

**UNIVERSIDADE ESTADUAL PAULISTA – UNESP
CÂMPUS DE JABOTICABAL**

**ANÁLISE DE ASSOCIAÇÃO GENÔMICA PARA
CARACTERÍSTICAS DE PRODUÇÃO EM BÚFALOS DE
LEITE**

Valentina Roldan Montes
Zootecnista

**UNIVERSIDADE ESTADUAL PAULISTA - UNESP
CÂMPUS DE JABOTICABAL**

**ANÁLISE DE ASSOCIAÇÃO GENÔMICA PARA
CARACTERÍSTICAS DE PRODUÇÃO EM BÚFALOS DE
LEITE**

Valentina Roldan Montes

Orientador: **Prof. Dr. Humberto Tonhati**

Coorientadora: **Dra. Sirlene Fernandes**

Lázaro

Coorientador: **Dr. Diercles Francisco Cardoso**

Tese de Doutorado apresentada à Faculdade de Ciências Agrárias e Veterinárias – Unesp, Câmpus de Jaboticabal, como parte das exigências para a obtenção do título de Doutor em Genética e Melhoramento Animal.

M779a Montes, Valentina Roldan
Análise de associação genômica para características de produção em búfalos de leite / Valentina Roldan Montes. -- Jaboticabal, 2021
55 f. : tabs.

Tese (doutorado) - Universidade Estadual Paulista (Unesp), Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal
Orientador: Humberto Tonhati
Coorientadora: Sirlene Fernandes Lazaro

1. SNPs. 2. Gwas. 3. Leite. 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



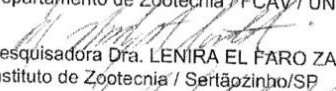
CERTIFICADO DE APROVAÇÃO

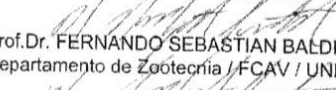
TÍTULO DA TESE: ANÁLISE DE ASSOCIAÇÃO GENÔMICA PARA CARACTERÍSTICAS DE PRODUÇÃO EM BÚFALOS DE LEITE

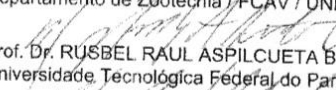
AUTORA: VALENTINA ROLDAN MONTES
ORIENTADOR: HUMBERTO TONHATI
COORIENTADORA: SIRLENE FERNANDES LÁZARO
COORIENTADOR: DIERCLES FRANCISCO CARDOSO

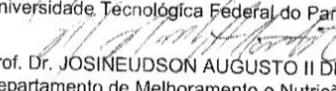
Aprovada como parte das exigências para obtenção do Título de Doutora em GENÉTICA E MELHORAMENTO ANIMAL, pela Comissão Examinadora:


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


Pesquisadora Dra. LENIRA EL FARO ZADRA (Participação Virtual)
Instituto de Zootecnia / Sertãozinho/SP


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


Prof. Dr. RUSBEL RAUL ASPILCUETA BORQUIS (Participação Virtual)
Universidade Tecnológica Federal do Paraná - Campus Dois Vizinhos / Dois Vizinhos/PR


Prof. Dr. JOSINEUDSON AUGUSTO II DE VASCONCELLOS SILVA (Participação Virtual)
Departamento de Melhoramento e Nutrição Animal / FMVZ/UNESP - Botucatu/SP

Jaboticabal, 14 de dezembro de 2020

DADOS CURRICULARES DO AUTOR

VALENTINA ROLDAN MONTES - nascida em 23 de dezembro de 1988, na cidade de Cali – Colômbia, filha de Júlio Cesar Roldan Ch. e Gloria Amparo Montes T. Iniciou em fevereiro de 2008 o curso de Tecnologia em produção animal na Corporação Universitária Santa Rosa de Cabal – UNISARC obtendo o título em junho de 2013, depois início a graduação em Zootecnia na mesma universidade obtendo o título de Zootecnista em abril de 2015. Em agosto de 2014, ingressou no Programa de Pós-graduação em Genética e Melhoramento Animal da Faculdade de Ciências Agrárias e Veterinárias da Universidade Estadual Paulista “Júlio de Mesquita Filho”, campus de Jaboticabal como bolsista da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), fundação do Ministério da Educação (MEC), sob orientação do Prof. Dr. Humberto Tonhati, obtendo o título de Mestre em 18 de Outubro de 2016. Em março de 2017 iniciou o doutorado na mesma instituição e no mesmo programa de pós-graduação, como bolsista da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), com orientação do Prof. Dr. Humberto Tonhati e coorientação da Dra. Sirlene Fernandes Lázaro e do Dr. Diercles Francisco Cardoso.

“La vida no es la que uno vivió, sino la que uno recuerda, y cómo la recuerda para contarla.”

Gabriel García Márquez

A meus pais Júlio César e Gloria Amparo; a meus avós Jorge Enrique e Aura Nelly vocês são minha base e meu incentivo, a meu namorado Oscar pelo amor, a compreensão, por sobre todas as coisas por ser meu grande apoio sempre. Vocês que não mediram esforços para que eu chegasse até aqui, Obrigada por sempre acreditarem e confiarem em mim.

OFEREÇO

AGRADECIMENTOS

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

Ao Prof. Humberto Tonhati, pela orientação, paciência, apoio e amizade em todo momento. Obrigada pela oportunidade.

Ao Diercles, Sirlene e Alexandra pelos ensinamentos, sugestões, pela amizade e pela enorme disposição em sempre me ajudar.

Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), pela bolsa de estudos concedida.

Aos membros de minha banca de qualificação, Prof. Roberto, Prof. Lenira, Prof. Henrique e Dra. Thaise, muito obrigada pelas dicas e correções. Foram muito uteis.

A meu namorado Oscar pelo grande apoio sempre, pela compreensão, pelo amor, e por tudo que você me oferece.

A meus pais Júlio César e Gloria Amparo; a meus avós Jorge Enrique e Aura Nelly; Obrigada por tudo, pela compreensão, carinho e amor. Essa conquista também é de vocês!

Muito obrigada!

SUMÁRIO

RESUMO –	ii
ABSTRACT-	iii
TABELA DE FIGURAS	iv
1. INTRODUÇÃO.....	1
2. REVISÃO DE LITERATURA	4
2.1 Bubalinocultura.....	4
2.2 Seleção genômica e Estudos de associação genômica ampla (GWAS).....	5
3. OBJETIVOS	9
4. MATERIAL E MÉTODOS	10
4.1 Dados fenotípicos	10
4.2 Genótipos e controle de qualidade.....	11
4.3 Predição de valores genéticos (EBVs).....	11
4.4 Pseudo-fenótipos (dEBV).....	12
4.5 Análises de associação genômica ampla (GWAS).....	13
4.5.1 Bayes C	13
4.5.2 Single SNP regression	14
4.5.3 Regiões de QTL e genes candidatos.....	15
5. RESULTADOS E DISCUSSÃO	16
5.1 Estimativas de componentes de variância	16
5.2 Genes candidatos em ambas abordagens.....	17
5.2.1 Gordura.....	26
5.2.2 Proteína e Mozzarella.....	29
5.3 BayesC	31
5.4 Single SNP regression	34
6. CONCLUSÃO	36
7. REFERÊNCIAS	37

ANÁLISE DE ASSOCIAÇÃO GENÔMICA PARA CARACTERÍSTICAS DE PRODUÇÃO EM BÚFALOS DE LEITE

RESUMO – A busca por animais com maior potencial de produção e maior eficiência reprodutiva tem contribuído para a implementação de diferentes tipos de programas de seleção. Com a genômica, torna-se possível a seleção dos animais sem a necessidade de mensuração de seus próprios fenótipos e/ou de seus parentes, e tem-se, também, a possibilidade de realizar a busca de genes ou regiões cromossômicas envolvidas com a expressão das características, por meio do estudo de associação genômica ampla (GWAS). A GWAS tem como objetivo associar regiões do genoma com fenótipos de interesse, identificar possíveis regiões de maior efeito sobre a característica e, posteriormente, investigar as funções biológicas dos genes que foram encontrados nestas regiões, visando assim, aumentar a compreensão da influência genética sobre a expressão fenotípica. O objetivo desse trabalho foi identificar regiões genômicas associadas com as características de produção de leite (PL), produção de gordura (PG), produção de proteína (PP), e produção de mozzarella (PMZ), mediante análise de duas diferentes abordagens: regressão single-SNP e Bayes C. Em adição diferenciar os genes encontrados nessas regiões que podem auxiliar na interpretação da expressão e fisiologia da característica. Foram utilizados dados fenotípicos e genotípicos provenientes da primeira lactação de 4068, 372, 488, 482 búfalas da raça Murrah pertencentes ao programa de controle leiteiro de bubalinos mantidos pelo Departamento de Zootecnia da UNESP/Jaboticabal, SP. No estudo de associação genômica ampla, foram identificadas algumas regiões cromossômicas polimórficas, sendo identificados 53 marcadores SNPs em comuns em ambas abordagens (Bayes C, regressão single-SNP), 30 SNPs para gordura, 18 para proteína e 5 para mozzarella. 9, 5 e 3 SNPs significativos para produção de gordura, proteína e mozzarella respectivamente pela abordagem Bayes C. 4 e 1 SNPs significativo para produção de gordura e proteína pela abordagem regressão single-SNP. Assim, os resultados têm permitido identificar diferentes regiões que continham genes importantes em diferentes processos biológicos e funções moleculares associadas pelo menos com um QTL, principalmente nos BBU6, BBU9, BBU3, BBU5 encontrassem os genes *CRTC2*, *ILF2*, *RPS27*, *RHOBTB3*, *S100A1*, *S100A13*, *TRNAG-CCC*, com efeito sob as diferentes características de qualidade do leite em búfalas avaliadas nessa pesquisa.

Palavras-chave: GWAS, LEITE, SNPs

GENOMIC ASSOCIATION ANALYSIS FOR PRODUCTION CHARACTERISTICS IN MILK BUFFALO

ABSTRACT- The search for animals with greater production potential and greater reproductive efficiency has contributed to the implementation of different types of selection programs. With genomics, it is possible to select animals without the need to measure their own phenotypes and / or their relatives, and there is also the possibility to search for genes or chromosomal regions involved with the expression characteristics, through the study of broad genomic association (GWAS). The objective of GWAS is to associate regions of the genome with phenotypes of interest, to identify possible regions of greatest effect on the trait and, subsequently, to investigate the biological functions of the genes that were found in these regions, thus seeking to increase the understanding of the genetic influence on phenotypic expression. The objective of this work was to identify genomic regions associated with the characteristics of milk production (PL), fat production (PG), protein production (PP), and mozzarella production (PMZ), by analyzing two different approaches: regression single-SNP and Bayes C, In addition to differentiate the genes found in these regions that can assist in the interpretation of the expression and physiology of the characteristic. Phenotypic and genotypic data from the first lactation of 4068, 372, 488, 482 Murrah buffaloes belonging to the buffalo milk control program maintained by the Department of Zootechnics at UNESP / Jaboticabal, SP, were used. In the study of broad genomic association, some polymorphic chromosomal regions were identified, with 53 SNPs markers in common in both approaches (Bayes C, regression single-SNP), 30 SNPs for fat, 18 for protein and 5 for mozzarella. 9, 5 and 3 significant SNPs for fat, protein and mozzarella production respectively by the Bayes C approach. 4 and 1 significant SNPs for fat and protein production by the regression single-SNP approach. Thus, the results have allowed to identify different regions that contained important genes in different biological processes and molecular functions associated with at least one QTL, mainly in BBU6, BBU9, BBU3, BBU5 to find the *CRTC2*, *ILF2*, *RPS27*, *RHOBTB3*, *S100A1*, *S100A13*, *TRNAG-CCC*, with effect on the different quality characteristics of milk in buffaloes evaluated in this research.

Keywords: GWAS, MILK, SNPs

TABELA DE FIGURAS

<i>Figura 1. Manhattan plots das associações entre cada SNP e produção de leite (A), produção de gordura (B), proteína (C) e produção de mozzarella (D) provenientes das abordagens Bayes C.</i>	<i>18</i>
<i>Figura 2. Manhattan plots das associações entre cada SNP e produção de leite (A), produção de gordura (B), proteína (C) e produção de mozzarella (D) proveniente da abordagem de single SNP regression.</i>	<i>19</i>
<i>Figura 3. QQ-plot para P-values (-Log10) estimados por meio da análise de single SNP regression.</i>	<i>20</i>
<i>Figura 4. Número de regiões significativas comuns entre as abordagens BayesC e single SNP regression para produção de gordura, proteína e mozzarella.</i>	<i>21</i>
<i>Figura 5. Desequilíbrio de ligação médio em Búfalos da raça Murrah de acordo com intervalos de distância (Kb).</i>	<i>22</i>
<i>Figura 6. Rede gênica de processo biológico (círculo), processos imunológicos (hexágono) e funções moleculares (octógono) para os potenciais genes candidatos identificados para produção de gordura (105 genes foram incluídos nesta análise). Os nomes em vermelho referem se a os genes observados.</i>	<i>29</i>
<i>Figura 7. Rede gênica de processo biológico (círculo), processos imunológicos (hexágono) e funções moleculares (octógono) para os potenciais genes candidatos identificados para produção de proteína (12 genes foram incluídos nesta análise). Os nomes em vermelho referem se a os genes observados..</i>	<i>31</i>
<i>Figura 8. Rede gênica de processo biológico (círculo), processos imunológicos (hexágono) e funções moleculares (octógono) para os potenciais genes candidatos identificados para produção de mozzarella (3 genes foram incluídos nesta análise). Os nomes em vermelho referem se a os genes observados.</i>	<i>31</i>

1. INTRODUÇÃO

A bubalinocultura tem sido identificada como uma atividade agropecuária de importância econômica que, além de fornecer carne e leite, contribui com força de tração para atividades no campo (Warriach et al., 2015). Recentemente a população mundial de búfalos foi estimada em 206 milhões de animais, dos quais aproximadamente 1,3 milhões estão distribuídos no território brasileiro (FAOSTAT, 2018). A produção leiteira em búfalos tornou-se uma atividade de grande importância econômica pelo rendimento na obtenção de derivados lácteos, principalmente do queijo *mozzarella*. O leite de búfala apresenta maiores teores de gordura, proteína e sólidos totais quando comparado ao leite bovino (Tonhati et al., 2008).

Dada a importância dessas características, compreender sua arquitetura genética poderá ser importante para definir melhores estratégias de melhoramento genético. Neste contexto, desde o desenvolvimento de chips de alta densidade de marcadores (SNP), diferentes abordagens de GWAS foram propostas (Liu et al, 2018; El-Halawany et al., 2017; De Camargo et al., 2015). Tais metodologias foram importantes para avanços na identificação de importantes genes em diversas espécies.

Diversos fatores, tais como tamanho da população, densidade dos marcadores (SNP), estrutura genética da população, desequilíbrio de ligação e a escolha de modelos estatísticos podem afetar os resultados de análises de GWAS (Zhang et al. 2012; Lu et al. 2013a). Diversas estratégias têm sido sugeridas para melhorar este tipo de problemas, entre elas aumentar o tamanho populacional a densidade de marcadores (SNPs) (Zhang et al. 2012; Sharma et al. 2015), além da utilização de modelos estatísticos com diferentes pressuposições (Richardson et al., 2016; Melo et al. 2019);

Um dos principais fatores associados à identificação falsas e erros nas associações do GWAS é o método estatístico utilizado para a análise. Portanto, é necessário desenhar o estudo de forma detalhada, dando atenção especial ao controle de qualidade dos dados, que é uma das etapas mais importantes para minimizar erros nos estudos de GWAS (Sharma et al. 2015). Inicialmente na área de

melhoramento animal foram desenvolvidos os modelos mistos, sendo um método eficaz para reduzir a apresentação de falsas associações, e têm sido aplicados com sucesso em GWAS (Wang et al. 2012; Wang et al. 2014), podendo reduzir significativamente o número de associações falsas relatadas como significativas (Sharma et al. 2015). Com isso, o estudo e a comparação entre diferentes metodologias devem ser realizados com o intuito de obter os métodos adequados para as diferentes características de interesse.

Algumas metodologias de predição genômica, sejam elas de caráter frequentista ou bayesiana, surgiram com o propósito de estimar os efeitos dos marcadores. As metodologias bayesianas aplicadas nos estudos genômicos constituem uma abordagem interessante comparado a métodos clássicos, uma vez que tais metodologias assumem que os marcadores podem explicar diferentes proporções da variação genética da característica analisada. Os efeitos preditos dos marcadores são então o efeito de um determinado locus ajustado para todos os outros loci incluídos no modelo (Garrick et al., 2013). A metodologia Bayesiana combina informações *a priori* dos parâmetros a serem preditos com as informações do banco de dados disponível, gerando uma distribuição *a posteriori* para cada parâmetro do modelo, a qual é usada nas inferências (Sorensen et al., 2007). Tais metodologias têm sido com sucesso implementado em análises de GWAS em diferentes espécies (Garrick et al., 2013; Zhang 2015; Dong et al., 2019).

Em búfalos, estudos de GWAS têm sido propostos para características associadas à produção de leite (Barros et al 2018; El Halawany et al. 2017). Barros et al. (2018) identificaram associação significativa para características de produção de leite, e porcentagens de gordura e proteína, em que os genes *ADAMTS12*, *ADAMTS8*, *ADAMTS15*, *ABCC4*, *BDNF* foram considerados importantes genes candidatos. Similarmente, El-Halawany et al. (2017) identificaram importantes regiões genômicas associadas com a produção de leite nos cromossomos BTA1, 5, 6 e 27, tal como os genes *LARGE1*, *ERC1*, *ADGRL3*, *PARM1* em búfalos egípcios. É importante mencionar que as metodologias usadas nesses estudos foram baseadas em abordagens do modelo de repetibilidade e modelos de regressão linear respectivamente. Portanto, realizar análises de GWAS para as características de

produção de leite de búfalos sob abordagens bayesianas será importante para avanços no conhecimento da arquitetura genética dessas características.

Além disso, a comunidade científica tem se interessado no estudo do genoma dos búfalos, a fim de melhorar o desempenho reprodutivo e o potencial de produção do búfalo, gerando informações genômicas como o mapeamento completo mediante genotipagem e sequenciamento do genoma bubalino (Michelizzi et al., 2010), e promovendo o interesse na descoberta de regiões e SNPs como ferramentas para o melhoramento genético para aumentar a produção dos animais. Assim, o objetivou-se com este estudo identificar regiões genômicas para características associadas a produção de leite em búfalos da raça Murrah. Adicionalmente, diferentes abordagens para o GWAS (single SNP regression e Bayes C) foram utilizadas.

2. REVISÃO DE LITERATURA

2.1 Bubalinocultura

A bubalinocultura tem sido identificada como uma atividade de importância econômica agropecuária que, além de fornecer carne e leite, contribui com força de tração para atividades no campo (Warriach et al., 2015). É compreendida principalmente pelas raças Murrah, Mediterrânea, Jafarabadi e seus cruzamentos. Atualmente, a população mundial de búfalos é estimada em aproximadamente 206 milhões de cabeças (FAOSTAT, 2018). A maior proporção do rebanho bubalino concentra-se principalmente na Ásia, sendo a Índia o primeiro país em número de cabeças com aproximadamente, 105 milhões (FAOSTAT, 2018).

A produção leiteira em búfalos tornou-se uma atividade de grande importância econômica pelo rendimento na obtenção de derivados lácteos, principalmente do queijo *mozzarella*. O leite de búfala apresenta maiores teores de gordura, proteína e sólidos totais quando comparado ao leite bovino (Tonhati et al., 2008). No Brasil, a região Sudeste apresenta o maior mercado de derivados de leite de búfalo, principalmente do queijo *mozzarella* devido à proximidade com os maiores centros urbanos do país (Tonhati et al., 2012), sendo o estado de São Paulo considerado o maior produtor de leite de búfala, com 68.307 cabeças (IBGE, 2017).

Os bubalinos apresentam superioridade na produção de leite em relação aos bovinos. Apresentando características que facilitam a sua identificação do ponto de vista físico-químico e sensorial, o leite bubalino difere do leite bovino principalmente pelo seu sabor levemente adocicado e coloração sempre muito branca devido à baixa quantidade de caroteno em sua gordura (Brito; Dias, 1998). Além disso, teores de proteína, gordura e lactose superam, consideravelmente, os teores do leite bovino (Tonhati et al, 2000; Duarte, 2001; Malavolta, 2006). Todos os componentes anteriormente mencionados, mais as vitaminas e minerais constituem os sólidos totais. O maior teor de sólidos totais confere ao leite de búfala maior aproveitamento final em comparação ao leite de vaca, já que em função da menor quantidade de água, haverá maior rendimento no laticínio.

De maneira geral, as vantagens do leite bubalino em relação ao leite bovino podem tornar mais atrativa à exploração da atividade leiteira em búfalos,

principalmente se observada à tendência mundial em pagar o leite pela quantidade e qualidade de seus constituintes (Madalena, 1999; Bueno et al., 2004).

2.2 Seleção genômica e Estudos de associação genômica ampla (GWAS)

O grande atrativo da genética molecular em benefício do melhoramento genético aplicado é a utilização direta das informações de DNA na seleção, de forma a permitir alta eficiência seletiva, grande rapidez na obtenção de ganhos genéticos com a seleção e baixo custo, em comparação com a tradicional seleção baseada em dados fenotípicos (Resende et al., 2008).

A Seleção Genômica (GS) tem sido utilizada em espécies de plantas e animais de todo o mundo (Hill, 2014), sendo bem-sucedida em bovinos leiteiros por possibilitar a redução dos intervalos de geração como resultado da alta acurácia da estimativa do valor genético (EBV) em animais jovens (Schaeffer, 2006). Os modelos de GS usam os marcadores que encontram-se em desequilíbrio de ligação (LD) com os QTL (*quantitative trait loci*), aproveitando os efeitos desses QTLs sem seleção prévia dos marcadores para sua associação com os fenótipos (Dekkers, 2007), além da relação entre indivíduos (Habier et al., 2007).

Os primeiros métodos estatísticos foram propostos por Meuwissen et al. (2001) em um estudo de simulação, abordando métodos dos quadrados mínimos, GBLUP, BayesA e BayesB. Dentre estes, apenas o método dos quadrados mínimos apresentou limitações, como a baixa habilidade preditiva e a impossibilidade de estimar todos os efeitos dos marcadores simultaneamente. Os demais métodos foram similares entre si, apresentando maior acurácia, sendo assim, o GBLUP, BayesA e BayesB tornaram-se amplamente utilizados.

Atualmente os métodos BayesC e BayesC π têm se destacado em função de sua particularidade de propiciar informação sobre a arquitetura genética do caráter quantitativo (Habier et al. 2011). O modelo BayesC, proposto por (Kizilkaya et al., 2010) assume a *priori* que os efeitos dos marcadores possuem distribuição com probabilidade π e sem efeito com probabilidade $1 - \pi$. Os autores observaram que este método apresentou vantagens em fornecer informações sobre a arquitetura

genética da característica e identificou as posições dos QTLs por modelagem da frequência de SNPs não nulos.

No entanto no modelo BayesC π também apresentado por Habier et al. (2011), a probabilidade π do SNP é assumida como parâmetro desconhecido, que pode ser estimada a partir dos dados. Nesse método, uma variância comum é especificada para todos os loci. Em contraste ao GBLUP, em que se assume que a variância é conhecida, no BayesC π essa variância pode ser estimada a partir dos dados.

As metodologias bayesianas aplicadas nos estudos genômicos apresentam vantagens em relação aos métodos clássicos, assumindo que os marcadores podem explicar diferentes proporções da variação genética da característica analisada. Desde então diferentes metodologias tem sido empregadas com sucesso na avaliação genética, inclusive em búfalos, por exemplo, Camargo et al. (2015) predisseram valores genéticos para características de produção de leite, produção de gordura e produção de proteína usando um modelo lineal misto, obtendo uma acurácia de 0.64, 0.61 e 0.61 respectivamente.

No contexto dos estudos de associação genômica ampla (GWAS), os quais são baseados no efeito dos SNPs que estão em desequilíbrio de ligação com mutações causais ou genes que controlam a expressão de uma característica de interesse, o objetivo é associar regiões do genoma com fenótipos de interesse a fim de identificar possíveis regiões de maior efeito sobre a característica e posterior investigação de funções biológicas, visando aumentar a compreensão da influência genética sobre a expressão fenotípica (Yang et al., 2013).

Os primeiros GWAS foram usados para análise de doenças humanas, permitindo grandes avanços no conhecimento do funcionamento de doenças complexas como o câncer (Hirschhorn and Daly 2005; Klein et al., 2005; Visscher et al., 2012; McCarthy et al., 2008; Lee et al., 2008). Logo após, esses tipos de análises foram utilizados na genética animal (Chan et al., 2009) como o sequenciamento do genoma das espécies domésticas, que tem permitido a identificação de muitos SNPs nos genomas dos animais (Zimin et al., 2009).

Atualmente, existem métodos eficientes para a genotipagem de SNPs, como os chips de alta densidade (Van Tassell et al., 2008; Matukumalli et al., 2009), isso mais o aperfeiçoamento dos métodos estatísticos para análise da informação genômica (Nicolazzi et al., 2015), tem fomentado a implementação desses análises.

Os GWAS exploram o desequilíbrio de ligação (LD), que é o nível de associação não aleatória entre os marcadores e as variantes causais (“*quantitative trait loci*-QTL”). Esta associação ocorre porque segmentos de cromossomos são herdados desde um ancestral comum mantendo-se idênticos por descendência, indicando naquela região que a taxa de recombinação é baixa. Se existem QTLs e marcadores no segmento, estes estarão em LD, sendo possível identificar o QTL através dos marcadores. Existem várias metodologias estatísticas que exploram essas associações. A mais simples dentre as metodologias é o teste de associação do genoma usando regressão de marcador único (single SNP regression) (Gondro; Van der Werf; Hayes, 2013). A identificação e análise dos genes localizados em tais regiões fornecem informações importantes sobre os mecanismos genéticos envolvidos na expressão fenotípica da característica, o que pode contribuir para a obtenção de avaliações genéticas acuradas.

Diversos resultados reportados em análises de GWAS mostram que os processos biológicos que controlam a expressão de uma característica ou risco de doença, são mais complexos do que se pensava (VISSCHER et al., 2012). Em alguns casos os GWAS conseguiram identificar com sucesso a variante genética que está associada com o fenótipo e descobrir o mecanismo de ação; em outros casos, os estudos evidenciaram que a característica é poligênica, tendo a influência de milhares de genes ao invés de alguns poucos (CANTOR et al., 2010).

Por exemplo, uma pesquisa feita para estudar a associação genética da suscetibilidade a infecção por *Mycobacterium bovis* em gado leiteiro, usando as metodologias regressão single-SNP e BayesC, identifico uma região candidata localizando associação com um gene que encontrasse envolvido na resposta imune (Richardson et al., 2016).

O método de análise de single SNP regression tem sido usado também para estimar associação genômica com características de produção de leite, gordura e proteína, porcentagem de proteína e gordura, e com fertilidade como característica de importância para desempenho ao parto em gado leiteiro e de corte (Purfield et al., 2015), identificando grande número de SNPs associados significativamente.

Da costa barros et al. (2018), em GWAS para características produtivas identificaram três genes da mesma família (ADAMTS12, ADAMTS8 e ADAMTS15) para a característica de produção de leite, estes genes participam do processo de angiogênese (Kumar et al., 2012). A relação da angiogênese com a produção de leite presumivelmente está relacionada ao suprimento de precursores. Assim, o aumento da vascularização dos alvéolos do úbere está potencialmente associado com a maior produção de leite. Em outra pesquisa para características produtivas e Contagem de Células Somáticas em búfalos, Venturini et al. (2014) conseguiram identificar 118 SNPs relacionados com produção de leite, gordura e proteína; também verifico associação significativa dos genes com as características como o gene POU4F1 que foi significativo para produção de leite, produção de proteína e produção de gordura.

3. OBJETIVOS

Objetivou-se com este estudo identificar regiões genômicas e genes candidatos associados às características de produção de leite, gordura, proteína, e produção de mozzarella, em búfalas da raça Murrah mediante análise de duas diferentes abordagens: single SNP regression e Bayes C.

4. MATERIAL E MÉTODOS

4.1 Dados fenotípicos

Os dados fenotípicos e genotípicos foram provenientes de animais participantes do programa de controle leiteiro mantido pelo Departamento de Zootecnia da FCAV/UNESP-Jaboticabal/SP. Foram usados 76757, 472, 2900, 2843 registros de test-days (TD) referentes às características de produções de leite (PL), gordura (PG), proteína (PP) e *mozzarella* (PMZ) respectivamente, provenientes da primeira lactação de 4068, 372, 488, 482 búfalas da raça Murrah correspondentes a cada característica avaliada. Os animais foram provenientes de seis rebanhos brasileiros localizados nos Estados do Ceará, Rio Grande do Norte e São Paulo, controladas mensalmente entre os anos de 1987 e 2018. A característica PMZ foi calculada como: $PMZ (kg) = PL * \{[(3.5 * \%P) + (1.23 * \%G) - 0.88] / 100\}$, utilizada anteriormente por Altiero et al. (1989).

O conjunto de dados foi editado para eliminar os seguintes registros de búfalas: com identificação inadequada; com menos de três registros de TD por lactação; sem informação de datas de nascimento ou parto. Os registros TD corresponderam à 5 a 305 dias. Foi assumido idade ao primeiro parto entre 21 e 86 meses. A idade da vaca ao parto foi agrupada em 7 classes. Os dias em lactação (DIM) foram agrupados em 30 classes com intervalos de 10 dias. Para remover *outliers*, registros de TD com médias acima ou abaixo de 3.5 desvios padrão fenotípico dentro de rebanho foram excluídos. Os grupos contemporâneos foram definidos pela combinação de rebanho ano e mês do TD (HTD), com a restrição de que cada HTD tivesse que conter pelo menos 4 animais. Após a edição dos dados, as estatísticas descritivas para PL, PG, PP e PMZ estão descritas na Tabela 1. O arquivo de pedigree consistiu de 5561 animais.

Tabela 1. Estatísticas descritivas para as características de produção de leite (PL), gordura (PG), proteína (PP), mozzarella (PMZ).

Características	N	Média (Kg)	Mín (Kg)	Máx (Kg)	Desvio padrão (Kg)
PL	76757	7,63	0,5	21,2	3,08

Características	N	Média (Kg)	Mín (Kg)	Máx (Kg)	Desvio padrão (Kg)
PG	472	0,47	0,09	1,38	0,21
PP	2900	0,35	0,034	10,04	0,22
PMZ	2843	1,75	0,15	5,23	0,67

4.2 Genótipos e controle de qualidade

Um total de 960 animais foram genotipados utilizando o painel 90K Axiom® Buffalo Genotipagem (Affymetrix). O controle de qualidade foi realizado adotando-se os seguintes critérios de exclusão dos SNPs: MAF (Frequência de Alelos Menores) < 0,05; Call rate < 0,95 e Equilíbrio de Hardy-Weinberg (HWE) p-value 10^{-6} . Os SNPs coincidentes também foram eliminados. Foi considerada Call rate < 0,98 como critério de exclusão da amostra. Após os critérios adotados, um total de 45.590 SNPs foi considerado para o presente estudo.

4.3 Predição de valores genéticos (EBVs)

Os EBVs foram preditos por meio de um modelo animal de repetibilidade. As análises foram realizadas separadamente para cada característica (PL, PG, PP e PMZ). Em notação matricial o modelo pode ser descrito como segue:

$$y = Xb + Za + Wp + e$$

Onde, $y \sim N(Xb, V)$ refere-se ao vetor de registros de TD com matriz de covariância V ; b refere-se ao vetor de efeitos fixos (combinação de rebanho ano e mês do TD - HTD, combinação de rebanho e classes de idade ao parto -HAGE, combinação de rebanho e classes de dias em lactação – HDIM (30 classes) e número de ordenhas por dia (1 ou 2)); $a \sim N(0, G)$ refere-se ao vetor de efeitos aleatórios genéticos aditivos, em que $G = A\sigma_a^2$, sendo A a matriz de parentesco baseada em pedigree e σ_a^2 a variância genética aditiva; $p \sim N(0, P)$ refere-se ao vetor de efeito

aleatório de ambiente permanente, em que $\mathbf{P} = I\sigma_{pe}^2$, sendo I uma matriz identidade e σ_{pe}^2 a variância de ambiente permanente; $e \sim N(0, \mathbf{R})$ refere-se ao vetor de efeitos de erro aleatório, em que $\mathbf{R} = I\sigma_e^2$, sendo σ_e^2 a variância residual; \mathbf{X} , \mathbf{Z} , e \mathbf{W} referem-se as matrizes de incidência de efeitos fixos e efeitos aleatórios genético aditivo e de ambiente permanente, respectivamente. Finalmente, \mathbf{V} é definida como: $\mathbf{V} = \mathbf{ZGZ}' + \mathbf{WPW}' + \mathbf{R}$. Os componentes de variância e a predição dos valores genéticos foram realizadas por meio do software REMLF90 (Misztal et al., 2002).

4.4 Pseudo-fenótipos (dEBV)

Os EBVs preditos no passo anterior foram desregredidos (dEBV) por meio do método proposto por VanRaden (2009). Este método de desregressão de dados longitudinais têm sido associados por apresentar melhores acurácia e menos viés (Oliveira et al. 2018) e pode ser descrito como:

$$dEBV_{ij} = PA_{ij} + \frac{(EBV_{ij} - PA_{ij})}{\frac{ERC_{(Ind-PA)i}}{ERC_{(Ind-PA)i} + ERC_{PAi+1}}},$$

Onde PA_{ij} e EBV_{ij} refere-se a média parental (PA) e ao EBV para o animal i no dia j . As contribuições efetivas de registros (ERC) para o animal i , excluindo a informação de PA do animal i foram estimadas de acordo com VanRaden & Wiggans, (1991), como:

$$ERC_{(Ind-PA)i} = \left(\lambda \frac{REL_{EBVi}}{1 - REL_{EBVi}} \right) - ERC_{PA},$$

$$ERC_{PAi} = \lambda \frac{REL_{PAi}}{1 - REL_{PAi}},$$

em que $\lambda = (1-h^2)/h^2$, REL_{EBVi} e REL_{PAi} referem-se as médias de confiabilidades para o EBV e PA do animal i para cada característica analisada.

As médias de confiabilidade do EBV foram 0.50 ± 0.05 para PL, 0.18 ± 0.05 para PG, 0.05 ± 0.06 para PP, e 0.06 ± 0.07 para PM. Somente fêmeas com confiabilidades > 0.02 foram selecionadas para análise de GWAS, totalizando 638 fêmeas para PL, 284 para PG, 236 para PP, e 237 para PM.

4.5 Análises de associação genômica ampla (GWAS)

Devido ao caráter poligênico das características (leite, gordura, proteína e mozzarella), as análises de GWAS foram baseadas em duas distintas abordagens: análise de regressão single-SNP e Bayes C.

4.5.1 Bayes C

A metodologia BayesC utilizada para as análises de GWAS foi proposta por Habier et al. (2011). Nessa metodologia, o efeito dos SNPs possui variância comum que segue uma distribuição chi-quadrado escalada invertida a *priori* (Habier et al., 2011) O modelo geral usado pode ser descrito como segue:

$$y = 1u + \sum_{i=1}^n g_i b_i \delta_i + e$$

onde y é o vetor dos dEBVs para cada uma das características (PL, PP, PG, PMZ) μ é a média geral, g refere-se ao vetor de genótipos dos animais para efeito de SNP i , b_i e δ_i referem-se a uma variável indicadora, que assume valor 0 ou 1, amostrada a partir de uma distribuição binomial com os parâmetros n e π , em que n é o número total de SNP e π a proporção de SNP com efeito nulo no modelo. Foi assumido que π segue uma distribuição beta a priori com parâmetros $\alpha = 10^8$ e $\beta = 10^{10}$, que equivale fixar π em 0,99 (Legarra et al., 2014).

As análises com a metodologia BayesC foram realizadas utilizando o software GS3 (Legarra et al., 2013), considerando um total de 500.000 iterações para a cadeia

de Markov Monte Carlo (MCMC), e período de *burn-in* de 200.000, retirando-se uma amostra a cada 50 ciclos. A convergência foi verificada através do pacote BOA (Smith, 2007), do software R.

O Fator de Bayes (BF) foi utilizado para avaliar a significância dos SNPs sobre as características, e foi definido de acordo com Varona et al., (2001) como:

$$BF = \frac{\binom{fi}{1-fi}}{\binom{\pi 1}{\pi}}$$

em que f_i refere-se ao número de vezes que SNP i foi incluído no modelo, π a probabilidade posteriori do SNP i não estar associado a característica e $\pi 1$ a probabilidade a posteriori do SNP i estar associado (representada como $1 - \pi$). Fatores de Bayes maiores que 3.2, 10 e 100 foram considerados como substancialmente, fortemente e decisivamente associados a uma característica, respectivamente.

4.5.2 Single SNP regression

As análises de regressão via marcas únicas foram baseadas em modelo misto conforme o modelo:

$$y = 1\mu + \mathbf{b}_x + g_j + e_j$$

em que y refere-se ao vetor dos dEBVs para cada uma das características (PL, PP, PG, PMZ); μ é a média geral; b é o efeito aditivo (efeito fixo) do SNP candidato a ser testado quanto à associação, x a variável indicadora do genótipos codificado como 0, 1 ou 2, , g_j o efeito acumulado de todos os SNPs (efeito aleatório poligênico) e e_j o erro residual. Este modelo foi implementado no software GCTA (Yang et al., 2013).

4.5.3 Regiões de QTL e genes candidatos

SNPs que foram significativos em ambas as abordagens Bayes C e single SNP regression considerando respectivamente um nível de significância de $BF > 3.2$ e $(3.0$ na base $-\log_{10})$ foram considerados importantes regiões a serem investigadas. Além disso, SNPs considerados fortemente associados em uma única abordagem ($BF > 10$ para Bayes C ou 5.0 na base $-\log_{10}$ para single SNP regression) também foram investigados.

As regiões de QTL de interesse foram definidas a partir de 100kb antes e depois da posição do SNP, definidas com base no desequilíbrio de ligação (LD) em torno de cada SNP selecionado. Para examinar o decaimento do LD, o r^2 em relação a distância física foi calculado usando o plink (Purcell et al., 2007). O nível de LD foi usado para indicar o início e o fim de um QTL. Esses QTLs foram então investigados quanto à presença de genes usando o Banco de dados de genes para *Bubalus bubalis* disponível em National Centro de Informações sobre Biotecnologia (NCBI, <https://www.ncbi.nlm.nih.gov/genome/gdv/?org=bubalus-bubalis&group=bovidae>).

Para todos os genes candidatos identificados, uma revisão na literatura foi realizada para melhor reportar a função biológica e molecular associada ao gene e as características estudadas. Além disso, para validar a relevância dos potenciais genes candidatos, análises de ontologia de genes foram realizadas para descrever as funções biológicas, processos imunológicos e funções moleculares associados aos genes usando o plug-in ClueGO do software Cytoscape (Bindea et al., 2013, 2009). O ClueGO calcula o teste de enriquecimento com base na distribuição hipergeométrica unilateral e na correção de Bonferroni, e a rede de genes criada representa os termos como nós que são vinculados com base nos níveis de escore kappa $\geq 0,4$.

5. RESULTADOS E DISCUSSÃO

5.1 Estimativas de componentes de variância

As estimativas de herdabilidade foram moderada para todas as características, variando de 0,19 a 0,23 (Tabela 2). Esses resultados apontaram que grande parte da variação fenotípica total se deve à ação genética aditiva dos genes. Portanto, um ganho genético considerável é esperado se a seleção for aplicada para qualquer característica.

A estimativa de herdabilidade para PL foi próxima às obtidas por Tonhati et al. (2000, 2008) e para PG, PP, PMZ foram iguais com as obtidas por Aspilcueta-Borquis et al. (2010a, 2010b), que também trabalharam com búfalos Murrah. No entanto, estimativas de herdabilidade para PL, PP e PG mais altas (0,28, 0,19 e 0,35) do que as obtidas no presente estudo, foram relatadas por Liu et al. (2018) com búfalos Mediterrâneos na Itália.

Em geral, as estimativas dos componentes de variância das características apresentaram menores valores quando comparados com outras pesquisas realizadas em búfalos da raça Murrah (Barros et al., 2016; Camargo et al., 2015, Liu et al. 2018). Diferenças na variabilidade genética da raça, na metodologia aplicada para estimar os componentes de variância, e também no sistema de produção (manejo e ambiente) (Aspilcueta-Borquis et al., 2010a) provavelmente explicam a diferença entre os resultados achados na literatura com aqueles obtidos no presente estudo.

Tabela 2. Estimativas dos componentes de variância e herdabilidade para as características de produção de leite (PL), gordura (PG), proteína (PP), mozzarella (PMZ).

	PL	PG	PP	PMZ
σ^2_a	1.065	0.005	0.001	0.056
σ^2_{pe}	1.933	0.005	0.000	0.000
σ^2_e	2.744	0.022	0.007	0.190
h^2	0.19	0.15	0.16	0.23

* σ^2_a = Variância aditiva. σ^2_{pe} = Variância de ambiente permanente. σ^2_e = Variância residual. h^2 = Herdabilidade.

5.2 Genes candidatos em ambas abordagens

Os Manhattan plots das associações entre cada SNP e as características estão demonstrados na Figura 1 para a abordagem Bayes C e na Figura 2 para a abordagem single SNP regression. SNPs que foram significativos em ambas abordagens Bayes C e single SNP regression, considerando respectivamente um nível de significância de $BF > 3.2$ e 3.0 na base $-\log_{10}$ respectivamente, representados pela linha azul, foram considerados importantes regiões a serem investigadas.

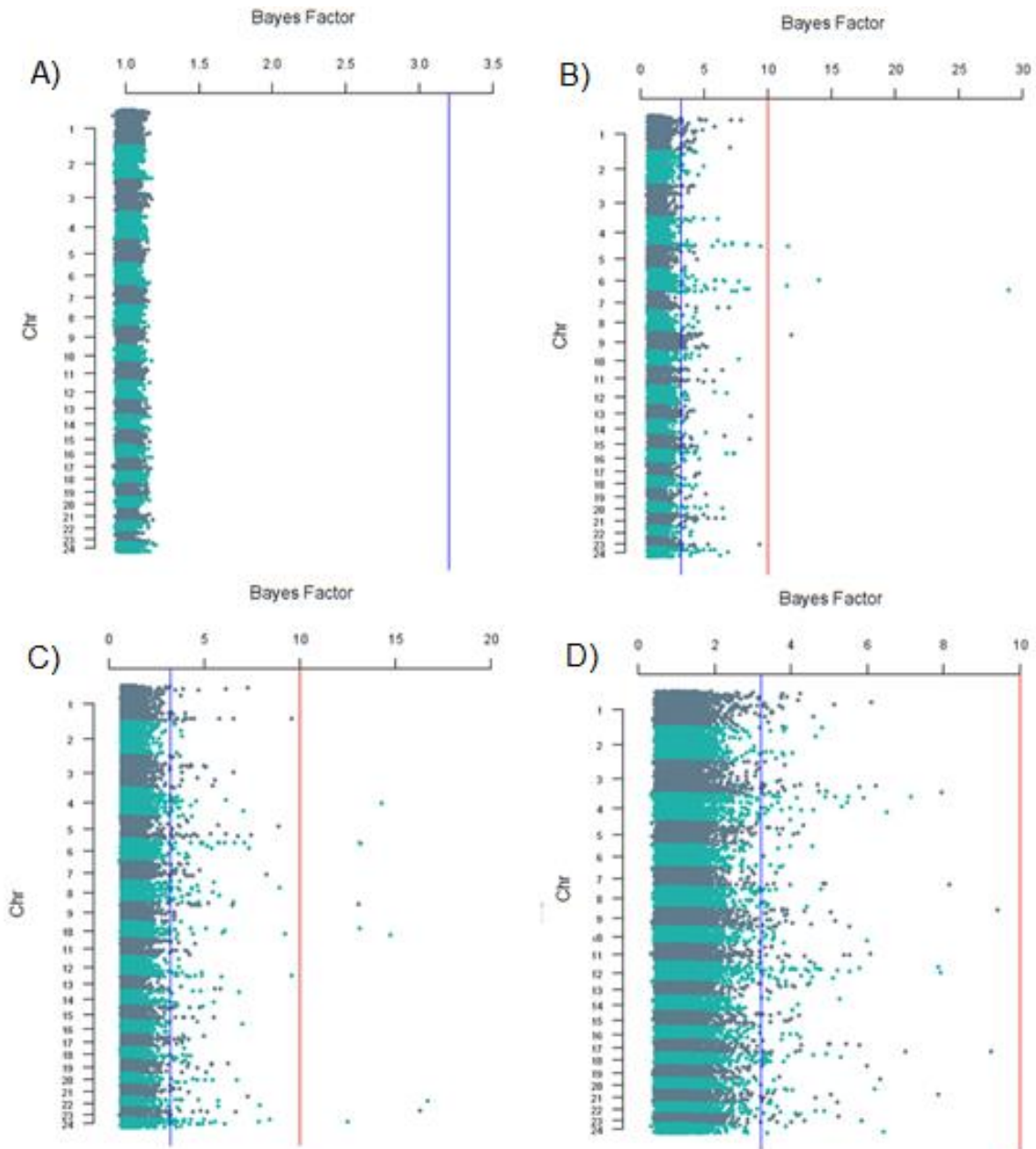


Figura 1. Manhattan plots das associações entre cada SNP e produção de leite (A), produção de gordura (B), proteína (C) e produção de mozzarella (D) provenientes das abordagens Bayes C.

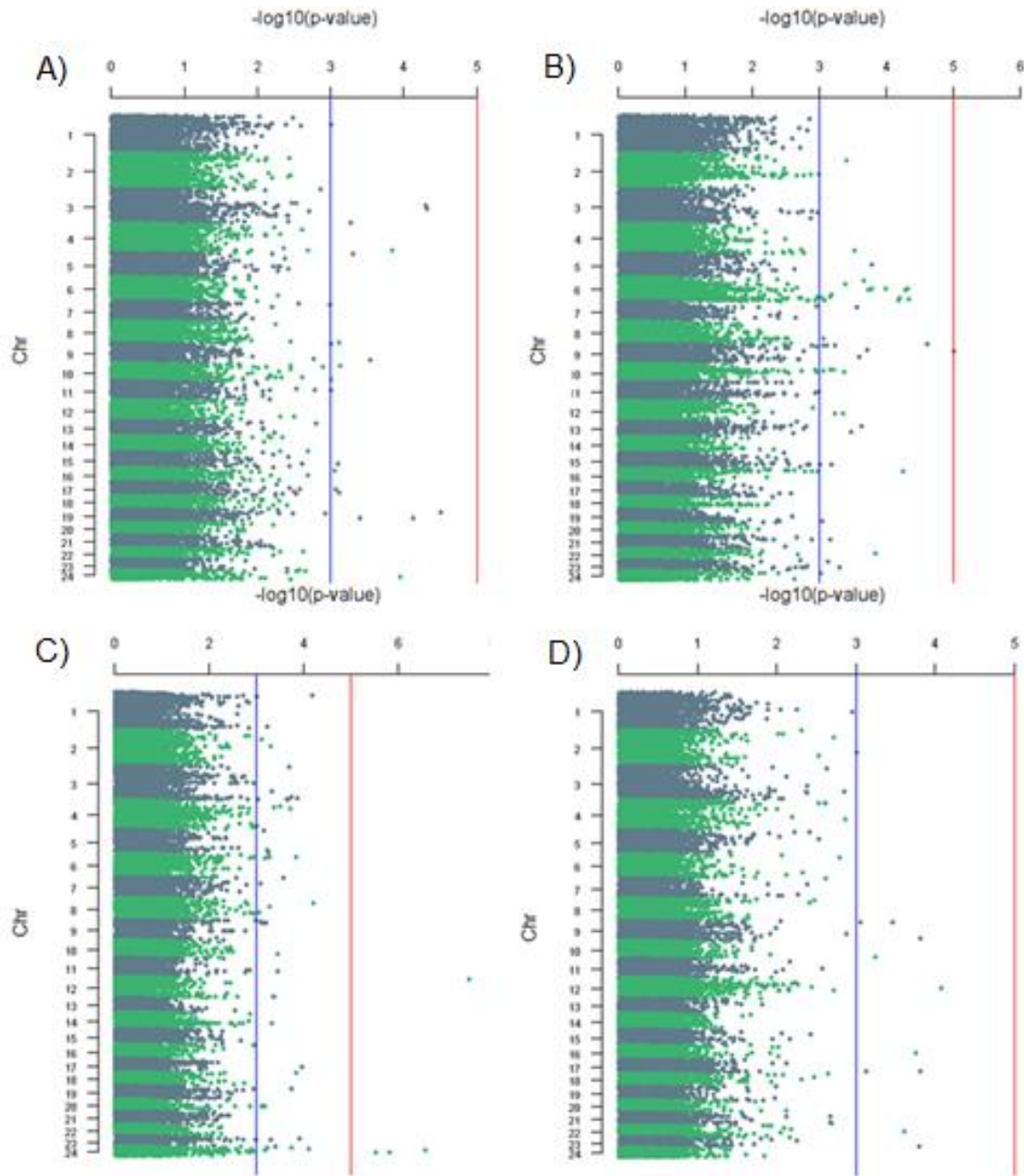


Figura 2. Manhattan plots das associações entre cada SNP e produção de leite (A), produção de gordura (B), proteína (C) e produção de mozzarella (D) proveniente da abordagem de single SNP regression.

Os gráficos *qq-plots* foram construídos (Figura 3) para avaliar magnitude das associações observadas entre marcadores e fenótipos (PL, PG, PP e PMZ), bem como identificar potenciais problemas de estrutura populacional. Desvios da linha de

identidade sugerem que a amostra contém valores que surgem nas extremidades, possivelmente devido a uma verdadeira associação.

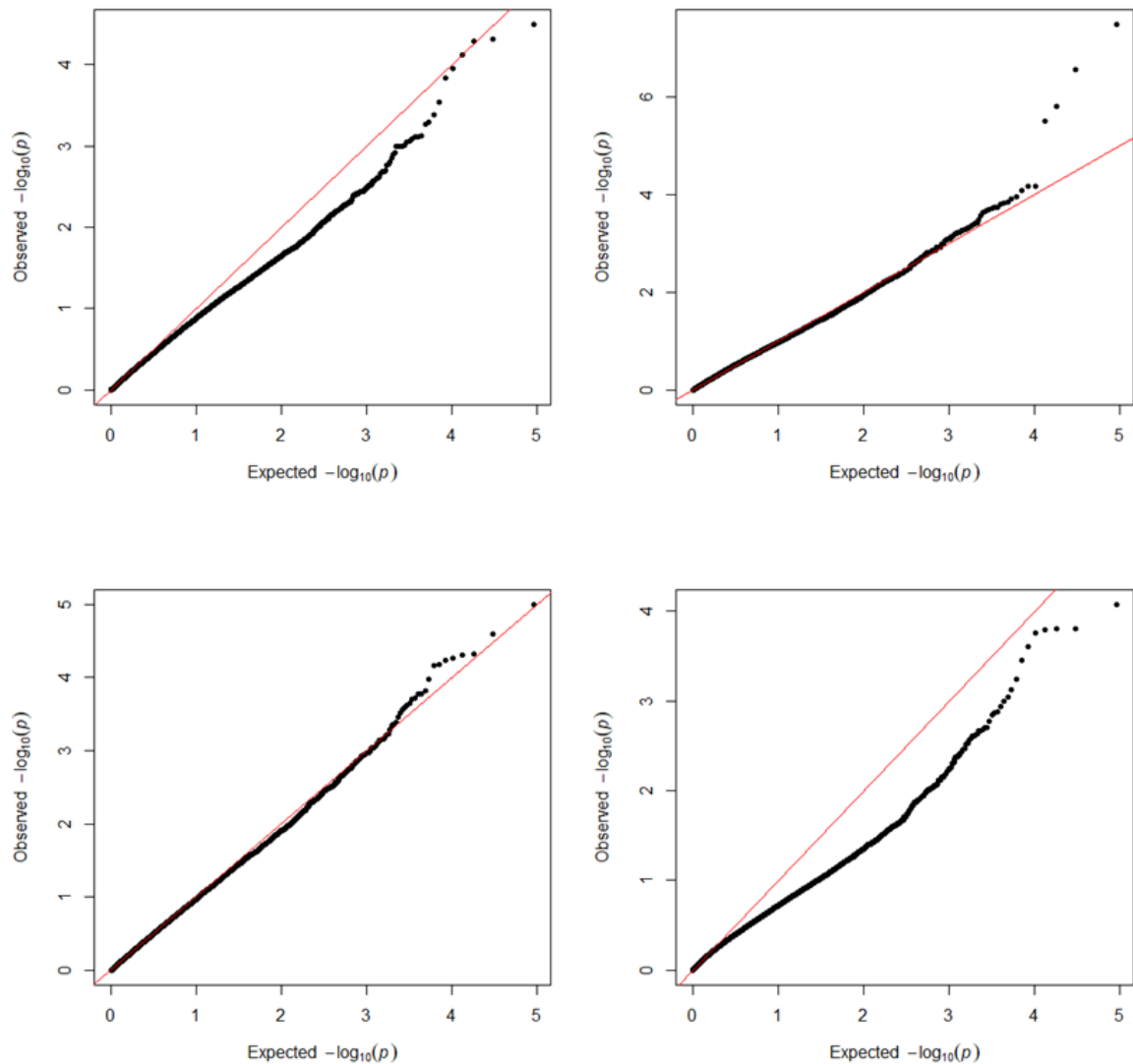


Figura 3. QQ-plot para P-values ($-\text{Log}_{10}$) estimados por meio da análise de single SNP regression.

Um diagrama de Venn também foi usado para apresentar o número de SNPs significativos que foram comuns entre as duas abordagens para PG, PP PM (Figura 4). Um total de 30 SNPs comuns em ambas abordagens (Bayes C, single SNP

regresion) foram selecionados para PG, 18 para PP e 5 para PMZ (Figura 4). Essas regiões foram consideradas importantes para pesquisa de gene candidato. No entanto, nenhum SNP foi selecionado para PL como mostrado nas figuras 2 e 3.

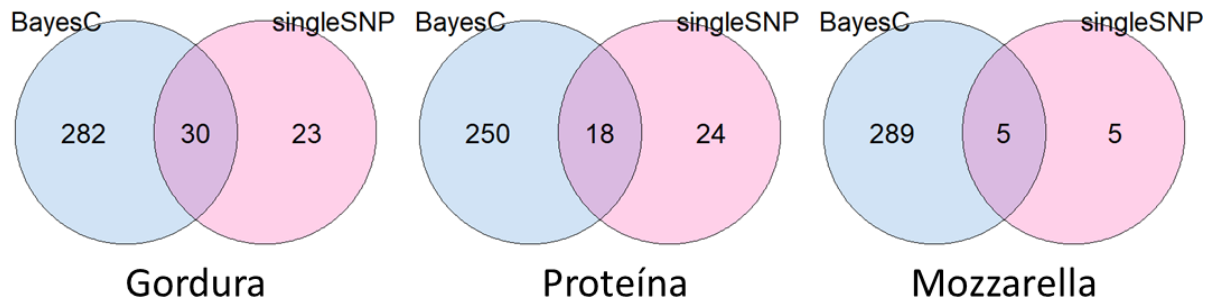


Figura 4. Número de regiões significativas comuns entre as abordagens BayesC e single SNP regression para produção de gordura, proteína e mozzarella.

O conceito de desequilíbrio de ligação (DL) foi importante para definir as regiões de QTL. O DL de acordo com intervalos de distância é demonstrado na Figura 5. Valores mais altos de DL foram observados somente em distâncias pequenas entre pares SNPs o que poderia indicar que são segregados juntos. O DL decai rapidamente conforme a distância entre os SNPs aumentam. Resultados semelhantes foram reportados por Deng et al. (2019) investigando a extensão do DL em búfalos da raça Mediterranea, onde uma tendência de declínio rápido dos valores de r^2 foi encontrada em curtas distâncias entre marcadores (<100 kb). Os valores de DL foram relevantes para informar um apropriado comprimento da janela para pesquisa de genes candidatos. Portanto um tamanho de 100kb foi definido para as janelas ($r^2 = 0.18$).

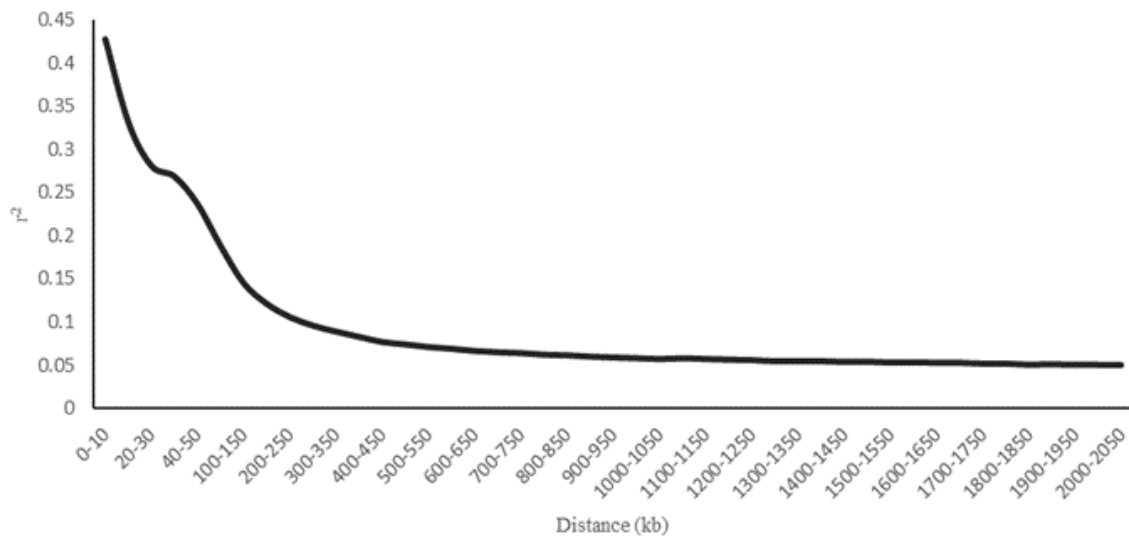


Figura 5. Desequilíbrio de ligação médio em Búfalos da raça Murrah de acordo com intervalos de distância (Kb).

Os genes anotados mais próximos aos marcadores significativamente identificados para as características estudadas comuns em ambas abordagens (Bayes C, single SNP regression) são apresentados na Tabela 3. A posição dos SNPs e dos genes anotados, utilizaram como referência o genoma bubalino *Bubalus bubalis* UOA_WB_1. Um total de 117, 14 e 8 genes foram identificados em regiões de QTL relevantes para PG, PP e PM, respectivamente.

Tabela 3. Regiões de QTL em que genes candidatos foram associados com produção de leite, gordura, proteína, e mozzarella em comum as análises de Bayes C e single SNP regression.

Chr	QTL (região bp)	SNP	BF	p value	Genes
Gordura					
1	17758439-17958439	AX-85051525	6,0651	4,1638	-
1	184672265-184872265	AX-85056686	6,5066	3,2156	<i>DSCAM</i>
2	49944180-50144180	AX-85048467	3,733	3,0973	-

Chr	QTL (região bp)	SNP	BF	p value	Genes
2	88336710-88536710	AX-85102005	3,7924	3,2782	<i>RBMS1</i>
3	133411215-133611215	AX-85051207	5,4781	3,3097	<i>NKX2-6, STC1</i>
3	164316280-164516280	AX-85077255	5,4069	3,7298	<i>PTBP3, LOC112577609, LOC112583976, HSDL2, KIAA1958</i>
3	164662616-164862616	AX-85099304	3,2118	3,8424	<i>LOC112583979, LOC102406480, SLC46A2</i>
5	108806558-109006558	AX-85108678	4,6038	3,2114	<i>JAM3, TRNAG-CCC, LOC102409316, LOC112584998</i>
6	16103573-16303573	AX-85066094	6,5565	3,2695	<i>NUP210L, LOC112585550, LOC112585549, RPS27, RAB13, JTB, CREB3L4, SLC39A1, CRTC2, DENND4B, GATAD2B</i>
6	16327896-16527896	AX-85051557	5,7375	3,2272	<i>INTS3, NPR1, TRNAM- CAU, ILF2, SNAPIN, CHTOP, S100A1, S100A13, S100A14, S100A16, LOC112585551, S100A2, S100A3</i>
6	20037608-20237608	AX-85117685	13,082	3,8226	<i>PRPF3, MRPS21, CIART, C6H1orf54, APH1A, CA14, ANP32E, PLEKHO1, LOC102390368, VPS45</i>
6	21713652-21913652	AX-85057069	13,134	3,1701	<i>LOC112585572, GJA5, ACP6, BCL9,</i>
9	4693488-4893488	AX-85125980	13,018	3,0798	-
9	15592079-15792079	AX-85098212	4,7065	3,0831	<i>RHOBTB3, SPATA9, RFESD, GPR150, ARSK, LOC112586798, TTC37</i>
9	16828988-17028988	AX-85094662	6,4384	3,1319	<i>KIAA0825</i>
9	16863918-17063918	AX-85073301	4,3015	3,1929	<i>KIAA0825</i>
10	76128449-76328449	AX-85110712	14,714	3,4262	<i>TRDN</i>
11	68751945-68951945	AX-85115761	4,0838	3,4274	-

Chr	QTL (região bp)	SNP	BF	p value	Genes
12	97555894-97755894	AX-85109191	4,8138	3,3289	GARNL3, LOC112578303, SLC2A8, ZNF79, RPL12, LOC112578261, LRSAM1, FAM129B
12	97664822-97864822	AX-85087066	4,4497	3,3673	GARNL3, SLC2A8, ZNF79, RPL12, LOC112578261, LRSAM1, FAM129B, LOC112578235, STXBP1
18	5241514-5441514	AX-85101967	3,8946	3,8157	WVOX, LOC112580377
19	14297927-14497927	AX-85112838	5,3174	3,7282	CWC27, LOC102399846, SREK1IP1, SHISAL2B
20	35620611-35820611	AX-85102033	4,9665	3,0883	CLK3, EDC3, LOC112580979, LOC102399021, LOC102399561, LOC112580792, CSK, LMAN1L, CPLX3, ULK3
20	35643700-35843700	AX-85077580	5,3665	3,1048	EDC3, LOC102399021, LOC102399561, LOC112580792, CSK, LMAN1L, CPLX3, ULK3, SCAMP2
24	3313165-3513165	AX-85070001	8,3916	3,7383	FOXK1, AP5Z1, RADIL, PAPOLB
24	3349351-3549351	AX-85085095	4,2214	3,2220	FOXK1, AP5Z1, RADIL, PAPOLB, WIPI2
24	11252885-11452885	AX-85079105	3,2565	4,0921	LOC112581917
24	14221985-14421985	AX-85057624	5,5083	3,1723	LOC102406545, SBDS, TMEM248, RABGEF1, LOC112581839, LOC102409786, TPST1
24	14823586-15023586	AX-85101478	12,489	6,5704	CHCHD2, PHKG1, SUMF2, LOC112581831, CCT6A, LOC112581924, LOC112581925, PSPH, TRNAE-CUC, NIPSNAP2, MRPS17, ZNF713, SEPT14, LOC102400066
24	26804293-27004293	AX-85047063	5,9975	5,5080	XYLT1
Proteína					

Chr	QTL (região bp)	SNP	BF	p value	Genes
4	153732634-153932634	AX-85099723	4,0910	3,6127	<i>LOC102391785</i>
6	107116059-107316059	AX-85100321	8,6383	3,4594	-
6	111794832-111994832	AX-85113021	6,6695	4,2396	<i>ZSCAN20, LOC112585732, EFHD1, GIGYF2</i>
6	114941551-115141551	AX-85111323	6,4781	3,1429	<i>AGAP1</i>
6	34449973-34449973	AX-85113845	3,9423	3,8225	<i>STXBP3</i>
6	63444857-63444857	AX-85108716	5,4032	3,0211	-
6	63514797-63514797	AX-85084361	8,3163	3,5168	-
6	63544111-63544111	AX-85085445	3,6056	4,1789	-
6	67542694-67542694	AX-85121042	28,8662	4,3184	<i>ST6GALNAC3</i>
6	67641068-67641068	AX-85080063	7,4676	4,1673	<i>ST6GALNAC3</i>
6	90175697-90175697	AX-85100962	4,0420	3,3735	-
9	40175225-40375225	AX-85126308	6,4526	3,7786	<i>EBF1</i>
13	31054969-31254969	AX-85114226	5,8799	3,9830	-
13	64851220-65051220	AX-85048676	5,7290	4,3089	<i>NBEA, DCLK1</i>
16	19407010-19607010	AX-85096786	6,5448	3,7116	<i>PDHX, APIP</i>
21	20401660-20601660	AX-85050563	7,6182	4,2705	-
22	24155234-24355234	AX-85095943	3,3337	3,1616	<i>DLGAP1</i>
24	11140672-11340672	AX-85083271	3,5971	4,9948	-
Mozzarella					
9	8784289-8784289	AX-85075354	9,39	3,4493	<i>PAM, LOC102405618, LOC102405384</i>
9	94944640-95144640	AX-85089465	3,23	3,8021	<i>ARHGEF18, INSR</i>
9	9690329-9690329	AX-85084147	3,88	3,0411	<i>LOC112587108</i>
12	47887900-65456478	AX-85046583	7,92	4,0713	<i>LOC112578244, LOC102393921</i>
17	55173873-72742451	AX-85124400	9,23	3,1199	-

Chr: cromossomo. BF: Bayes Factor.

Diversos genes identificados em nosso estudo foram previamente descritos em estudos semelhantes (De Camargo et al., 2015; Venturini et al., 2014; Du et al., 2019). No entanto, novos potenciais genes candidatos foram associados com importantes funções biológicas e moleculares, que podem ser importantes para compreender os principais mecanismos de ação envolvidos nas características de produção de leite. Portanto, suas principais funções serão exploradas na sequência. Além disso, os principais processos biológicos, imunes e funções moleculares foram sintetizados em uma rede gênica para PG, (Figura 6), PP (Figura 7) e PM (Figura 8).

5.2.1 Gordura

Um total de 117 genes foram reportados em nosso estudo como associados com a característica de produção de gordura distribuídos em 13 regiões. Na tabela 3, pode-se identificar que o cromossomo 6 continha a maior quantidade de genes, e o valor de BF mais alto (13,13). Também foi observado que o cromossomo 24 foi o segundo em apresentar maior quantidade de genes com um total de 32, além disso, nessa região o p value foi de 6,57, sendo o mais alto para essa característica (Tabela 3).

. Alguns desses genes já foram reportados e associados com a característica em outros estudos de GWAs em búfalos (Venturini et al., 2014; Camargo 2015). Sendo que a reprodução do búfalo luta com uma alta incidência de mortalidade embrionária precoce, um estudo feito por Panigrahi et al. (2020) para tentar compreender as vias moleculares dos perfis transcriptômicos endometriais de búfalos grávidas durante o início da gravidez, descreve o gene *CREB3L4* que em nosso estudo foi associado com produção de gordura, como um gene com maior expressão regulada negativamente em tecidos endometriais de animais grávidas contribuindo na afetação de funções que potencialmente modulam o ambiente endometrial para estabelecer e manter uma gravidez bem-sucedida, diminuindo a quantidade de partos nos rebanhos, baixando a produção de leite e seus derivados, assim como também levando ao produtor a perdas econômicas, pela perda de crias e a baixa produção leiteira.

Como o volume de leite secretada pelas fêmeas bovinas, encontra-se regulado na glândula mamaria pela quantidade de lactose sintetizada nas células (Vivas 2017), Hill et al. (2016) tem descrito o gene *CRTC2* como coativador na gluconeogenesis (Figura 6), processo indispensável para sintetizar a glicose e aumentar os níveis de lactose que influenciam a quantidade de leite produzida. Uma pesquisa feita em camundongos para explorar o papel do gene *CRTC2* na síntese de gordura e proteína no leite induzida por aminoácidos, demonstraram que o gene *CRTC2* se expressa maiormente na glândula mamaria no período da lactação, sendo um coativador de transcrição na síntese de gordura do leite nas células epiteliais mamárias (Li et al., 2019), dando sentido que em nosso estudo este gene tivesse associação significativa com a produção de gordura.

A cetose bovina é uma doença metabólica onde a demanda energética é superior a capacidade de ingestão de alimentos pelo animal, contribuindo a perdas econômicas associadas à morte, serviços veterinários, descartes, redução da taxa de concepção, aumento do intervalo entre partos situação que afeta fortemente a produção de leite (Garzon, et al., 2018) e consequentemente, componentes como a gordura. Um estudo realizado por Fugate et al. (2014) para identificar polimorfismos de nucleotídeo único (SNP) que estão associados à cetose subclínica em gado Jersey, identificou o gene *ILF2* como associado significativamente com essa doença.

Usando as homologias cromossômicas búfalo-bovinas Venturini et al. (2014) realizou um estudo de GWAs para identificar SNPs em búfalos, associados a características de produção e qualidade do leite e contagem de células somáticas, identificando o gene *RPS27* associado significativamente com produção de leite, gordura e proteína; e o gene *RHOBTB3* associado com porcentagem de gordura além de ter associação com produção de leite e proteína, genes que em nosso estudo tiveram a mesma associação.

Uma pesquisa de Xu et al. (2019) para descobrir proteínas hepáticas críticas que afetam o metabolismo da gordura, proteína e lactose do leite em gado Holstein chinês, traçaram o perfil dos proteomas do fígado em três períodos da lactação, descobrindo os genes funcionais potenciais para a síntese do leite, sendo identificado entre eles o gene *S100A1* associado com produção de proteína, gordura e leite.

O gene *S100A13* associado nesse estudo com produção de gordura, corresponde à ativação das células endoteliais no processo de angiogênese na endometriose (Hayrabyan et al., 2005). Uma vez que a angiogênese favorece na viabilidade dos embriões em ovinos (Rivas et al., 2006), possivelmente esse processo fisiológico aconteça do mesmo jeito em búfalos, sendo assim a viabilidade embrionária um fator importante na reprodução, influenciando diretamente na produção de leite e seus componentes.

Durante a lactação, o epitélio mamário requer um suprimento contínuo de glicose para a síntese dos diferentes componentes do leite: lactose, lipídios e proteínas (Reynolds et al., 1994; Xiao e Cant, 2005). Uma pesquisa para determinar o padrão de expressão de importantes transportadores e seus reguladores na glândula mamária durante diferentes estágios fisiológicos de búfalos, identificou o gene *SLC2A1* como um dos principais transportadores associados à captação de glicose na glândula mamária do búfalo (Sharma et al., 2014), podendo explicar a associação desse gene com produção de gordura nesse estudo.

Diferentes estudos feitos em camundongos descrevem genes identificados nesse estudo como implicados em diferentes funções na glândula mamaria, podendo ser que aconteçam do mesmo jeito no metabolismo bubalino. Numa pesquisa, Kambe et al. (2008) para identificar as funções específicas e adaptação dos genes da família *Zips* durante a deficiência de Zinco, demonstraram a importância do gene *SLC39A1* quando esse mineral é deficiente, sendo que baixos níveis de Zinco afetam todas as fases do processo reprodutivo nas fêmeas bovinas, desde o estro e gestação até a lactação (Valencia, 2016). Assim mesmo, outra pesquisa em camundongos sugere que o gene *STC1* tem efeitos no desenvolvimento da glândula mamaria (Figura 6), o teor de gordura no leite, o crescimento pós-natal do filhote e comportamento materno (Zaidi et al., 2006).

Outros genes como o *TRNAG-CCC* associado significativamente com gordura nesse estudo de GWAS, foi identificado em outros estudos realizados para identificação de SNPs, regiões genômicas e genes que afetam as características de produção em búfalas (De Camargo et al., 2015) e em bovinos encontrasse associado

significativamente com contagem de células somáticas e produção de lactose respectivamente (Costa et al., 2019).

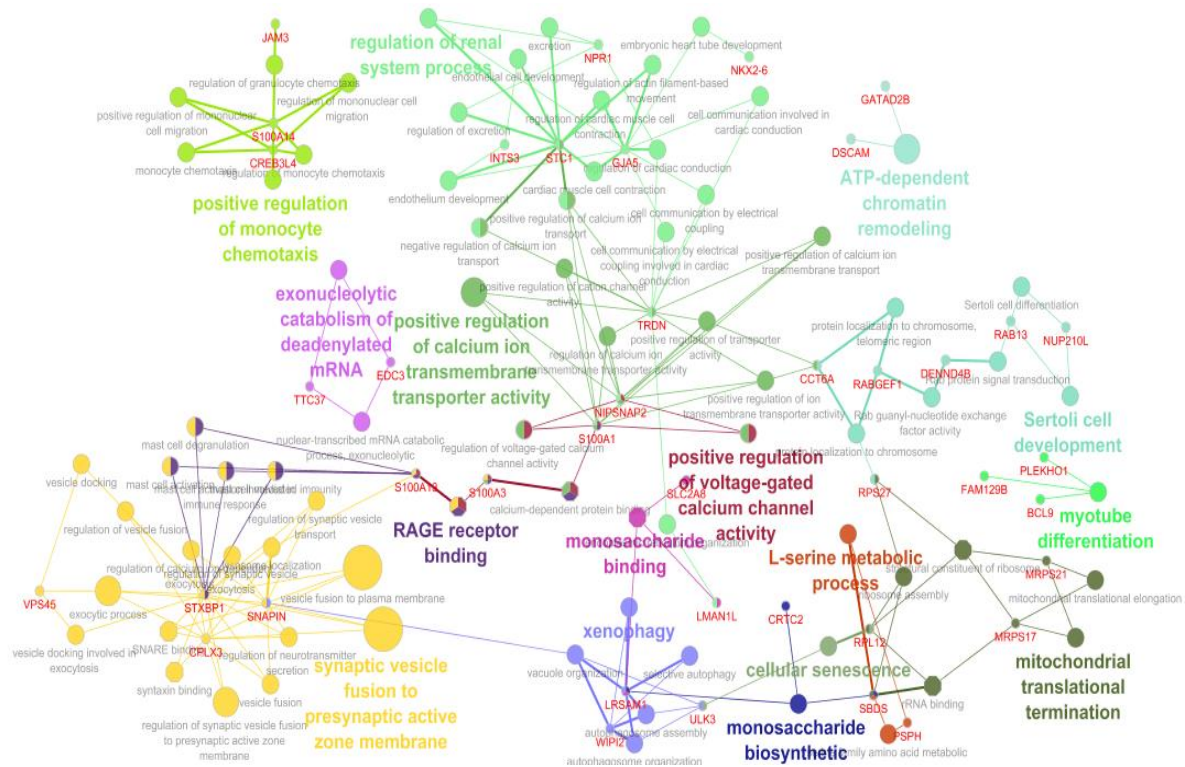


Figura 6. Rede gênica de processo biológico (círculo), processos imunológicos (hexágono) e funções moleculares (octógono) para os potenciais genes candidatos identificados para produção de gordura (105 genes foram incluídos nesta análise). Os nomes em vermelho referem-se aos genes observados. Os termos mais enriquecidos por grupo são mostrados em destaque.

5.2.2 Proteína e Mozzarella

Para a característica de produção de proteína foram identificadas 8 regiões associadas, sendo o cromossomo 6 quem apresentava o BF mais alto (28,86) e continha a maior quantidade de genes (Tabela 3). Para a característica de produção de Mozzarella, foram identificadas 3 regiões, onde os cromossomos 9 e 12 apresentavam os valores mais altos de BF e p-value respectivamente. Os genes anotados mais próximos dos SNPs associados com as características de produção de proteína e produção de mozzarella, ainda não foram identificados e descritos na

Figura 7. Rede gênica de processo biológico (círculo), processos imunológicos (hexágono) e funções moleculares (octógono) para os potenciais genes candidatos identificados para produção de proteína (12 genes foram incluídos nesta análise). Os nomes em vermelho referem-se aos genes observados. Os termos mais enriquecidos por grupo são mostrados em destaque.

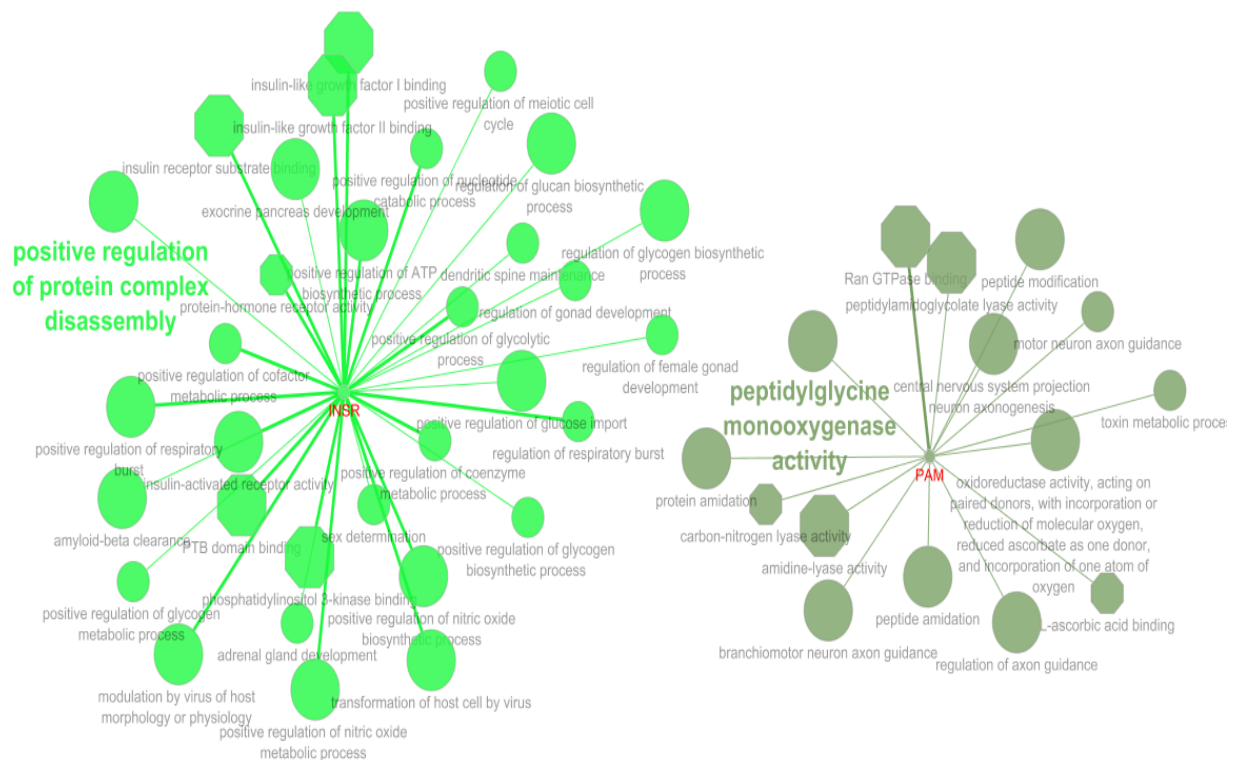


Figura 8. Rede gênica de processo biológico (círculo), processos imunológicos (hexágono) e funções moleculares (octógono) para os potenciais genes candidatos identificados para produção de mozzarella (3 genes foram incluídos nesta análise). Os nomes em vermelho referem-se aos genes observados. Os termos mais enriquecidos por grupo são mostrados em destaque.

5.3 BayesC

Na Tabela 4 encontrassem descritos os SNPs fortemente associados ($BF > 10$) pela abordagem BayesC. Foram identificados 9 SNPs para gordura, 5 para proteína e 3 para mozzarella.

Tabela 4. Regiões de QTL em que genes candidatos foram associados com produção de leite, gordura, proteína, e mozzarella baseado na abordagem Bayes C.

Chr	QTL (região bp)	SNP	BF	Genes
Gordura				
4	91246634-91446634	AX-85077742	14.25	<i>TMPRSS12, LOC102400376, LOC102399706, LOC102411837, LOC102411498, LOC102407231</i>
6	20037608-20237608	AX-85117685	13.08	<i>PRPF3, MRPS21, CIART, C6H1orf54, APH1A, CA14, ANP32E, PLEKHO1, LOC102390368, VPS45</i>
6	21713652-21913652	AX-85057069	13.13	<i>LOC112585572, GJA5, ACP6, BCL9</i>
9	4693488-4893488	AX-85125980	13.02	-
10	32731586-32931586	AX-85080891	13.08	<i>EYA4</i>
10	76128449-76328449	AX-85110712	14.71	<i>TRDN</i>
22	17428429-17628429	AX-85065144	16.65	<i>TRNAY-AUA, SETBP1</i>
23	11216510-11416510	AX-85126194	16.24	<i>PANK1, KIF20B</i>
24	14823586-15023586	AX-85101478	12.49	<i>CHCHD2, PHKG1, SUMF2, LOC112581831, CCT6A, LOC112581924, LOC112581925, PSPH, TRNAE-CUC, NIPSNAP2, MRPS17, ZNF713, SEPT14, LOC102400066, LOC102400066</i>
Proteína				
4	160795448-160995448	AX-85055817	11.53	<i>BMPR1A, LOC112584859, MMRN2, SNCG, ADIRF, FAM25A, GLUD1</i>
6	57899918-58099918	AX-85124549	13.94	<i>LOC102405956, DDAH1</i>
6	91477870-91677870	AX-85044140	11.42	<i>SSBP3, MRPL37, LOC102389062, CDCP2, TCEANC2, LOC112585863, TMEM59</i>
6	111794832-111994832	AX-85113021	28.87	<i>ZSCAN20, LOC112585732, EFHD1, GIGYF2</i>
9	16110909-16310909	AX-85119900	11.80	<i>MCTP1</i>
Mozzarella				
7	101143327-101343327	AX-85119852	8.13	<i>EIF4E, LOC102410899, LOC102410556</i>

Chr	QTL (região bp)	SNP	BF	Genes
9	8684289-8884289	AX-85075354	9.39	<i>PAM, LOC102405618, LOC102405384</i>
17	63858162-64058162	AX-85124400	9.23	-

Chr: cromossomo. BF: Bayes Factor.

Um total de 68 genes foram reportados nesse estudo para as características de gordura, proteína e mozzarella analisadas nessa pesquisa. Sendo que em nosso estudo o gene *GJA5* foi descrito como associado significativamente com a característica de produção de gordura, em estudo de associação genômica para características de produção e conteúdo do leite em búfalas, Venturini et al. (2014) também identificaram o gene *GJA5* como significativo com produção de gordura, além de também achar associação com proteína e leite.

Uma vez que doenças como a cetose que induzem o catabolismo das gorduras para obter energia, quando apresentadas nas vacas pode afetar o rendimento do leite e seus constituintes, entre eles a gordura, o gene *PSPH* identificado nesse estudo como associado com produção de gordura, foi identificado por Nayeri et al. (2019), como gene candidato funcionalmente importante associado com essa doença.

O gene *SETBP1* que foi associado com produção de gordura nesse estudo, encontrasse relacionado com a proliferação do câncer de mama em ovelhas (Peng et al., 2017), sendo que qualquer alteração na fisiologia da mama reduz a eficiência produtiva, afetando diretamente à produção leite e, por conseguinte a produção de gordura no leite. De acordo com Sutera (2018), em outro estudo de GWAS feito em ovelhas, descreve os genes *TMPRSS12* e *TRNAE-CUC*, como parte da região associada com as características de porcentagem e gordura e proteína em ovelhas da raça Valle de Belice.

Outros estudos de GWAS realizados para identificação de SNPs, regiões genômicas e genes que afetam as características de produção em búfalos, Camargo et al. (2015) e Du et al. (2019) descrevem o gene *TRNAY-AUA* como associado significativamente com a característica de porcentagem de gordura, sendo que em nossa pesquisa o mesmo gene encontrasse associado com a produção de gordura.

Genes que em nosso estudo foram identificados perto dos SNPs associados com produção de proteína, encontrassem já identificados em outras pesquisas. Por exemplo, Raven et al. (2014), usaram informações sobre o desenvolvimento mamário, vias de sinalização e involução da prolactina para identificar genes candidatos em regiões que poderiam estar associadas a características de produção do leite. Assim, o gene *BMPR1A*, foi identificado como envolvido no desenvolvimento mamário e associado com porcentagem de proteína.

Assim mesmo, Breckenridge et al. (2009) realizaram uma pesquisa para identificar a função do gene *DDAH1* nos mamíferos, eles descreveram esse gene como expresso no desenvolvimento embrionário e com alta atividade na síntese do óxido nítrico. Uma vez que o óxido nítrico tem um papel importante como vasodilatador e angiogênico além de ser chave na síntese de proteínas, e na lactogênese no tecido mamário podendo aumentando o fluxo lácteo (Lacasse e Prosser, 2003).

5.4 Single SNP regression

Os SNPs fortemente associados pela abordagem single SNP regression encontrassem descritos na Tabela 5. Foram identificados 4 SNPs para gordura e um para proteína.

Tabela 5. Regiões de QTL em que genes candidatos foram associados com produção de leite, gordura, proteína, e mozzarella baseado na análise de single SNP regression.

Chr	QTL (região bp)	SNP	p value	Genes
Gordura				
12	14380870-14580870	AX-85119414	7.4849	<i>MEMO1</i>
24	14823586-15023586	AX-85101478	6.5704	<i>CHCHD2, PHKG1, SUMF2, LOC112581831, CCT6A, LOC112581924, LOC11258192, PSPH, TRNAE-CUC, NIPSNAP2, MRPS17, ZNF713, SEPT14</i>

24	26804293-27004293	AX-85047063	5.5080	<i>XYLT1</i>
24	27497184-27697184	AX-85101472	5.8025	-
Proteína				
9	40225225-40325225	AX-85126308	4.99	<i>EBF1</i>

Chr: cromossomo.

Um total de 16 genes foram reportados nesse estudo para as características de produção de proteína e produção de mozzarella. Uma pesquisa de associação genômica para características de lactação, produção de leite e idade ao primeiro parto (EPP) feita em gado tailandês, descreve o gene *XYLT1* como associado significativamente com idade ao primeiro parto (Yodklaew et. al., 2017); alguns autores como Pirlo et al. (2000) e Ettema & Santos (2004), tem observado variações positivas na porcentagem de gordura no leite, quando variações na EPP podendo explicar a associação de esse SNP com a produção de gordura em nosso estudo; segundo esses autores, a porcentagem de gordura e proteína total do leite tem aumento em vacas com EPP maior (>25 meses) em comparação das vacas que pariram com idade média (23-25 meses) a baixa (<23 meses).

Assim, nossos resultados têm permitido identificar diferentes regiões que continham genes importantes em diferentes processos biológicos e funções moleculares associadas pelo menos com um QTL, principalmente nos BBU6, BBU9, BBU3, BBU5 encontrassem os genes *CRTC2*, *ILF2*, *RPS27*, *RHOBTB3*, *S100A1*, *S100A13*, *TRNAG-CCC*, com efeito sob as diferentes características de qualidade do leite em búfalas avaliadas nessa pesquisa. Muitas dessas regiões genômicas com efeito para as características em análise correspondem a regiões previamente identificadas como QTLs para produção de leite, produção de gordura, produção de proteína e produção de mozzarella sugerindo que os dados foram bem modelados pelas duas metodologias BayesC e single SNP regression para análises de associação de todo o genoma. No entanto, novos potenciais genes candidatos foram associados com importantes funções biológicas e moleculares, que pode ser importante para compreender os principais mecanismos de ação envolvidos nas características de produção de leite.

6. CONCLUSÃO

O presente estudo, revelou 53 marcadores SNPs em comuns em ambas abordagens (Bayes C, single SNP regression), 30 SNPs para gordura, 18 para proteína e 5 para mozzarella que influenciam nas características de qualidade do leite. Os genes *CRTC2*, *ILF2*, *RPS27*, *RHOBTB3*, *S100A1*, *S100A13*, *TRNAG-CCC* que foram encontrados e associados com essas características produtivas podem ser candidatos para pesquisar mutações causais. Um estudo mais aprofundado com a integração dessas descobertas certamente promoverá uma melhor compreensão da arquitetura genética global das características da produção de leite em gado leiteiro.

7. REFERÊNCIAS

- Altiero V, Moio L and Addeo F. (1989). Previsione della resa in “mozzarella” sulla base del contenuto in grasso e proteine del latte di bufala. **Sci. Tec.** Lattiero-Casearia 40: 425-433.
- Ardlie, K. G., Kruglyak, L., & Seielstad, M. (2002). Patterns of linkage disequilibrium in the human genome. **Nature Reviews Genetics**, 3(4), 299-309.
- Aspilcueta-Borquis, R. R., Di Palo, R., Araujo Neto, F. R., Baldi, F., De Camargo, G. M. F., Albuquerque, L. G. D., & Tonhati, H. (2010a). Genetic parameter estimates for buffalo milk yield, milk quality and mozzarella production and Bayesian inference analysis of their relationships. **Genetics and Molecular Research**, 1636-1644.
- Aspilcueta-Borquis, R. R., Sesana, R. C., Berrocal, M. H. M., Seno, L. D. O., Bignardi, A. B., El Faro, L., ... & Tonhati, H. (2010b). Genetic parameters for milk, fat and protein yields in Murrah buffaloes (*Bubalus bubalis* Artiodactyla, Bovidae). *Genetics and Molecular Biology*, 33(1), 71-77.
- Barros, C. D. C., Aspilcueta-Borquis, R. R., Fraga, A. B., & Tonhati, H. (2016). Genetic parameter estimates for production and reproduction traits in dairy buffaloes. **Revista Caatinga**, 29(1), 216-221.
- Beal, R., Fernandez, A. A. C., Grammatopoulos, D. K., Matter, K., & Balda, M. S. (2020). ARHGEF18/p114RhoGEF coordinates PKA/CREB signaling and actomyosin remodeling to drive trophoblast cell-cell fusion during placenta morphogenesis. **bioRxiv**.
- Breckenridge, R. A., Kelly, P., Nandi, M., Vallance, P. J., Ohun, T. J., & Leiper, J. (2009). A role for Dimethylarginine Dimethylaminohydrolase 1 (DDAH1) in mammalian development. **International Journal of Developmental Biology**, 54(1), 215-220.
- Brito, J.R.F.; Dias, J.C. (1998). A qualidade do leite. Juiz de Fora: **Embrapa/Tortuga**. 98p.
- Bueno, P. R. D. B. U., Rorato, P. R. N., Dürr, J. W. U., & Krug, E. E. B. U. (2004). Valor econômico para componentes do leite no estado do Rio Grande do Sul. **Revista Brasileira de Zootecnia**.
- Cantor, R. M.; Lange, K.; Sinsheimer, J. S. (2010). Prioritizing GWAS results: a review of statistical methods and recommendations for their application. **The American Journal of Human Genetics**, v. 86, n. 1, p. 6–22.
- Chan EKF, Hawken R and Reverter A (2009). The combined effect of SNP-marker and phenotype attributes in genome-wide association studies. **Animal Genetics** 40:149–56. doi: 10.1111/j.1365-2052.2008.01816.
- Costa, A., Schwarzenbacher, H., Mészáros, G., Fuerst-Waltl, B., Fuerst, C., Sölkner, J., & Penasa, M. (2019). On the genomic regions associated with milk lactose in Fleckvieh cattle. **Journal of dairy science**, 102(11), 10088-10099.

- Crisà, A., Ferrè, F., Chillemi, G., & Moioli, B. (2016). RNA-Sequencing for profiling goat milk transcriptome in colostrum and mature milk. **BMC veterinary research**, 12(1), 264.
- Da Costa Barros, C., de Abreu Santos, D. J., Aspilcueta-Borquis, R. R., De Camargo, G. M. F., de Araújo Neto, F. R., & Tonhati, H. (2018). 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**, 85(4), 402-406.
- De Camargo, G. M. F., Aspilcueta-Borquis, R. R., Fortes, M. R. S., Porto-Neto, R., Cardoso, D. F., Santos, D. J. A., & Tonhati, H. (2015). Prospecting major genes in dairy buffaloes. **BMC genomics**, 16(1), 872.
- Dekkers, J. C. M. (2007). Prediction of response to marker-assisted and genomic selection using selection index theory. **Journal of Animal Breeding and Genetics** 124(6):331-341.
- De Resende, M. D. V., Lopes, P. S., da Silva, R. L., & Pires, I. E. (2008). Seleção genômica ampla (GWS) e maximização da eficiência do melhoramento genético. **Pesquisa florestal brasileira**, (56), 63.
- Deng, T., Liang, A., Liu, J., Hua, G., Ye, T., Liu, S., & Salzano, A. (2019). 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**, 9, 688.
- Dong, L., Han, Z., Fang, M., Xiao, S., & Wang, Z. (2019). Genome-wide association study identifies loci for body shape in the large yellow croaker (*Larimichthys crocea*). **Aquaculture and Fisheries**, 4(1), 3-8.
- Du, C., Deng, T., Zhou, Y., Ye, T., Zhou, Z., Zhang, S., ... & Yang, L. (2019). Systematic analyses for candidate genes of milk production traits in water buffalo (*Bubalus Bubalis*). **Animal genetics**, 50(3), 207-216.
- Duarte, J.M.C.; Tonhati, H.; Ceron-Muñoz, M.F. et al., (2001). Efeitos ambientais sobre a produção no dia do controle e características físico-químicas do leite em um rebanho bubalino no estado de São Paulo, Brasil. **Revista do Instituto de Laticínios Cândido Tostes**, v.56, n.322, p.17-20.
- Dubey, P.C., Suman, C.L., Sanyal, M.K. (1996). Factors affecting composition of milk of buffaloes. **Indian Journal of Animal Science**. 67(9):802-804.
- El-Halawany, N., Abdel-Shafy, H., Abd-El-Monsif, A. S., Abdel-Latif, M. A., Al-Tohamy, A. F., & El-Moneim, O. M. A. (2017). Genome-wide association study for milk production in Egyptian buffalo. **Livestock Science**, 198, 10-16.
- Ettema, J. F., & Santos, J. E. P. (2004). Impact of age at calving on lactation, reproduction, health, and income in first-parity Holsteins on commercial farms. **Journal of dairy science**, 87(8), 2730-2742.

FAOSTAT. Food and Agriculture Organization of the United Nations. **Live animals**. (2018). Disponível em: <http://www.fao.org/faostat/es/#data/QA/visualize>

Fernández, M. H., & Vrba, E. S. (2005). A complete estimate of the phylogenetic relationships in Ruminantia: a dated species-level supertree of the extant ruminants. **Biological reviews**, 80(2), 269-302.

Garrick, D. J., & Fernando, R. L. (2013). Implementing a QTL detection study (GWAS) using genomic prediction methodology. In **Genome-wide association studies and genomic prediction** (pp. 275-298). Humana Press, Totowa, NJ.

Garzón Audor, A. M., & Oliver Espinosa, O. J. (2018). Epidemiología de la cetosis en bovinos: una revisión. **CES Medicina Veterinaria y Zootecnia**, 13(1), 42-61.

Giovannone, B., Tsiaras, W. G., de la Monte, S., Klysik, J., Lautier, C., Karashchuk, G., ... & Smith, R. J. (2009). GIGYF2 gene disruption in mice results in neurodegeneration and altered insulin-like growth factor signaling. **Human molecular genetics**, 18(23), 4629-4639.

Gondro, C., Van der Werf, J., & Hayes, B. (Eds.). (2013). **Genome-wide association studies and genomic prediction** (p. 566). Totowa, NJ, USA: Humana Press.

Habier, D., R. L. Fernando, AND J. C. M. Dekkers. (2007). The impact of genetic relationship information on genome-assisted breeding values. **Genetics** 177(4):2389-2397.

Habier D, Fernando RL, Kizilkaya K, Garrick DJ. (2011). Extension of the Bayesian alphabet for genomic selection. **BMC Bioinformatics**, v. 12, n. 186.

Hayrabedian, S., Kyurkchiev, S., & Kehayov, I. (2005). Endoglin (cd105) and S100A13 as markers of active angiogenesis in endometriosis. **Reprod Biol**, 5(1), 51-67.

Hill, W. G. (2014). Applications of Population Genetics to Animal Breeding, from Wright, Fisher and Lush to Genomic Prediction. **Genetics** 196(1):1-16. doi: 10.1534/genetics.112.147850.

Hill, M. J., Suzuki, S., Segars, J. H., & Kino, T. (2016). CRT2 is a coactivator of GR and couples GR and CREB in the regulation of hepatic gluconeogenesis. **Molecular endocrinology**, 30(1), 104-117.

Hirschhorn, J. N. Daly, M. J. (2005). Genome-wide association studies for common diseases and complex traits. v.6. **Nature reviews/genetics**.

IBGE. Instituto Brasileiro de Geografia e Estatística, 2017. Disponível em: https://biblioteca.ibge.gov.br/visualizacao/periodicos/3093/agro_2017_resultados_preliminares.pdf

Kambe, T., Geiser, J., Lahner, B., Salt, D. E., & Andrews, G. K. (2008). Slc39a1 to 3 (subfamily II) Zip genes in mice have unique cell-specific functions during adaptation to zinc deficiency. **American Journal of Physiology-Regulatory, Integrative and Comparative Physiology**, 294(5), R1474-R1481.

Kizilkaya, K., Fernando, R. L., & Garrick, D. J. (2010). Genomic prediction of simulated multibreed and purebred performance using observed fifty thousand single nucleotide polymorphism genotypes. **Journal of animal science**, 88(2), 544-551.

Klein RJ, Zeiss C, Chew EY, Tsai J-Y, Sackler RS, Haynes C, Henning AK, SanGiovanni JP, Mane SM, Mayne ST et al. (2005) Complement factor H polymorphism in age-related macular degeneration. **Science** 308:385–9. doi: 10.1126/science.1109557.

Kumar, S., Rao, N., & Ge, R. (2012). Emerging roles of ADAMTSs in angiogenesis and cancer. **Cancers**, 4(4), 1252-1299.

Lacasse, P., & Prosser, C. G. (2003). Mammary blood flow does not limit milk yield in lactating goats. **Journal of dairy science**, 86(6), 2094-2097.

Lee, S. H., Van Der Werf, J. H., Hayes, B. J., Goddard, M. E., & Visscher, P. M. (2008). Predicting unobserved phenotypes for complex traits from whole-genome SNP data. **PLoS genetics**, 4(10).

Lee, E. J., Oh, B., Lee, J. Y., Kimm, K., Lee, S. H., & Baek, K. H. (2008). A novel single nucleotide polymorphism of INSR gene for polycystic ovary syndrome. **Fertility and sterility**, 89(5), 1213-1220.

Li, P., Zhou, C., Li, X., Yu, M., Li, M., & Gao, X. (2019). CRT2 is a key mediator of amino acid-induced milk fat synthesis in mammary epithelial cells. **Journal of agricultural and food chemistry**, 67(37), 10513-10520.

Liu, J. J., Liang, A. X., Campanile, G., Plastow, G., Zhang, C., Wang, Z., & Yang, L. G. (2018). Genome-wide association studies to identify quantitative trait loci affecting milk production traits in water buffalo. **Journal of dairy science**, 101(1), 433-444.

Lu D, Miller S, Sargolzaei M, Kelly M, Vander Voort G, Caldwell T, Wang Z and Moore S (2013a) Genome-wide association analyses for growth and feed efficiency traits in beef cattle. **Journal Animal Science** 91:3612–3633. doi: 10.2527/jas2012-5716

McCarthy, M. I., Abecasis, G. R., Cardon, L. R., Goldstein, D. B., Little, J., Ioannidis, J. P., & Hirschhorn, J. N. (2008). Genome-wide association studies for complex traits: consensus, uncertainty and challenges. **Nature reviews genetics**, 9(5), 356-369.

Madalena, F.E. (1999). Consequências econômicas da seleção para gordura e proteína do leite no Brasil. **Revista da Sociedade Brasileira de Zootecnia**, v.29, n.3, p.685- 692.

Malavolta, F.C. (2006). **Métodos indiretos de avaliação da saúde da glândula mamária nas espécies bovina e bubalina**. Dissertação (Mestrado). 2006. Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista, “Júlio de Mesquita Filho”, Jaboticabal.

Matukumalli, L. K., Lawley, C. T., Schnabel, R. D., Taylor, J. F., Allan, M. F., Heaton, M. P., & Van Tassell, C. P. (2009). Development and characterization of a high density SNP genotyping assay for cattle. **PloS one**, 4(4).

- Meuwissen, T.H.E.; Hayes, B.J.; Goddard, M.E. (2001). Prediction of total genetic value using genome-wide dense marker maps. **Genetics**. v. 157, p.1819–1829.
- Michelizzi, V. N., Dodson, M. V., Pan, Z., Amaral, M. E. J., Michal, J. J., McLean, D. J., & Jiang, Z. (2010). Water buffalo genome science comes of age. **International journal of biological sciences**, 6(4), 333.
- Nayeri, S., Schenkel, F., Fleming, A., Kroezen, V., Sargolzaei, M., Baes, C., ... & Miglior, F. (2019). Genome-wide association analysis for β -hydroxybutyrate concentration in Milk in Holstein dairy cattle. **BMC genetics**, 20(1), 58.
- Nicolazzi EL, Biffani S, Biscarini F, Orozco Ter Wengel P, Caprera A, Nazzicari N and Stella A. (2015). Software solutions for the livestock genomics SNP array revolution. **Animal Genetic** 46:343–353. doi: 10.1111/age.12295.
- Patel, R.S., Mistry, V.V. (1997). Physicochemical and structural properties of ultrafiltered buffalo milk and milk powder. **Journal Dairy Science**. 80:812-817.
- Panigrahi, M., Kumar, H., Sah, V., Verma, A. D., Bhushan, B., & Parida, S. (2020). Transcriptome profiling of buffalo endometrium reveals molecular signature distinct to early pregnancy. **Gene**, 144614.
- Peng, W. F., Xu, S. S., Ren, X., Lv, F. H., Xie, X. L., Zhao, Y. X., ... & Shen, M. (2017). A genome-wide association study reveals candidate genes for the supernumerary nipple phenotype in sheep (*Ovis aries*). **Animal genetics**, 48(5), 570-579.
- Pirlo, G., Miglior, F., & Speroni, M. (2000). Effect of age at first calving on production traits and on difference between milk yield returns and rearing costs in Italian Holsteins. **Journal of dairy science**, 83(3), 603-608.
- Poretsky, L. (1999). The insulin-related ovarian regulatory system in health and disease. **Endocr Rev**, 20, 535-582.
- Pryce JE, Bolormaa S, Chamberlain AJ, Bowman PJ, Savin K, Goddard ME and Hayes BJ (2010) A validated genome-wide association study in 2 dairy cattle breeds for milk production and fertility traits using variable length haplotypes. **J Dairy Sci** 93:3331–45. doi: 10.3168/jds.2009-2893
- Purfield, D. C., Bradley, D. G., Evans, R. D., Kearney, F. J., & Berry, D. P. (2015). Genome-wide association study for calving performance using high-density genotypes in dairy and beef cattle. **Genetics Selection Evolution**, 47(1), 47.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., & Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. **The American journal of human genetics**, 81(3), 559-575.
- Raven, L. A., Cocks, B. G., Goddard, M. E., Pryce, J. E., & Hayes, B. J. (2014). Genetic variants in mammary development, prolactin signalling and involution pathways explain considerable variation in bovine milk production and milk composition. **Genetics Selection Evolution**, 46(1), 29.

Reynolds, C. K., Harmon, D. L., & Cecava, M. J. (1994). Absorption and delivery of nutrients for milk protein synthesis by portal-drained viscera. **Journal of Dairy Science**, 77(9), 2787-2808.

Richardson, I. W., Berry, D. P., Wiencko, H. L., Higgins, I. M., More, S. J., McClure, J., ... & Bradley, D. G. (2016). A genome-wide association study for genetic susceptibility to Mycobacterium bovis infection in dairy cattle identifies a susceptibility QTL on chromosome 23. **Genetics Selection Evolution**, 48(1), 19.

Rivas, P. C., Rodríguez-Márquez, J. M., & Hernández, A. 2006. Número de vasos sanguíneos y expresión del factor de crecimiento del endotelio vascular y de la óxido nítrico sintasa endotelial e inducida en la alantoides ovina. **Revista Científica**, 16(4), 315-324.

Seno, L. O., Cardoso, V. L., El Faro, L., Sesana, R. C., Aspilcueta-Borquis, R. R., De Camargo, G. M. F., & Tonhati, H. (2010). Genetic parameters for milk yield, age at first calving and interval between first and second calving in milk Murrah buffaloes. **Livestock Research for Rural Development**, 22(2), 1-8.

Schaeffer, L. R. (2006). Strategy for applying genome-wide selection in dairy cattle. **Journal of Animal Breeding and Genetics**. v.123, p.218-223.

Sharma, A., Aggarwal, J., Sodhi, M., Kishore, A., Mishra, B. P., Mohanty, A. K., & Mukesh, M. (2014). Stage specific expression of ATP-binding cassette and solute carrier superfamily of transporter genes in mammary Gland of riverine buffalo (Bubalus bubalis). **Animal Biotechnology**, 25(3), 200-209.

Sharma A, Lee JS, Dang CG, Sudrajad P, Kim HC, Yeon SH, Kang HS and Lee S-H (2015) Stories and Challenges of Genome Wide Association Studies in Livestock - A Review. **Asian-Australasian Journal Animal Science** 28:1371–9. doi: 10.5713/ajas.14.0715

Sorensen, D., & Gianola, D. (2007). Likelihood, Bayesian, and MCMC methods in quantitative genetics. **Springer Science & Business Media**.

Sutera, A. (2018). Comparison of Genome Wide Association Studies for milk production traits in Valle del Belice dairy sheep.

Tonhati, H., Albuquerque, L.G., Oliveira, J.F.S. (1996). Melhoramento genético em bubalinos. Programa Vale do Ribeira, SP. In: **Simpósio Nacional De Melhoramento Genético Animal**, 1, Ribeirão Preto, SP.

Tonhati, H.; Muñoz, M.F.C.; Oliveira, J.A.; Duarte, J.M.C.; Furtado, T.P.; Tseimazides, S.P. (2000). Parâmetros genéticos para produção de leite, gordura e proteína em bubalinos. **Revista Brasileira de Zootecnia**. Viçosa. v. 29, n. 6, p. 2051-2056.

Tonhati, H., Mendoza-Sánchez, G., Sesana, R. C., Aspilcueta Borquis, R. R., & Galvão De Albuquerque, L. (2008). Programa de mejoramiento genético de búfalos. **Revista Argentina de Producción Animal**, 28(1), 53-67.

- Tonhati, H.; Camargo, G. M. F.; Aspilcueta-Borquis, R. R.; Hutardo-Lugo, N. A. (2012). Melhoramento Genético de bubalinos. In: Pereira, J. C. C.. **Melhoramento genético aplicado à produção animal**. 6 ed. Belo Horizonte: FEPMVZ Editora. cap 19, p.468-483.
- Van Tassell, C. P., Smith, T. P., Matukumalli, L. K., Taylor, J. F., Schnabel, R. D., Lawley, C. T., & Sonstegard, T. S. (2008). SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. **Nature methods**, 5(3), 247.
- Valencia Hernández, A. F. (2016). **Efecto de la aplicación de suplemento mineral inyectable sobre la condición corporal, dinámica folicular y tasa de concepción en novillas cebú comercial en el departamento del Caquetá**. Dissertação (mestrado). Facultad de Ciencias Veterinarias. Universidad de La Salle
- Valle, J.L.E. (1990). Características e usos do leite de bubalinos. In: **Reunião Anual Da Sociedade Brasileira de Zootecnia**, 27, Campinas, SP. Anais... Campinas: SBZ. 1990. p.739-743.
- Venturini, G. C., Cardoso, D. F., Baldi, F., Freitas, A. C., Aspilcueta-Borquis, R. R., Santos, D. J. A., & Tonhati, H. (2014). Association between single-nucleotide polymorphisms and milk production traits in buffalo. **Genetics and Molecular Research**, 13(4), 10256-68.
- Visscher, P.M.; Brown, M.A.; Mccarthy, M.I.; Yang, J. (2012). Five Years of GWAS Discovery. **The American Journal of Human Genetics**, v.90, p.7–24.
- Vivas, W. A. C. (2017). **Factores limitantes de la Gluconeogenesis en el periodo de transición de la vaca lechera**. Documentos de Trabajo ECAPMA, (2).
- Wang, H., Misztal, I., Aguilar, I., Legarra, A., & Muir, W. M. (2012). Genome-wide association mapping including phenotypes from relatives without genotypes. **Genetics Research**, 94(2), 73-83.
- Wang, H., Misztal, I., Aguilar, I., Legarra, A., Fernando, R. L., Vitezica, Z., ... & Muir, W. M. (2014). 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**, 5, 134.
- Warriach, H. M., Mcgill, D. M., Bush, R. D., Wynn, P. C., & Chohan, K. R. (2015). A review of recent developments in buffalo reproduction—a review. **Asian-Australasian journal of animal sciences**, 28(3), 451.
- Wise, A., Tenezaca, L., Fernandez, R. W., Schatoff, E., Flores, J., Ueda, A., & Venkatesh, T. (2015). Drosophila mutants of the autism candidate gene neurobeachin (rugose) exhibit neuro-developmental disorders, aberrant synaptic properties, altered locomotion, and impaired adult social behavior and activity patterns. **Journal of neurogenetics**, 29(2-3), 135-143.

Xiao, C. T., & Cant, J. P. (2005). Relationship between glucose transport and metabolism in isolated bovine mammary epithelial cells. **Journal of dairy science**, 88(8), 2794-2805.

Xu, L., Shi, L., Liu, L., Liang, R., Li, Q., Li, J., ... & Sun, D. (2019). Analysis of Liver Proteome and Identification of Critical Proteins Affecting Milk Fat, Protein, and Lactose Metabolism in Dairy Cattle with iTRAQ. **Proteomics**, 19(12), 1800387.

Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2013). Genome-wide complex trait analysis (GCTA): methods, data analyses, and interpretations. In *Genome-wide association studies and genomic prediction* (pp. 215-236). **Humana Press**, Totowa, NJ.

Yodklaew, P., Koonawootrittriron, S., Elzo, M. A., Suwanasopee, T., & Laodim, T. (2017). Genome-wide association study for lactation characteristics, milk yield and age at first calving in a Thai multibreed dairy cattle population. **Agriculture and Natural Resources**, 51(3), 223-230.

Zaidi, D., James, K. A., & Wagner, G. F. (2006). Passive immunization of lactating mice with stanniocalcin-1 antiserum reduces mammary gland development, milk fat content, and postnatal pup growth. **American Journal of Physiology-Endocrinology and Metabolism**, 291(5), E974-E981.

Zhang H, Wang Z, Wang S and Li H (2012) Progress of genome wide association study in domestic animals. **Journal Animal Science Biotechnology** 3:26. doi: 10.1186/2049-1891-3-26

Zhang Xingue. (2015). **Efficiency of single-step GBLUP in genomic evaluation and GWAS in broiler chickens**. Doctoral dissertation, university of Georgia.

Zimin, A. V.; Delcher, A.L.; Florea, L.; Kelley, D.R.; Schatz, M.C.; Puiu, D.; Hanrahan, F.; Pertea, G.; Van Tassell, C.P.; Sonstegard, T.S. (2009). A whole genome assembly of the domestic cow, *Bos taurus*. **Genome Biology**, v.10, n.4.