC102413394</i> , <i>LOC112586893</i> , <i>LOC102406833</i> , <i>LOC112586895</i> , <i>PCDHAC1</i> , <i>LOC112586894</i> , <i>LOC102393295</i> , <i>LOC112586891</i> , <i>LOC112587200</i> , <i>LOC112586897</i> , <i>LOC112586896</i> , <i>LOC102409093</i> , <i>ZMAT2</i> , <i>HARS2</i> , <i>HARS</i> , <i>DND1</i> , <i>WDR55</i> , <i>IK</i> , <i>NDUFA2</i> , <i>TMCO6</i> , <i>CD14</i> , <i>LOC112586900</i> , <i>LOC102398505</i> , <i>SLC35A4</i> , <i>LOC112586898</i> , <i>APBB3</i> , <i>LOC112586899</i> , <i>SRA1</i> , <i>EIF4EBP3</i> , <i>LOC102409452</i> , <i>LOC112587218</i>
21	38.244.251	38.710.893	0,23	<i>CADPS</i>
4	97.720.215	97.943.648	0,21	
2	136.451.314	136.826.115	0,20	

Chr: Cromossomo; Início: Posição inicial da janela em pares de base; Final: Posição final da janela em pares de bases; Var_e : Proporção da variância explicada.

Tabela 2A. Informações relacionadas as janelas com maior porcentagem da variância explicada estruturada em ordem decrescente para as análises pós-GWAS para Produção de Proteína com todos os genes encontrados em suas respectivas regiões.

Produção de Proteína				
Chr	Início	Final	Var_e	Genes
9	59.203.043	59.664.067	0,57	LOC102413394, LOC112586893, LOC102406833, LOC112586895, PCDHAC1, LOC112586894, LOC102393295, LOC112586891, LOC112587200, LOC112586897, LOC112586896, LOC102409093, ZMAT2, HARS2, HARS, DND1, WDR55, IK, NDUFA2, TMCO6, CD14, LOC112586900, LOC102398505, SLC35A4, LOC112586898, APBB3, LOC112586899, SRA1, EIF4EBP3, LOC102409452, LOC112587218
13	51.905.104	52.865.113	0,45	LOC102394645, LOC102396621, LOC102415771
5	22.882.327	23.367.722	0,39	ASTN1, LOC112585067, PAPPA2, TRNAG-CCC
13	32.055.911	32.632.965	0,39	
2	136.451.314	136.826.115	0,35	
2	74.040.263	74.385.057	0,34	LOC102411690, LOC112583050, ATP5MC3, ATF2, CHN1, LOC112583094
2	135.144.109	135.762.880	0,28	
2	124.233.692	124.713.447	0,26	CFAP221, TRNAE-UUC, TRNAC-GCA, TMEM177, PTPN4, LOC112582552, EPB41L5, LOC112582997, LOC112583143
3	130.877.405	131.230.304	0,26	
11	4.242.175	4.970.016	0,24	LOC102394035

Chr: Cromossomo; Início: Posição inicial da janela em pares de base; Final: Posição final da janela em pares de bases; Var_e : Proporção da variância explicada.

Tabela 3A. Informações relacionadas as janelas com maior porcentagem da variância explicada estruturada em ordem decrescente para as análises pós-GWAS para Porcentagem de Proteína com todos os genes encontrados em suas respectivas regiões.

Porcentagem de Proteína				
Chr	Início	Final	Var_e	Genes
14	44.796.559	45.216.106	0,84	SLC24A3, LOC112578807, LOC112578809, LOC102415547, SCP2D1, DTD1
16	48.010.995	48.399.590	0,37	LOC112579518, TRNAH-GUG, SOX6
1	182.181.195	182.495.987	0,30	CPNE4, TRNAS-GCU, LOC102415238
8	54.583.552	55.025.566	0,29	LOC102403299
11	20.812.433	21.360.230	0,28	MAP3K9, LOC102399077, TTC9, LOC112577694, LOC102398227, MED6, LOC102396933, ADAM20, LOC102398545, ADAM21, LOC102393166, LOC112587951, SYNJ2BP
11	7.499.217	7.924.058	0,27	NUBP1, TEK5, EMP2, ATF7IP2, LOC112581805, LOC112581909
24	33.059.598	33.523.710	0,22	
8	1.390.411	1.917.371	0,22	LOC112586670
3	138.689.052	139.153.333	0,22	LOC102410739, LOC112583936, RASEF, FRMD3
4	73.183.227	73.812.097	0,22	LOC102414765, LOC112584566, CAND1

Chr: Cromossomo; Início: Posição inicial da janela em pares de base; Final: Posição final da janela em pares de bases; Var_e : Proporção da variância explicada.

Tabela 4A. Informações relacionadas as janelas com maior porcentagem da variância explicada estruturada em ordem decrescente para as análises pós-GWAS para Produção de Gordura com todos os genes encontrados em suas respectivas regiões.

Produção de Gordura				
Chr	Início	Final	Var_e	Genes
13	51.905.104	52.865.113	0,55	LOC102394645, LOC102396621, LOC102415771
13	32.055.911	32.632.965	0,47	

13	49.134.889	50.040.735	0,42	LOC102402869
5	22.882.327	23.367.722	0,39	ASTN1, LOC112585067, PAPPA2, TRNAG-CCC
9	59.203.043	59.664.067	0,38	LOC102413394, LOC112586893, LOC102406833, LOC112586895, PCDHAC1, LOC112586894, LOC102393295, LOC112586891, LOC112587200, LOC112586897, LOC112586896, LOC102409093, ZMAT2, HARS2, HARS, DND1, WDR55, IK, NDUFA2, TMCO6, CD14, LOC112586900, LOC102398505, SLC35A4, LOC112586898, APBB3, LOC112586899, SRA1, EIF4EBP3, LOC102409452, LOC112587218
2	74.040.263	74.385.057	0,33	LOC102411690, LOC112583050, ATP5MC3, ATF2, CHN1, LOC112583094
3	81.840.631	82.380.929	0,32	LOC112583540
21	28.197.211	28.493.766	0,31	PDZRN3, LOC112581277, LOC112581279
11	1.999.363	2.471.913	0,30	FOXN3, LOC102396186
3	83.902.255	84.250.027	0,30	

Chr: Cromossomo; Início: Posição inicial da janela em pares de base; Final: Posição final da janela em pares de bases; Var_e : Proporção da variância explicada.

Tabela 5A. Informações relacionadas as janelas com maior porcentagem da variância explicada estruturada em ordem decrescente para as análises pós-GWAS para Porcentagem de Gordura com todos os genes encontrados em suas respectivas regiões.

Porcentagem de Gordura				
Chr	Início	Final	Var_e	Genes
2	124.548.843	125.016.355	0,4	PTPN4, EPB41L5, LOC112582997, LOC112583143, TMEM185B, RALB, LOC112578475, INHBB
6	66.245.168	66.946.549	0,4	MIGA1, USP33, ZZZ3, AK5, PIGK

19	67.465.425	67.817.734	0,31	<i>ICE1, ADAMTS16</i>
15	32.167.001	33.032.940	0,29	<i>CDH20, LOC112581460, LOC112581459</i>
22	1.915.805	2.229.915	0,28	<i>LOC102402155, LOC112586188, LOC112586189</i>
7	44.474.034	44.799.967	0,28	<i>PTPN13, MAPK10, LOC102395404</i>
7	16.021.230	16.287.888	0,26	<i>LOC102398867, RAP2B, LOC112578081, LOC112586095</i>
1	158.670.966	159.169.276	0,24	<i>NSMAF, SDCBP, LOC112579182, LOC102415293, LOC112579183, UBXN2B</i>
15	57.525.732	57.892.793	0,23	<i>C1QL2, LOC102397943, STEAP3, LOC102390404, C2H2orf76, DBI, TMEM37, SCTR, CFAP221</i>
2	123.755.686	124.233.692	0,22	

Chr: Cromossomo; Início: Posição inicial da janela em pares de base; Final: Posição final da janela em pares de bases; Var_e : Proporção da variância explicada.

Tabela 6A. Informações relacionadas as janelas com maior porcentagem da variância explicada estruturada em ordem decrescente para as análises pós-GWAS para Escore de Células Somáticas com todos os genes encontrados em suas respectivas regiões.

Escore de Células Somáticas				
Chr	Início	Final	Var_e	Genes
3	57.055.515	57.723.147	0,55	<i>LOC112583815, LOC102413458, LOC112583816, LOC102412899, LOC112583481, LOC112584098, LOC102397853, LOC112583817, LOC112583818, LOC112583819, LOC112584096, LOC102412952, LOC102415089, LOC112584097, TMEM100, LOC102412706</i>
19	64.950.232	65.319.092	0,30	<i>MTRR, FASTKD3, C19H5orf49, ADCY2</i>
12	82.278.725	82.550.881	0,28	<i>LOC112578115, DDX1, NBAS, LOC112578297</i>
6	45.725.219	46.165.125	0,28	<i>DPYD</i>
10	4.876.936	5.216.651	0,25	<i>QKI</i>

23	15.901.734	16.583.571	0,25	LOC102407362, LOC102408577, LOC102396619, LOC112581525, LOC102401347, LOC102392701, LOC102391944, LOC112581647, LOC102402137, PDLIM1, LOC102390247
10	22.683.031	22.962.597	0,25	STX11, SF3B5, PLAGL1, LOC102414660, ZC2HC1B, LTV1, PHACTR2
12	67.224.166	67.517.096	0,25	ANTXR1, GFPT1, NFU1, AAK1
4	63.963.147	64.628.584	0,24	LOC102396058, LOC112584798, LRP1, NXPH4, SHMT2, NDUFA4L2, LOC112584553, STAC3, R3HDM2, LOC112584554, INHBC, INHBE, GLI1, ARHGAP9, MARS, DDIT3, MBD6, DCTN2, KIF5A, PIP4K2C, DTX3, ARHGEF25, SLC26A10, B4GALNT1, LOC102403500, LOC112584402, OS9, LOC112584555, AGAP2, TSPAN31, CDK4, MARCH9, LOC102414554, METTL1, EEF1AKMT3, TSFM, AVIL, CTDSP2, LOC112584556
14	44.827.731	45.248.556	0,23	SLC24A3, LOC112578807, LOC112578809, LOC102415547, SCP2D1, DTD1

Chr: Cromossomo; Início: Posição inicial da janela em pares de base; Final: Posição final da janela em pares de bases; Var_e : Proporção da variância explicada.

Tabela 7A. Informações relacionadas aos blocos com maior porcentagem da variância explicada estruturada em ordem decrescente para as análises pós-GWAS para Produção de Leite com todos os genes encontrados em suas respectivas regiões.

Produção de Leite				
Chr	Início	Final	Var_e	Genes
17	17.883.840	17.911.336	0,07	SH2B3
10	14.870.334	14.905.564	0,07	SYNE1
17	71.927.676	71.951.418	0,06	
12	74.453.460	74.482.613	0,05	ITSN2
21	28.220.032	28.240.175	0,05	PDZRN3
5	23.339.586	23.367.722	0,04	PAPPA2
21	28.325.698	28.351.030	0,04	PDZRN3
9	87.901.882	87.911.410	0,04	LOC112586951

14	65.222.836	65.261.082	0,04	<i>PARD3</i>
13	32.221.326	32.251.064	0,04	

Chr: Cromossomo; Início: Posição inicial da janela em pares de base; Final: Posição final da janela em pares de bases; Var_e : Proporção da variância explicada.

Tabela 8A. Informações relacionadas aos blocos com maior porcentagem da variância explicada estruturada em ordem decrescente para as análises pós-GWAS para Produção de Proteína com todos os genes encontrados em suas respectivas regiões.

Produção de Proteína				
Chr	Início	Final	Var_e	Genes
19	11.077.957	11.104.688	0,06	<i>SH2B3</i> <i>LOC102415771</i> <i>TCF4</i> <i>ITSN2</i> <i>SYNE1</i>
17	17.883.840	17.911.336	0,06	
13	52.712.877	52.747.153	0,05	
22	7.255.648	7.283.121	0,05	
12	74.453.460	74.482.613	0,05	<i>PAPPA2</i> <i>CDH18</i>
10	14.870.334	14.905.564	0,05	
13	32.221.326	32.251.064	0,04	
5	23.339.586	23.367.722	0,04	
20	2.947.635	2.968.271	0,04	
19	53.640.371	53.667.741	0,04	

Chr: Cromossomo; Início: Posição inicial da janela em pares de base; Final: Posição final da janela em pares de bases; Var_e : Proporção da variância explicada.

Tabela 9A. Informações relacionadas aos blocos com maior porcentagem da variância explicada estruturada em ordem decrescente para as análises pós-GWAS para Porcentagem de Proteína com todos os genes encontrados em suas respectivas regiões.

Porcentagem de Proteína				
Chr	Início	Final	Var_e	Genes
20	10.243.689	10.269.523	0,05	<i>LOC112580784</i> , <i>LOC112580783</i> , <i>SERPINA4</i> <i>PDZRN3</i> <i>SLC26A7</i> <i>MCC</i> <i>LOC112581805</i>
21	28.220.032	28.240.175	0,05	
15	8.980.779	8.994.091	0,05	
11	101.755.349	101.789.795	0,04	
24	33.378.098	33.401.488	0,04	<i>STPG2</i> <i>KCTD1</i>
12	20.349.856	20.378.216	0,04	
7	101.951.592	101.977.604	0,04	
4	27.490.121	27.526.234	0,04	
22	31.824.785	31.859.201	0,04	
12	27.013.203	27.041.472	0,04	

Chr: Cromossomo; Início: Posição inicial da janela em pares de base; Final: Posição final da janela em pares de bases; Var_e : Proporção da variância explicada.

Tabela 10A. Informações relacionadas aos blocos com maior porcentagem da variância explicada estruturada em ordem decrescente para as análises pós-GWAS para Produção de Gordura com todos os genes encontrados em suas respectivas regiões.

Produção de Gordura				
Chr	Início	Final	Var_e	Genes
13	52.712.877	52.747.153	0,07	<i>LOC102415771</i>
13	32.221.326	32.251.064	0,06	
3	82.001.158	82.031.525	0,06	
21	28.325.698	28.351.030	0,05	<i>PDZRN3</i>
5	23.339.586	23.367.722	0,05	<i>PAPPA2</i>
21	28.220.032	28.240.175	0,05	<i>PDZRN3</i>
11	2.140.106	2.173.979	0,05	<i>FOXN3</i>
12	74.453.460	74.482.613	0,05	<i>ITSN2</i>
17	17.883.840	17.911.336	0,05	<i>SH2B3</i>
22	7.255.648	7.283.121	0,05	<i>TCF4</i>

Chr: Cromossomo; Início: Posição inicial da janela em pares de base; Final: Posição final da janela em pares de bases; Var_e : Proporção da variância explicada.

Tabela 11A. Informações relacionadas aos blocos com maior porcentagem da variância explicada estruturada em ordem decrescente para as análises pós-GWAS para Porcentagem de Gordura com todos os genes encontrados em suas respectivas regiões.

Porcentagem de Gordura				
Chr	Início	Final	Var_e	Genes
4	93.949.244	93.983.912	0,07	<i>CALCOCO1</i>
6	66.482.572	66.503.767	0,07	<i>AK5</i>
4	114.708.327	114.733.026	0,06	
15	74.043.693	74.061.982	0,06	<i>TG</i>
20	7.943.940	7.973.300	0,05	<i>LOC112580848</i>
5	47.370.665	47.399.072	0,05	<i>SDCCAG8</i>
5	11.277.275	11.334.109	0,05	<i>LOC102390734</i>
16	62.932.274	62.953.872	0,04	<i>SIK2</i>
6	87.417.070	87.439.439	0,04	<i>DAB1</i>
5	105.145.310	105.165.625	0,04	<i>PUS3, DDX25</i>

Chr: Cromossomo; Início: Posição inicial da janela em pares de base; Final: Posição final da janela em pares de bases; Var_e : Proporção da variância explicada.

Tabela 12A. Informações relacionadas aos blocos com maior porcentagem da variância explicada estruturada em ordem decrescente para as análises pós-GWAS para Escore de Células Somáticas com todos os genes encontrados em suas respectivas regiões.

Escore de Células Somáticas				
Chr	Início	Final	Var_e	Genes
12	82.278.725	82.284.980	0,06	<i>LOC112578115</i>
4	4.791.027	4.821.686	0,06	<i>PRR5</i>
1	142.558.386	142.595.355	0,05	<i>MECOM, LOC112585169</i>
1	134.669.557	134.744.473	0,05	<i>TBL1XR1</i>
10	45.887.118	45.910.218	0,05	
3	66.131.195	66.174.010	0,04	
7	17.719.204	17.743.720	0,04	
14	30.744.325	30.779.491	0,04	
24	17.788.040	17.817.079	0,04	
3	57.099.594	57.135.931	0,04	<i>LOC102413458</i>

Chr: Cromossomo; Início: Posição inicial da janela em pares de base; Final: Posição final da janela em pares de bases; Var_e : Proporção da variância explicada.