

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

**ESTUDO DE SELEÇÃO GENÔMICA PARA
CARACTERÍSTICAS DE PRODUÇÃO E QUALIDADE DO
LEITE DE BÚFALAS**

**Camila da Costa Barros
Zootecnista**

2017

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

**ESTUDO DE SELEÇÃO GENÔMICA PARA
CARACTERÍSTICAS DE PRODUÇÃO E QUALIDADE DO
LEITE DE BÚFALAS**

Camila da Costa Barros

Orientador: Prof. Dr. Humberto Tonhati

Coorientador: Dr. Rusbel Raul Aspilcueta-Borquis

Coorientador: Dr. Daniel Jordan de Abreu Santos

**Tese 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.**

2017

B277e Barros, Camila da Costa
Estudo de seleção genômica para características de produção e
qualidade do leite de búfalas / Camila da Costa Barros. --
Jaboticabal, 2017
viii, 38 p. ; 29 cm

Tese (doutorado) - Universidade Estadual Paulista, Faculdade de
Ciências Agrárias e Veterinárias, 2017

Orientador: Humberto Tonhati

Coorientador: Rusbel Raul Aspilcueta-Borquis, Daniel Jordan de
Abreu Santos

Banca examinadora: Roberto Carvalheiro, Francisco Ribeiro de
Araujo Neto, Leonardo de Oliveira Seno, Guilherme Costa Venturini

Bibliografia

1. Búfalos. 2. GWAS. 3. Predição genômica. 4. SNP. I. Título. II.
Jaboticabal-Faculdade de Ciências Agrárias e Veterinárias.

CDU 636.082:636.293.2

CERTIFICADO DE APROVAÇÃO

TÍTULO DA TESE: ESTUDO DE SELEÇÃO GENÔMICA PARA CARACTERÍSTICAS DE PRODUÇÃO E QUALIDADE DO LEITE DE BÚFALAS

AUTORA: CAMILA DA COSTA BARROS

ORIENTADOR: HUMBERTO TONHATI

COORIENTADOR: RUSBEL RAUL ASPILCUETA BORQUIS

COORIENTADOR: DANIEL JORDAN DE ABREU SANTOS

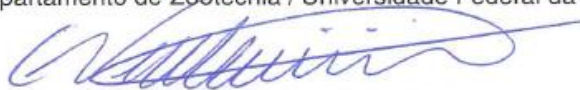
Aprovada como parte das exigências para obtenção do Título de Doutora em GENÉTICA E MELHORAMENTO ANIMAL, pela Comissão Examinadora:


Prof. Dr. RUSBEL RAUL ASPILCUETA BORQUIS
Departamento de Zootecnia / Universidade Federal da Grande Dourados - Dourados/MS


Prof. Dr. ROBERTO CARVALHEIRO
Departamento de Zootecnia / FCAV / UNESP - Jaboticabal


Prof. Dr. FRANCISCO RIBEIRO DE ARAUJO NETO
Departamento de Zootecnia / Instituto Federal Goiano / Rio Verde/GO


Prof. Dr. LEONARDO DE OLIVEIRA SENO - VIDEOCONFERÊNCIA
Departamento de Zootecnia / Universidade Federal da Grande Dourados - Dourados/MS


Prof. Dr. GUILHERME COSTA VENTURINI
Departamento de Zootecnia / FMVZ / UNESP - Botucatu

Jaboticabal, 21 de julho de 2017

DADOS CURRICULARES DO AUTOR

CAMILA DA COSTA BARROS nasceu em 17 de março de 1988 na cidade de Maceió – AL, filha de Robson Araújo Barros e Sandra da Costa Barros. Iniciou o curso de Zootecnia no ano de 2006 no Centro de Ciências Agrárias – CECA da Universidade Federal de Alagoas – UFAL, obtendo o título de Zootecnista em 13 de maio de 2011. Em março do mesmo ano, ingressou no Programa de Pós-Graduação em Zootecnia na Universidade Federal de Alagoas como bolsista da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES, obtendo o título de Mestra em Zootecnia em 4 de junho de 2013, sob orientação da Professora Dra. Angelina Bossi Fraga. Em agosto de 2013, ingressou no Programa de Pós-Graduação em Genética e Melhoramento Animal na Faculdade de Ciências Agrárias e Veterinárias, UNESP, Campus de Jaboticabal – SP, inicialmente como bolsista da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES, e em seguida bolsista da Fundação de Amparo à Pesquisa do Estado de São Paulo – FAPESP, sob a orientação do Professor Dr. Humberto Tonhati. Durante o período de janeiro a setembro de 2016, realizou estágio sanduíche na University of Wisconsin – EUA como bolsista FAPESP, sob a supervisão do Professor Dr. Guilherme Jordão de Magalhães Rosa.

Veja

*Não diga que a canção está perdida
Tenha fé em Deus, tenha fé na vida
Tente outra vez*

Beba

*Pois a água viva ainda está na fonte
Você tem dois pés para cruzar a
ponte
Nada acabou, não não não não*

Tente

*Levante sua mão sedenta e recomece
a andar
Não pense que a cabeça aguenta se
você parar,
não não não não
Há uma voz que canta,
uma voz que dança,
uma voz que gira
Bailando no ar*

Queira

*Basta ser sincero e desejar profundo
Você será capaz de sacudir o mundo,
vá
Tente outra vez*

Tente

*E não diga que a vitória está
perdida
Se é de batalhas que se vive a vida
Tente outra vez*

*Raul Seixas/ Paulo Coelho/ Marcelo
Motta*

Aos meus pais, Robson e Sandra, que me proporcionaram condições para que eu pudesse chegar até aqui.

DEDICO

À minha mãe, Sandra, exemplo de determinação, positividade e resiliência.

OFEREÇO

AGRADECIMENTOS

A Deus, por me dar forças para seguir essa caminhada.

Ao meu orientador Prof. Dr. Humberto Tonhati, pela oportunidade e confiança.

Aos meus coorientadores, Rusbel R. Aspilcueta-Borquis e Dr. Daniel J. A. Santos, pela coorientação e dedicação, os quais foram fundamentais para o andamento deste trabalho.

Ao Prof. Dr. Guilherme Rosa, pela oportunidade de participar do seu grupo nos EUA, apoio, auxílio e valiosas sugestões neste trabalho.

Ao Dr. Fernando Brito, pelos ensinamentos e sugestões que foram fundamentais para o andamento desta pesquisa, paciência e amizade.

A Capes e a FAPESP (Processo n°: 2013/24427-3) pela concessão da bolsa de estudos no Brasil e no exterior (Processo n°: 2015/18614-0).

A FCAV/UNESP Jaboticabal e seus funcionários.

Aos membros da banca de qualificação, Prof. Dr. Henrique Nunes de Oliveira, Prof. Dr. Roberto Carneiro, Dr. Rodrigo P. Savegnago e Dra. Daiane C. B. Scalet, pelas valiosas sugestões.

Aos membros da banca de defesa, Prof. Dr. Roberto Carneiro, Prof. Dr. Francisco Ribeiro de Araújo Neto, Prof. Dr. Guilherme Costa Venturini e Prof. Dr. Leonardo de Oliveira Seno, pela disponibilidade e relevantes sugestões.

A todos os amigos da UNESP Jaboticabal que fiz durante a minha passagem nesta instituição.

Aos amigos de Madison que foram essenciais para a minha adaptação nos EUA e que me proporcionaram também momentos de alegria e descontração.

Aos meus pais, Robson e Sandra, por todo o apoio e incentivos para buscar o aprendizado e poder me aprofundar nos conhecimentos acerca da profissão que escolhi ter.

À minhas irmãs, Patrícia e Rafaela, meu sobrinho Emanuel e minha avó Mariinha, por compreenderem a ausência nos momentos importantes e únicos da nossa família.

Ao meu namorado, Ivan, não só pelo apoio e incentivo a todo momento, mas também pela compreensão no momento de ausência, conversas, desabafos e conselhos que foram fundamentais para continuar nessa caminhada.

À família Moura Lapera, que me recebeu com tanto carinho e me acolheu nesta cidade. Diante da ausência da minha verdadeira família por motivo de distância, senti-me abraçada e membro da família de vocês também.

Ao meu companheirinho de quatro patas, Bento, pelo amor incondicional que apenas um cão pode oferecer a um ser humano.

Muito obrigada!

SUMÁRIO

RESUMO	vi
CAPÍTULO 1 – Considerações Gerais	1
1. INTRODUÇÃO	1
1.1 Revisão de literatura	2
1.1.1 Seleção genômica	5
1.1.2 Associação genômica ampla	6
1.3 Objetivos	8
2. REFERÊNCIAS BIBLIOGRÁFICAS	9
CAPÍTULO 2 – Métodos Bayesianos aplicados à predição genômica de características de produção e qualidade do leite de búfalas	11
2.1 Introdução	12
2.2 Material e métodos	13
2.3 Resultados e discussão	18
2.4 Conclusão	21
2.5 Referências bibliográficas	21
CAPÍTULO 3 – Estudo de associação genômica ampla para características de produção e qualidade do leite de búfalas	23
3.1 Introdução	24
3.2 Material e métodos	26
3.3 Resultados e discussão	30
3.4 Conclusão	34
3.5 Referências bibliográficas	34
CAPÍTULO 4 - Considerações Finais	37

ESTUDO DE SELEÇÃO GENÔMICA PARA CARACTERÍSTICAS DE PRODUÇÃO E QUALIDADE DO LEITE DE BÚFALAS

RESUMO – Objetivou-se com o presente trabalho comparar diferentes métodos Bayesianos de predição genômica para as características de produção de leite (PL) e as porcentagens de gordura (%G) e proteína (%P) no leite de búfalas e, realizar um estudo de associação genômica ampla, a fim de identificar regiões cromossômicas e genes possivelmente relacionados às mesmas, utilizando informações de indivíduos genotipados e não genotipados. O número de animais com fenótipo foi 3.355, o arquivo de pedigree continha 15.495 animais, dos quais 322 foram genotipados com o 90 K Axiom® Buffalo Genotyping array. Os seguintes critérios de controle de qualidade dos SNPs foram utilizados: MAF < 0,05; Call Rate < 0,95 e Equilíbrio de Hardy-Weinberg p -value < 10^{-6} . Em relação à amostra foi considerado call rate < 0,90. Para as predições genômicas, os seguintes modelos Bayesianos foram utilizados: Bayes A (BA), Bayes B (BB), Bayes C (BC) e Bayes LASSO (BL). O fenótipo corrigido para os efeitos fixos (Y^*) foi utilizado como variável resposta nas análises genômicas. A habilidade de predição dos diferentes modelos foi avaliada usando o método leave-one-out de validação cruzada. As acurácias de predição foram calculadas através da correlação de Pearson entre o valor genético genômico estimado (GEBV) e a variável resposta (Y^*) para cada modelo e característica avaliados. Em relação ao estudo de associação genômica ampla, um processo iterativo foi realizado para calcular os pesos dos marcadores em função do quadrado dos efeitos dos SNPs e das frequências alélicas (ssGWAS). Em geral, todos os modelos Bayesianos demonstraram semelhantes acurácias de predição, variando de 0,41 a 0,42, 0,38 a 0,39 e 0,39 a 0,40 para a PL, %G e %P, respectivamente. Portanto, os métodos BA, BB, BC e BL podem ser utilizados nas predições dos efeitos dos SNPs, obtendo-se, praticamente, as mesmas acurácias de predição. Os dez SNPs de maiores efeitos para a PL, %G e %P explicaram 7,48, 9,94 e 6,56% da proporção da variância genética, respectivamente. Os resultados do ssGWAS revelaram regiões cromossômicas e genes que podem estar relacionados com as características

analisadas. Tais regiões e genes identificados poderão contribuir para o melhor entendimento sobre a influência dos mesmos nas características produção de leite e as porcentagens de gordura e proteína no leite de búfalas.

Palavras-chave: búfalos, GWAS, predição genômica, SNP

GENOMIC SELECTION STUDIES FOR PRODUCTION AND QUALITY TRAITS OF MILK BUFFALOES

ABSTRACT – The aim of this study was to compare different Bayesian methods of genomic prediction for milk yield (MY), fat (%F) and protein (%P) percentages in dairy buffaloes in Brazil, and to perform a genome-wide association study for the purpose of identify chromosomal regions and genes possibly related to the these traits, using information from genotyped and non-genotyped individuals. The number of animals with phenotype was 3,355, the pedigree file contained 15,495 animals, of which 322 were genotyped. The animals were genotyped using a 90K SNP panel (Axiom® Buffalo Genotyping Array). The following criteria for quality control of SNPs were used: MAF < 0.05, Call Rate < 0.95 and Hardy-Weinberg Equilibrium p-value < 10^{-6} . In relation to the sample, a Call Rate < 0.90 was used. Four methods for genomic prediction were used: Bayes A (BA), Bayes B (BB), Bayes C (BC) and Bayes LASSO (BL). Phenotypes for the fixed effects (Y^*) were used as response variables. The predictive ability of the different models was evaluated using a leave-one-out cross-validation approach. The prediction accuracy was calculated by Pearson's correlation between estimated genomic genetic value (GEBV) and response variable (Y^*) for each model. In relation to genome-wide association studies, an iterative process was performed to derive SNP weights as function of squares of SNP effects and allele frequencies (ssGWAS). In general, all Bayesian models showed similar prediction accuracy, ranging from 0.41 to 0.42, 0.38 to 0.39 and 0,39 to 0,40 for MY, %F and %P, respectively. Therefore, the methods BA, BB, BC and BL can be used in the predictions of the effects of SNPs, obtaining, practically, the same prediction accuracy. The proportions of variance explained by the top 10 SNPs for MY, %F and %P were: 7.48, 9.94 and 6.56%, respectively. The results of ssGWAS revealed chromosomal regions and genes that may be related with the analyzed traits. These regions and genes may contribute to a better understanding of their influence on milk yield and fat and protein percentages in buffalo milk.

Key-words: buffaloes, genomic prediction, GWAS, SNP

CAPÍTULO 1 – Considerações Gerais

1. INTRODUÇÃO

Os búfalos (*Bubalus bubalis*) são animais domésticos oriundos do continente asiático e não possuem relação direta com as espécies agressivas e selvagens do bisão (*Bos bison bison*), também conhecido como búfalo americano, ou do búfalo africano (*Syncerus caffer caffer*). Possuem temperamento dócil, quando manejados diariamente, se adaptam muito bem às regiões úmidas e têm grande importância socioeconômica, principalmente em países em desenvolvimento, devido à sua aptidão tanto para a produção de carne e leite, quanto para o trabalho.

Na Itália, país onde originalmente a *mozzarella* de búfala surgiu, houve uma intensificação da seleção genética desses animais voltado a produção de leite. No Brasil, a produção de leite de búfalas ganhou destaque nas últimas décadas devido à sua composição, uma vez que é possível obter maiores rendimentos dos derivados lácteos, e por isso, a seleção por meio da genética quantitativa vem sendo realizada.

Dentre as dificuldades de se avaliar animais com superioridade genética para a produção de leite e seus constituintes, estão o fato de que tais características são mensuradas apenas em animais do sexo feminino e de expressão tardia. Tornando necessária a implementação de testes de progênie para avaliação de touros, fazendo com que o ganho genético seja mais lento devido aos longos intervalos de geração. Contudo, os recentes adventos das tecnologias genômicas tornou possível incluir as informações dos genótipos dos animais no processo de melhoramento genético das características de produção e qualidade do leite de búfalas.

A seleção genômica de bovinos leiteiros já é uma realidade em países como Estados Unidos, Nova Zelândia, Austrália e Holanda. Tal processo se baseia em duas etapas fundamentais. A primeira, consiste na estimação dos efeitos dos marcadores através de uma população de referência em que todos os animais genotipados possuem fenótipos ou pseudo fenótipos. Deste modo, os efeitos de todos os loci que contribuem para a variação da característica são capturados. Nas gerações seguintes, com a estimação do efeito dos SNPs, animais jovens podem ser selecionados apenas

com base nos seus valores genômicos (*Genomic Estimated Breeding Value* – GEBVs) (HAYES et al., 2009).

A fim de minimizar as dificuldades de inserir a informação genômica na avaliação genética, a criação de modelos que utilizem este tipo de informação para melhorar a predição do valor genético dos animais, correspondentes às características de importância econômica, vem sendo realizada pelos pesquisadores da área. Com isso, o estudo e a comparação destas diferentes metodologias devem ser realizados com o intuito de obter os métodos adequados para as diferentes características de interesse econômico.

Além disso, devido à disponibilidade da sequência do genoma de várias espécies animais e ao grande número de SNPs descobertos, estudos de associação genômica ampla (GWAS – *Genome Wide Association Studies*) têm contribuído para a identificação de regiões cromossômicas ou genes relacionados a algumas características de importância econômica em bovinos, como por exemplo, produção e qualidade do leite, fertilidade, crescimento e características de carcaça (ZHANG et al., 2012).

1.1 Revisão de literatura

Os búfalos (*Bubalus bubalis*) são originários da Ásia, mais precisamente da Índia. Sua domesticação ocorreu há aproximadamente 5.000 anos, tempo relativamente recente em relação à espécie bovina. Por terem se tornado importantes fontes de carne, leite e trabalho, esses animais foram levados à África, à Europa e, somente depois introduzidos no continente Americano (BORGHESE, 2005).

No Brasil, sua entrada se deu pela Ilha de Marajó, no Estado do Pará, em meados de 1895. Entretanto, somente em 1962 com a importação de vinte reprodutores selecionados puros das raças Jafarabadi e Murrah, houve grandes estímulos aos criadores e selecionadores devido à contribuição desses animais para o melhoramento genético do rebanho nacional. Este fato representou um marco na exploração e expansão dos búfalos no país (SANTIAGO, 2000). Quatro são as raças oficialmente reconhecidas pela Associação Brasileira de Criadores de Búfalos (Murrah, Mediterrâneo, Jafarabadi e Carabao).

A adaptabilidade desses animais aos mais variados ambientes, sua elevada fertilidade e longevidade produtiva, permitiram que o rebanho bubalino brasileiro tivesse uma evolução significativa na década de 80. No início, sua exploração destinava-se, em sua maioria, à produção de carne, porém, a partir dos anos 90, verificou-se um aumento no interesse em sua exploração leiteira e pela produção dos derivados de leite de búfalas devido a um maior valor agregado, permitindo remunerar a matéria-prima a preços cerca de duas vezes maiores que aqueles pagos pelo leite bovino (BERNARDES, 2007).

O leite de búfala apresenta maiores porcentagens de gordura, proteína e sólidos totais quando comparado ao leite de vaca, o que confere maior rendimento na obtenção de derivados lácteos e maior remuneração aos produtores. Além da já conhecida e tradicional *mozzarella*, é possível produzir também outros tipos de queijos como minas frescal, ricota, coalho, provolone, entre outros (TONHATI et al., 2008). Dentro desse contexto, a constituição do leite de búfalas em termos de gordura, proteína e sólidos totais é tão importante quanto a produção de leite em valores econômicos.

A estimativa da herdabilidade permite prever o progresso a ser obtido por meio da seleção. Alguns autores, trabalhando com búfalas leiteiras, obtiveram estimativas moderadas das herdabilidades para a produção de leite e seus constituintes, variando de 0,20 a 0,38 (MALHADO et al., 2007; ASPILCUETA-BORQUIS et al., 2010; RODRIGUES et al., 2010; SENO et al., 2010). Estes resultados sugerem que tais características possuem variabilidade genética aditiva suficiente para responderem a programas de seleção.

A identificação de genes responsáveis pela variação fenotípica das características produção de leite e seus constituintes é um passo importante para o desenvolvimento de métodos de seleção de indivíduos portadores de genótipos superiores. No intuito de contribuir com os programas de melhoramento genético de bubalinos no Brasil, o Departamento de Zootecnia da Unesp/Jaboticabal vem mantendo um programa de controle leiteiro com o objetivo de desenvolver pesquisas direcionadas à espécie bubalina.

Algumas características de importância econômica nos animais de produção, assim como a produção de leite e seus constituintes, podem ser determinadas pelos

efeitos ambientais e pela combinação de um grande número de genes. Estudos importantes têm sido realizados na identificação de genes associados à proteína e ao desempenho na produção de leite, tais como a Kappa caseína, Alpha-Lactoglobulina e Beta-Lactoglobulina (OTAVIANO et al., 2005; RIAZ et al., 2008; RAMESHA et al., 2008; MEIGNANALAKSHMI e MAHALINGA NAINAR, 2009).

A produção de leite e seus constituintes são características medidas somente em animais do sexo feminino e, portanto, para a avaliação dos touros é comum a realização de testes de progênie, que além de tornar a avaliação genética onerosa, também resulta em longos intervalos de gerações e, conseqüentemente, ganho genético mais lento. Contudo, os marcadores moleculares, definidos de acordo com Ferreira e Grattapaglia (1998) como características do DNA que diferenciam dois ou mais indivíduos, vêm contribuir como uma ferramenta auxiliar nos programas de seleção. Uma vez que algumas de suas vantagens é a possibilidade de se predizer o mérito genético dos animais sem a necessidade de mensurar o seu próprio fenótipo ou de parentes próximos (MEUWISSEN; HAYES; GODDARD, 2001).

A seleção assistida por marcadores moleculares (MAS) utiliza simultaneamente dados fenotípicos e de marcadores moleculares em ligação gênica próximas a uma região do DNA que afeta uma característica quantitativa (*Quantitative Traits Loci*: QTL). De acordo com Resende et al. (2008), tais marcadores são utilizados como covariáveis na explicação dos valores fenotípicos dos indivíduos ou como efeitos aleatórios incluídos no modelo para o fenótipo. Após modelagem estatística, esses marcadores são eleitos ou não como determinantes dos efeitos dos QTLs.

Entretanto, a seleção baseada na MAS implica em algumas limitações, tais como a detecção de um pequeno número de QTLs de efeito maior. Tais QTLs não explicam suficientemente toda a variação genética das características, pelo fato destas serem determinadas por um grande número de loci de pequenos efeitos e sofrerem influência ambiental. Além do mais, a MAS somente apresenta resultados consideráveis quando a seleção é baseada em dados fenotípicos e quando o tamanho da população genotipada é muito grande (DEKKERS, 2004).

Os avanços tecnológicos da última década permitiram o desenvolvimento de metodologias de alto desempenho e acurácia na caracterização e genotipagem de polimorfismos de nucleotídeo único (SNPs). Tais marcadores têm como base as

alterações mais elementares da molécula de DNA, ou seja, mutações de bases únicas da cadeia de bases nitrogenadas e, são extremamente abundantes nos genomas de espécies não endogâmicas (CAETANO, 2009). Esses marcadores são bastante úteis, pois permitem delimitar uma determinada região genômica que possui um QTL. Portanto, avanços importantes têm sido alcançados na genotipagem de animais com o uso de *chips* de SNPs, com plataformas já disponíveis para a detecção de milhares de SNPs simultaneamente em várias espécies animais.

A Affymetrix iniciou a comercialização de um painel que identifica 90.000 SNPs espalhados pelo genoma bubalino. O Axiom® Buffalo Genotyping Array é atualmente a única ferramenta disponível no mercado para a genotipagem de alta densidade em búfalos, possibilitando a genotipagem de duas espécies (*Bubalus bubalis bubalis* e *Bubalus bubalis carabanensis*). Tal tecnologia pode promover avanços nos estudos de localização e associação de genes que afetam características complexas, tais como a produção de leite e a porcentagem de gordura no leite de búfalas.

1.1.1 Seleção genômica

A seleção genômica ampla (GWS), proposta por Meuwissen, Hayes e Goddard (2001), enfatiza a predição simultânea dos efeitos genéticos de milhares de marcadores do DNA dispersos no genoma de um organismo, de forma a capturar os efeitos de todos os loci e explicar a variação genética da característica quantitativa. Sua condição primordial é que haja desequilíbrio de ligação, a nível populacional, entre alelos dos marcadores e alelos dos genes associados ao caráter. A predição do valor genético é feita com base nas informações fenotípicas e genotípicas de indivíduos pertencentes a uma amostra da população de seleção.

A superioridade da GWS em relação à seleção por meio da genética quantitativa (aquela baseada apenas no fenótipo e pedigree) pode ser atribuída a alguns fatores, como o uso da matriz de parentesco genômica e a combinação dos três tipos de informação (fenotípica, genotípica e genealógica), fatos que contribuem para o aumento da acurácia (RESENDE et al., 2010).

Algumas metodologias de predição genômica, sejam elas de caráter frequentista ou bayesiana, surgiram com o propósito de estimar os efeitos dos

marcadores. Na abordagem bayesiana, cada modelo assume uma pressuposição diferente acerca da distribuição *a priori* para os efeitos dos SNPs.

Comparações entre os modelos bayesianos para a predição genômica de algumas características em bovinos leiteiros têm sido realizadas. A superioridade do método Bayes B em relação aos métodos Bayes A e GBLUP foi relatada (MEUWISSEN; HAYES; GODDARD, 2001; HABIET et al. 2010). Em relação ao Bayes C, Habier et al. (2011) 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.

1.1.2 Associação genômica ampla

Os estudos de associação genômica ampla (*Genome-wide Association Studies* – GWAS) fazem o uso da combinação do genótipo e do fenótipo do indivíduo para identificar regiões no genoma que estejam possivelmente relacionadas às características de interesse.

Tais estudos se tornaram possível a partir do uso das tecnologias que utilizam os polimorfismos de nucleotídeo único (SNPs). Por sua vez, os SNPs são definidos como uma variação na sequência do DNA que afeta apenas uma base e ocorrem com frequência em humanos e outras espécies animais (BUSH e MOORE, 2012).

De acordo com Peters et al. (2012), as pesquisas que associam os SNPs, provenientes de painéis de alta densidade, às características que são objeto de estudo possuem muitas aplicações práticas, tais como a detecção de QTLs e construção de bibliotecas de genes. Estes estudos também são indispensáveis para o conhecimento da estrutura genética da característica de interesse numa determinada população.

No entanto, segundo Dekkers et al. (2004), o tamanho da amostra deve ser suficientemente grande e representativo da população de interesse para que potenciais resultados errôneos não aconteçam, tais como resultados falso-positivos e a superestimação dos efeitos dos QTLs significativos, além de inconsistências na fase ligação marcador-QTL.

Com a finalidade de aumentar o número de animais avaliados, Wang et al. (2012) propuseram um método em que são utilizadas nas análises as informações de

genótipo dos animais que foram genotipados, assim como o fenótipo e pedigree de indivíduos genotipados e não genotipados, em um único passo (ssGWAS). Nessa abordagem, os efeitos de todos os SNPs são considerados. Portanto, a utilização de um modelo que faz uso de todas as informações disponíveis pode trazer ganhos, uma vez que uma amostra mais representativa da população está sendo analisada.

1.2 Objetivos

Comparar diferentes métodos Bayesianos de predição genômica para a produção de leite e as porcentagens de gordura e proteína no leite de búfalas, com o intuito de apontar o melhor modelo de predição genômica para estas características e realizar um estudo de associação genômica ampla a fim de identificar regiões cromossômicas e genes possivelmente relacionadas às mesmas, e assim, obter um melhor entendimento acerca do controle genético destas características.

2. REFERÊNCIAS BIBLIOGRÁFICAS

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, 2010.

BERNARDES, O. Bubalinocultura no Brasil: situação e importância econômica. **Revista Brasileira de Reprodução Animal**, Belo Horizonte, v. 31, n. 3, p. 293-298, 2007.

BORGHESE, A. Buffalo Production and Research, Rome 2005. FAO. Food Agriculture Organization of The United Nations. Disponível em: <<http://www.fao.org/docrep/010/ah847e/ah847e00.htm>> Acesso em 30/11/2016.

BUSH, W. S.; MOORE, J. H. Genome-wide association studies. *PLoS Comput Biol*, v. 8, n. 12, p. e1002822-e1002822, 2012.

CAETANO, A. R. Marcadores SNP: Conceitos básicos, aplicações no manejo e no melhoramento animal e perspectivas para o futuro. **Revista Brasileira de Zootecnia**, v. 38, n. SUPPL. 1, p. 64–71, 2009.

DEKKERS, J.C.M. Commercial application of marker and gene assisted selection in livestock: strategies and lessons. **Journal of Animal Science**, v. 82, p.313-328, 2004.

FERREIRA, M.E.; GRATTAPALIA, D. Introdução ao uso de marcadores moleculares em análise genética. 3d. Brasília: Embrapa-Cenargen, 1998. 220p.

HABIER, D. et al. The impact of genetic relationship information on genomic breeding values in German Holstein cattle. **Genetics, selection, evolution : GSE**, v. 42, n. Ld, p. 5, 2010.

HABIER, D. et al. Extension of the bayesian alphabet for genomic selection. **BMC bioinformatics**, v. 12, n. 1, p. 186, 2011.

HAYES, B. J. et al. Genomic selection in dairy cattle: Progress and challenges. **Journal of Dairy Science**, v. 92, p. 433-443, 2009.

MALHADO, C.H.M. et al. Parâmetros e tendências da produção de leite em bubalinos da raça Murrah no Brasil. **Revista Brasileira de Zootecnia**, v. 36, n. 2, p. 376-379, 2007.

MEIGNANALAKSHMI, S.; MAHALINGA NAINAR, A. PCR-RFLP analysis of beta-lactoglobulin gene in Murrah buffaloes. **Tamilnadu J. Veterinary & Animal Sciences**, v. 5, n. 5, p. 194-197, 2009.

MEUWISSEN, T.H.E.; HAYES, B.J.; GODDARD, M.E. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. **Genetics**, n. 157, p. 1819-1829, 2001.

- OTAVIANO, A.R. et al. Kappa-casein gene study with molecular markers in female buffaloes (*Bubalus bubalis*). **Genetics and Molecular Biology**, v. 28, n. 2, p. 237-241, 2005.
- PETERS, S.O et al. Bayesian genome-wide association analysis of growth and yearling ultrasound measures of carcass traits in Brangus heifers. **Journal of Animal Science**, v.90, p.3398–3409, 2012.
- RAMESHA, K.P. et al. Alpha-Lactalbumin Gene Polymorphism: A preliminary study on two breeds of the river Buffalo (*Bubalus bubalis*). **AsPac J. Mol. Biol. Biotechnol.**, v. 16, n. 2, p. 47-52, 2008.
- RESENDE, M.D.V. et al. Seleção genômica ampla (GWS) e maximização da eficiência do melhoramento genético. **Pesquisa Florestal Brasileira**, n. 56, p. 63, 2008.
- RESENDE, M.D.V. et al. Computação da seleção genômica ampla (GWS). **Empresa Brasileira de Pesquisa Agropecuária**. Colombo: Embrapa Florestas, 2010.
- RIAZ, M.N. et al. Molecular marker assisted study of kappa-casein gene in Nili-Ravi (Buffalo) breed of Pakistan. **Pakistan Vet. J.**, v. 28, n. 3, p. 103-106, 2008.
- RODRIGUES, A.E. et al. Estimação de parâmetros genéticos para características produtivas em búfalos na Amazônia Oriental. **Arq. Bras. Med. Vet. Zootec.**, v.62, n.3, p.712-717, 2010.
- SANTIAGO, A. A. Introdução dos Búfalos no Brasil. Associação Brasileira dos Criadores de Búfalos. 13p., 2000.
- SENO, L. et al. 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*. Disponível em: <<http://www.lrrd.org/lrrd22/2/seno22038.htm>> Acesso em 30/11/2016.
- TONHATI, H. et al. Programa de mejoramiento genético de búfalos. **Revista Argentina de Producción Animal**, v. 28, n. 1, p. 53-67, 2008.
- WANG, H. et al. Genome-wide association mapping including phenotypes from relatives without genotypes. **Genetics Research**, v.94, n.2, p.73-83, 2012.
- ZHANG et al. Progress of genome wide association study in domestic animals. **Journal of Animal Science and Biotechnology**, v. 3, n. 26, 2012.

CAPÍTULO 2 – Métodos Bayesianos aplicados à predição genômica de características de produção e qualidade do leite de búfalas

RESUMO – Um dos derivados do leite de búfala mais requisitado pelos consumidores é a *mozzarella*. Tradicionalmente, o melhoramento genético de búfalos tem sido realizado utilizando-se informações fenotípicas e de pedigree. No entanto, com o advento das tecnologias que utilizam as informações dos marcadores do tipo polimorfismos de nucleotídeo único (SNPs), é possível realizar a seleção genômica. Algumas de suas vantagens são a possibilidade da avaliação genética antecipada de animais jovens e a consequente redução dos intervalos de gerações. O objetivo deste estudo foi comparar diferentes métodos Bayesianos de predição genômica para as características produção de leite truncada aos 305 dias (PL) e as porcentagens de gordura (%G) e proteína (%P) do leite de búfalas. Foram genotipados 289 animais utilizando o painel 90 K Axiom® Buffalo Genotyping (Affymetrix). Os seguintes modelos Bayesianos foram utilizados para a predição dos valores genéticos genômicos: Bayes A (BA), Bayes B (BB), Bayes C (BC) e Bayes LASSO (BL). O fenótipo corrigido para os efeitos fixos (Y^*) foi utilizado como variável resposta nas análises genômicas. A habilidade de predição dos diferentes modelos foi avaliada usando o método leave-one-out de validação cruzada. As acurácias de predição foram calculadas através da correlação de Pearson entre o valor genético genômico estimado (GEBV) e a variável resposta (Y^*) para cada modelo e característica avaliados. Em geral, todos os modelos Bayesianos demonstraram semelhantes acurácias de predição, variando de 0,41 a 0,42; 0,38 a 0,39 e 0,39 a 0,40 para a PL, %G e %P, respectivamente. Portanto, os métodos BA, BB, BC e BL podem ser utilizados nas predições dos efeitos dos SNPs, obtendo-se, praticamente, as mesmas acurácias de predição.

Palavras-chave: acurácia de predição, búfalos, GEBV, seleção genômica, SNP

2.1 INTRODUÇÃO

O rebanho bubalino brasileiro apresentou elevado crescimento a partir da década de 80 como uma atividade voltada, quase que exclusivamente, a produção de carne. No entanto, somente a partir dos anos 90 houve um grande interesse por parte dos produtores na produção de leite. Este fato estar relacionado ao atrativo retorno econômico que esta atividade proporciona devido ao maior rendimento na obtenção dos derivados lácteos. Tal rendimento é atribuído a constituição do leite de búfalas, uma vez que apresenta maiores teores de gordura, proteína e sólidos totais, quando comparado ao leite de vacas (TONHATI et al., 2008).

Um dos derivados do leite de búfala mais requisitado pelos consumidores é a *mozzarella*. Este produto, além de apresentar sabor peculiar, demonstra qualidades nutricionais superiores ao do queijo oriundo do leite de vaca (SAMEEN et al., 2008). Contudo, para alavancar a produção de queijo *mozzarella*, faz-se necessário melhorar os índices produtivos da bubalinocultura leiteira no Brasil.

A seleção, por meio da genética quantitativa, de animais superiores para as características de produção e qualidade do leite de búfalas tem sido realizada. Porém, tais características não podem ser medidas diretamente em touros, uma vez que apenas fêmeas expressam as mesmas. No entanto, com o advento das tecnologias que utilizam as informações dos marcadores do tipo polimorfismos de nucleotídeo único (SNPs), é possível realizar a seleção genômica. Algumas de suas vantagens são a possibilidade da avaliação genética antecipada de animais jovens e a consequente redução dos intervalos de gerações.

A seleção genômica consiste nas informações de um grande número de SNPs espalhados pelo genoma, assumindo-se que todos os QTLs (*Quantitative Trait Loci*) estejam em desequilíbrio de ligação com pelo menos um marcador. Sua implementação se dá por meio de uma equação de estimação dos efeitos dos marcadores que é obtida através de uma população com dados fenotípicos e genotípicos conhecidos. Em seguida, a estimativa dos efeitos dos marcadores é utilizada para prever o valor genético genômico (GEBV – *Genomic Estimated Breeding Value*) dos animais candidatos à seleção, que, por sua vez, possuem apenas genótipos conhecidos (MEUWISSEN; HAYES e GODDARD, 2001).

Muitas são as metodologias de seleção genômica que têm sido propostas com o objetivo de incorporar as informações dos efeitos dos marcadores genéticos nos modelos. A escolha da metodologia, que pode diferir entre si em relação à distribuição dos efeitos dos QTLs, pode ter impacto direto nas acurácias das predições genômicas a depender do tipo de característica e população analisada (DE LOS CAMPOS et al., 2013). Portanto, a escolha de métodos apropriados para a predição genômica pode contribuir para o melhoramento genético das características de produção e qualidade do leite de búfalas no Brasil.

Comparações entre alguns modelos que utilizam a abordagem bayesiana têm sido realizadas em estudos com características de bovinos leiteiros (Hayes et al., 2009; Moser et al., 2009; Habier et al. 2010). Resultados discrepantes foram observados quando alguns autores trabalharam com dados simulados. No entanto, quando os estudos foram realizados com dados reais, poucas diferenças acerca da habilidade de predição para as metodologias bayesianas foram verificadas (DAETWYLER et al., 2013; DE LOS CAMPOS et al., 2013). Portanto, nenhum dos métodos foi claramente superior em todas as situações.

Visto que há escassez das informações na literatura a respeito da aplicação da seleção genômica em rebanhos bubalinos, o objetivo do presente estudo foi comparar diferentes métodos bayesianos de predição genômica com o propósito de identificar o melhor modelo para a produção de leite e as porcentagens de gordura e proteína no leite de búfalas no Brasil.

2.2 MATERIAL E MÉTODOS

2.2.1 Arquivos de dados fenotípicos e genotípicos

Foram utilizadas informações fenotípicas e genotípicas de 289 búfalas da raça Murrah, provenientes de dois rebanhos brasileiros localizados nos Estados do Rio Grande do Norte e São Paulo. Estes animais fazem parte do programa de controle leiteiro mantido pelo Departamento de Zootecnia da FCAV/UNESP Jaboticabal. As características analisadas foram: produção de leite truncada aos 305 dias (PL), porcentagem de gordura (%G) e a porcentagem de proteína (%P) no leite de búfalas.

Os dados analisados foram baseados em 1.381 lactações de 289 búfalas mensalmente registradas entre os anos de 2004 e 2016. Foram consideradas informações de controles obtidos a partir do quinto dia de produção, sendo mantidas apenas as lactações com duração superior a 90 dias. Os grupos de contemporâneos (GC) foram formados de acordo com o rebanho, ano e estação de parto, com a restrição de que cada GC deveria conter no mínimo quatro animais, assim como registro das características entre ± 3 desvios padrão da média do grupo. As estatísticas descritivas das características analisadas estão presentes na tabela 1.

Tabela 1. Estatísticas descritivas para a produção de leite (PL) e para as porcentagens de gordura (%G) e proteína (%P).

Característica	N	Média	Desvio padrão	GC
PL (Kg)	1.381	2.428,00	842,89	34
%G	979	6,16	1,08	22
%P	988	4,15	0,27	22

N = número, GC = Grupos de contemporâneos

Os componentes de variância (Tabela 2) foram estimados pelo método de Máxima Verossimilhança Restrita, em análise unicaracterística, utilizando-se o programa REMLF90 (MISZTAL, 2005). Foram considerados como efeitos fixos os grupos de contemporâneos, como covariável a idade da búfala ao parto (linear e quadrático) e como efeito aleatório os efeitos genético aditivo, ambiente permanente e residual. A notação matricial do modelo animal utilizado pode ser assim apresentada:

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

Em que y é o vetor da característica observada, X é a matriz de incidência dos efeitos fixos, β é o vetor dos efeitos fixos, Z é a matriz de incidência dos efeitos aleatórios genéticos aditivos, a é o vetor dos efeitos aleatórios genéticos aditivos, W é a matriz de incidência do efeito aleatório de ambiente permanente, p é o vetor do efeito aleatório de ambiente permanente e e é o vetor dos efeitos aleatórios residuais.

Tabela 2. Variância genética aditiva (σ_a^2), ambiente permanente (σ_{pe}^2), residual (σ_e^2), herdabilidade (h^2) e repetibilidade (r) para as características produção de leite (PL) e as porcentagens de gordura (%G) e proteína (%P).

Característica	σ_a^2	σ_{pe}^2	σ_e^2	h^2	r
PL	78.930	77.710	189.100	0,23	0,50
%G	0,1486	0,1187	0,4891	0,20	0,56
%P	0.0158	0,0115	0,0337	0,26	0,58

Os animais foram genotipados utilizando o painel 90 K Axiom® Buffalo Genotyping (Affymetrix). Embora a seleção dos SNPs incluídos neste chip seja baseada em dados de sequenciamento dos búfalos, a posição dos SNPs e os genes anotados utilizam como referência o genoma bovino, uma vez que não há disponível, até o momento, referência de genoma completo público para esta espécie (DE CAMARGO et al., 2015).

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 p -value < 10^{-6} . Os SNPs coincidentes também foram eliminados. Foi considerado call rate < 0,90 como critério de exclusão da amostra. A imputação dos marcadores perdidos foi realizada utilizando-se o programa FImpute (SARGOLZAEI et al., 2014). Em seguida um novo controle de qualidade foi realizado: MAF (Frequência de Alelos Menores) < 0,05 e Call Rate < 0,95. Após os critérios adotados, um total de 45.125 SNPs foi considerado para o presente estudo.

O fenótipo corrigido para os efeitos fixos e ambiente permanente (Y^*) foi utilizado como variável resposta nas análises genômicas. Este foi estimado utilizando-se o pacote do R *Bayesian Generalized Linear Regression* - BGLR (PÉREZ E DE LOS CAMPOS, 2014), empregando-se um modelo animal considerando como efeitos fixos os grupos de contemporâneos, como covariável a idade da búfala ao parto (linear e quadrático) e como efeito aleatório os efeitos genético aditivo, ambiente permanente e residual.

2.2.2 Predições genômicas

Foram utilizadas quatro metodologias de predição genômica, as quais diferem de acordo com as pressuposições para os efeitos dos marcadores e suas variâncias: Bayes A, Bayes B, Bayes C e LASSO Bayesiano.

O modelo geral para as predições genômicas pode ser descrito na forma matricial como:

$$y = \mu + Wg + e$$

Em que \mathbf{y} é o vetor do fenótipo corrigido para os efeitos fixos e ambiente permanente (Y^*); μ é a média geral; \mathbf{W} é uma matriz ($n \times p$) que consiste nos genótipos de p marcadores SNP para cada um dos n animais; \mathbf{g} é o vetor dos efeitos dos marcadores para os quais as pressuposições dependem da metodologia a ser utilizada; \mathbf{e} é o vetor dos efeitos residuais.

i) Bayes A

A metodologia Bayes A foi proposta por Meuwissen, Hayes e Goddard, (2001). Neste caso, os efeitos dos marcadores são assumidos como amostras de uma distribuição normal com média zero e variância de cada marcador dada por uma distribuição qui-quadrado escalonada invertida *a priori*, com ν_{β} graus de liberdade e S^2_{β} como parâmetro de escala da distribuição. O modelo Bayes A apresenta distribuição marginal *a priori* dos efeitos genéticos marcadores como uma distribuição t de Student.

ii) Bayes B

A metodologia Bayes B, também proposta por Meuwissen, Hayes e Goddard, (2001), apresenta as mesmas pressuposições da Bayes A. No entanto, a diferença entre as duas metodologias consiste na pressuposição de que uma fração π dos SNPs em Bayes B são as mesmas em Bayes A, no entanto outra fração $(1-\pi)$ assume que os SNPs apresentam efeitos nulos, considerando que em muitos loci não existe variação genética, ou seja, não há segregação.

iii) Bayes C

A metodologia Bayes C foi proposta por Habier et al. (2011). Neste método, assume-se que todos os SNPs têm variância comum seguindo distribuição qui-quadrado invertida escalonada *a priori*, com v_g graus de liberdade e um parâmetro de escala S_g^2 . Assumindo-se ponto de massa em zero para o efeito dos SNPs com probabilidade π , e amostrando-se de uma distribuição gaussiana os efeitos dos SNPs com probabilidade $(1-\pi)$. O parâmetro π é tratado como desconhecido e tem como *priori* uma distribuição Beta, $\pi \sim \text{Beta}(p_0, \pi_0)$, onde $p_0 > 0$ e $\pi_0 \in [0, 1]$. Por *default* $p_0 = 10$ e $\pi_0 = 0,5$.

iv) Least Absolut Shrinkage and Selection Operator - LASSO

No LASSO Bayesiano (PARK e CASELLA, 2008) a distribuição marginal *a priori* para os efeitos dos SNPs é uma dupla exponencial com parâmetro $\lambda^2 \sim \text{Gamma}$. No qual a variância residual (σ_e^2) assume uma distribuição qui-quadrado invertida escalonada *a priori*, com df_e graus de liberdade e um parâmetro de escala S_e .

λ é o parâmetro “sharpness”, com *priori* vaga, sendo uniforme entre 0 e 1.000.000. A parametrização consiste em estimar as variâncias individuais dos SNPs, condicional à regularização do parâmetro λ (VANRADEN, 2008). Somente para a característica PL, o λ obteve valor zero após 300 interações. Portanto, o modelo foi ajustado usando diferentes valores de lambda variando de 0,01 a 300 e o valor de lambda que forneceu a maior acurácia de predição (5,0) foi escolhido.

2.2.4 Critérios de comparação entre os modelos

O método *leave-one-out* de validação cruzada (*Leave-one-out Cross Validation-LOOCV*) foi utilizado para comparar a habilidade de predição dos diferentes modelos e características estudados. No LOOCV, um modelo foi construído utilizando o número total de indivíduos menos um ($n-1$) na população de treinamento e validado no único indivíduo da amostra que não foi utilizado na população de treinamento. Este procedimento foi repetido n vezes, assim cada animal na amostra foi utilizado para validar o modelo testado. De acordo com Resende et al. (2008), este método maximiza o tamanho da população de treinamento. As acurácias de predição foram calculadas através da correlação de Pearson entre o valor genético genômico estimado (GEBV)

e a variável resposta (Y^*) para cada modelo e característica. O quadrado médio do erro de predição (*Mean Squared Error* – MSE) também foi utilizado como medida de verificação do ajuste do modelo e, portanto, quanto menor o valor de MSE melhor é o ajuste do modelo.

As análises de predição dos valores genéticos genômicos foram realizadas utilizando o pacote do R *Bayesian Generalized Linear Regression* (BGLR) desenvolvido por Pérez e De los Campos (2014). Um total de 300 mil iterações foi utilizado para a cadeia de Markov (MCMC), com um *burn-in* de 100 mil ciclos, retirando-se uma amostra a cada 50 ciclos. A convergência foi verificada através de inspeção visual. As análises foram realizadas utilizando o *Center for High Throughput Computing* pertencente ao Departamento de Ciências Computacionais da Universidade de Wisconsin (Madison – EUA).

2.3 RESULTADOS E DISCUSSÃO

As acurácias de predição variaram de 0,41 a 0,42, 0,38 a 0,39 e 0,39 a 0,40 para PL, %G e %P, respectivamente (Tabela 3). Em relação ao MSE, os valores variaram de 412.355,50 a 414.495,60 para a PL. No que se refere às características %G e %P, os mesmos valores foram encontrados para as mesmas: 1,0 e 0,06, respectivamente.

Tabela 3. Acurácias de predição medidas pela correlação de Pearson entre o Y corrigido e o valor genético genômico $r(Y^*, \text{GEBV})$ e o quadrado médio do erro de predição (MSE) para as características produção de leite (PL) e as porcentagens de gordura (%G) e proteína (%P).

		PL	%G	%P
$r(Y^*, \text{GEBV})$	BA	0,42	0,39	0,40
	BB	0,41	0,38	0,40
	BC	0,42	0,39	0,40
	BL	0,42	0,38	0,39
MSE	BA	413.219,50	1,00	0,06
	BB	414.495,60	1,00	0,06
	BC	412.355,50	1,00	0,06
	BL	413.529,10	1,00	0,06

BA = Bayes A, BB = Bayes B, BC = Bayes C, BL = Bayes LASSO

De modo geral, as acurácias de predição obtidas para os diferentes modelos analisados apresentaram valores semelhantes para todas as características. Estes resultados estão de acordo com os relatados na literatura quando Moser et al., (2009) trabalharam com bovinos de leite. De acordo com os mesmos autores, os diferentes métodos de predição genômica precisam lidar com o problema da super parametrização resultante do ajuste de muitos parâmetros para um relativo pequeno número de dados. Segundo Gianola (2013), quando o número de parâmetros a serem estimados é bem maior que o número de amostras ($n \ll p$), não se dispõe de número suficiente de informações para que o aprendizado bayesiano ocorra de modo pleno, tornando mínima a influência da distribuição *a priori*. Por isso, resultados similares entre os diferentes métodos são esperados.

Outro fator que pode ter contribuído para que os resultados das acurácias de predição não tenham sido discrepantes entre si, é o tipo de dado a ser utilizado nas análises. De acordo com de los Campos et al. (2013), os estudos com dados simulados têm demonstrado a superioridade de alguns modelos, como por exemplo, o modelo Bayes B. No entanto, isto não tem sido plenamente confirmado em análises com dados reais, ou seja, as diferenças apresentadas entre os modelos não foram relevantes.

Além disso, a complexidade da arquitetura genética e o tipo das características analisadas parece influenciar nos valores das acurácias de predição quando os modelos de predição genômica são comparados. De acordo com Daetwyler et al. (2013), trabalhando com dados simulados, os modelos apresentaram similaridades em relação as suas capacidades preditivas para as características afetadas por muitos loci de pequenos efeitos. O contrário ocorreu quando poucos loci contribuíam para a variação genética da característica. No entanto, ao trabalharem com dados reais, todos os modelos alcançaram acurácias semelhantes para ambas características de diferentes complexidades.

O pequeno número de animais genotipados no presente estudo (289 búfalas) pode também ter contribuído para a semelhança dos valores das acurácias de predição. Estes resultados estão de acordo com os relatados previamente na literatura, quando os autores trabalharam com um reduzido número de indivíduos genotipados, para diferentes características de baixa herdabilidade em bovinos de corte (FERNANDES JÚNIOR et al., 2016; MAGNABOSCO et al., 2016).

No presente estudo, as herdabilidades observadas para as características foram de moderada magnitude (variando de 0,21 a 0,22). Vale ressaltar que, maiores acurácias de predição são alcançados quando se trabalha com características que sofrem menores interferências ambientais, ou seja, as que apresentam herdabilidades de elevada magnitude. De acordo com Goddard e Hayes (2009), é necessário um grande número de animais para detectar associações de QTLs com características de baixa herdabilidade, e que, portanto, a seleção genômica para estas características podem ser menos acuradas e poucos QTLs serão encontrados.

De maneira semelhante às acurácias de predição, os MSE para as características estudadas não apresentaram diferenças consideráveis entre si (Tabela 3). Em relação a PL, o menor valor verificado foi para o modelo Bayes C. No entanto, devido às semelhanças apontadas entre os modelos estudados, os resultados apresentados aqui não puderam ser levados em conta na hora da escolha do melhor modelo bayesiano para a predição genômica das características analisadas.

As acurácias observadas nesta pesquisa, para as características analisadas, foram consideradas de baixa magnitude quando comparadas aos demais estudos em bovinos de leite da raça holandesa (HAYES et al., 2009). Porém, uma vez que tais

características só podem ser mensuradas em animais do sexo feminino, faz-se necessário realizar testes de progênie a fim de avaliar os touros. Com isso, a seleção genômica poderia promover ganhos para a bubalinocultura leiteira no Brasil devido a possibilidade da avaliação antecipada do genótipo de animais candidatos à seleção, com a consequente redução no intervalo de gerações e diminuição dos custos com testes de progênie.

2.4 CONCLUSÃO

Os modelos BA, BB, BC e BL podem ser utilizados para a implementação da seleção genômica de búfalos leiteiros, obtendo-se praticamente as mesmas acurácias de predição.

2.5 REFERÊNCIAS BIBLIOGRÁFICAS

DAETWYLER, H. D. et al. Genomic prediction in animals and plants: Simulation of data, validation, reporting, and benchmarking. **Genetics**, v. 193, n. 2, p. 347–365, 2013.

DE CAMARGO, G. M. F. . M. F. et al. Prospecting major genes in dairy buffaloes. **BMC Genomics**, v. 16, n. 1, p. 872, 2015.

DE LOS CAMPOS, G. et al. Whole-genome regression and prediction methods applied to plant and animal breeding. **Genetics**, v. 193, n. 2, p. 327–345, 2013.

FERNANDES JÚNIOR, G. A. et al. Genomic prediction of breeding values for carcass traits in Nellore cattle. **Genetics, selection, evolution : GSE**, v. 48, n. 1, p. 7, 2016.

GIANOLA, D. Priors in Whole-Genome Regression: The Bayesian Alphabet Returns. **Genetics**, v.194, p.573–596, 2013.

GODDARD, M. E.; HAYES, B. J. Mapping genes for complex traits in domestic animals and their use in breeding programmes. **Nature Reviews Genetics**, v. 10, n. 6, p. 381–391, 2009.

HABIER, D. et al. The impact of genetic relationship information on genomic breeding values in German Holstein cattle. **Genetics, selection, evolution : GSE**, v. 42, n. 1, p. 5, 2010.

HABIER, D. et al. Extension of the bayesian alphabet for genomic selection. **BMC bioinformatics**, v. 12, n. 1, p. 186, 2011.

HAYES, B. J. et al. Genomic selection in dairy cattle: Progress and challenges. **Journal of Dairy Science**, v. 92, p. 433-443, 2009.

MAGNABOSCO, C. U. et al. Accuracy of genomic breeding values for meat tenderness in {Polled} {Nelore} cattle. **Journal of Animal Science**, n. 020802006, p. 1–9, 2016.

MEUWISSEN, T. H. E.; HAYES, B. J.; GODDARD, M. E. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. **Genetics**, v. 157, p. 1819-1829, 2001.

MISZTAL, I. [2005]. REMLF90 manual. Disponível em: <<ftp://nce.ads.uga.edu/pub/ignacy/blupf90/>>. Acesso em: 11/10/2016.

MOSER, G. et al. A comparison of five methods to predict genomic breeding values of dairy bulls from genome-wide SNP markers. **Genetics Selection Evolution**, v. 41, n. 1, p. 56, 2009.

PARK, T.; CASELLA, G. The Bayesian Lasso. **Journal of the American Statistical Association**, v. 103, n. 482, p. 681–686, 2008.

PÉREZ, P.; DE LOS CAMPOS, G. Genome-wide regression and prediction with the BGLR statistical package. **Genetics**, v. 198, n. 2, p. 483–495, 2014.

RESENDE, M.D.V. et al. Seleção genômica ampla (GWS) e maximização da eficiência do melhoramento genético. **Pesquisa Florestal Brasileira**, n. 56, p. 63, 2008.

SAMEEN, A. et al. Quality evaluation of mozzarella cheese from different milk sources. **Pakistan Journal of Nutrition**, v. 7, n. 6, p. 753-756, 2008.

SARGOLZAEI, M.; CHESNAIS, J. P.; SCHENKEL, F. S. A new approach for efficient genotype imputation using information from relatives. **BMC genomics**, v. 15, n. 1, p. 478, 2014.

TONHATI, H. et al. Programa de mejoramiento genetico de bufalos. **Revista Argentina de Producción Animal**, v. 28, n. 1, p. 53–67, 2008.

VANRADEN, P.M. Efficient Methods to Compute Genomic Predictions. **Journal of Dairy Science**, 91, p. 4414–4423, 2008.

CAPÍTULO 3 - Estudo de associação genômica ampla para características de produção e qualidade do leite de búfalas

RESUMO – Os estudos de associação genômica ampla (GWAS) são um importante passo na identificação de regiões cromossômicas e genes relacionados às características de interesse econômico das espécies domésticas. No entanto, geralmente, apenas uma fração de animais numa população é genotipada. Nesses casos, uma alternativa é o uso de modelos estatísticos que combinam a informação fenotípica de indivíduos genotipados e não genotipados, como o *single-step* GWAS (ssGWAS). O objetivo deste trabalho foi identificar regiões do genoma bubalino possivelmente relacionadas às características de produção de leite (PL) e as porcentagens de gordura (%G) e proteína (%P) do leite de búfalas no Brasil, utilizando informações de indivíduos genotipados e não genotipados. O número de animais com fenótipo foi 3.355, o arquivo de pedigree continha 15.495 animais, dos quais 322 foram genotipados com o painel 90 K Axiom® Buffalo Genotyping array. Um modelo de repetibilidade foi utilizado. Um processo iterativo foi realizado para calcular os pesos dos marcadores em função do quadrado dos efeitos dos SNPs e das frequências alélicas. Os dez SNPs de maiores efeitos para a PL, %G e %P explicaram 7,48, 9,94 e 6,56% da proporção da variância genética, respectivamente. Tais SNPs estão localizados nos cromossomos 6, 7, 16, 17, 20 e 29 para PL, 1, 5, 6, 17, 18 e 26 para %G e 1, 2, 6, 8, 9, 15, 26 e 27 para %P. Essas regiões abrigam genes com funções biológicas que podem estar relacionados com as características analisadas. Tais regiões e genes identificados poderão contribuir para o melhor entendimento sobre a influência dos mesmos nas características produção de leite e as porcentagens de gordura e proteína no leite de búfalas.

Palavras-chave: búfalos, GWAS, *single-step*, SNP

3.1 INTRODUÇÃO

Os búfalos são animais de grande importância socioeconômica, principalmente nos países em desenvolvimento, devido a sua contribuição na produção de carne, leite e trabalho. A adaptabilidade desses animais ao clima brasileiro permitiu um aumento significativo do rebanho a partir da década de 80. Atualmente, o rebanho bubalino brasileiro apresenta 1,3 milhões de animais (IBGE, 2015) e quatro raças oficialmente reconhecidas pela Associação Brasileira de Búfalos: Mediterrâneo, Murrah, Jafarabadi (*Bubalus bubalis*, $2n=50$) e Carabao (*Bubalus carabensis*, $2n=48$).

Na Itália, a população de búfalos leiteiros sofreu uma forte seleção genética, especialmente para aumentar a produção de *mozzarella*, o produto lácteo mais requisitado e famoso do mercado. No Brasil, esse tipo de queijo também ganhou destaque na culinária local e, portanto, há uma grande demanda no mercado nacional. O leite de búfalas, apresenta elevados teores de gordura, proteína e sólidos totais. Este fato, confere maior rendimento dos produtos derivados do leite e, conseqüentemente, maior remuneração aos produtores (TONHATI et al., 2008).

As características de importância econômica nos animais de produção, assim como a produção de leite e seus constituintes, podem ser determinadas pela combinação de um grande número de genes. Alguns estudos têm demonstrado que alguns genes estão relacionados à expressão da produção e qualidade do leite de búfalas tais como a Kappa caseína, Alpha-Lactoglobulina e Beta-Lactoglobulina (OTAVIANO et al., 2005; RIAZ et al., 2008; RAMESHA et al., 2008; MEIGNANALAKSHMI e MAHALINGA NAINAR, 2009).

Os estudos de associação genômica ampla (GWAS) possibilitam a investigação da arquitetura genética das características através da identificação de polimorfismos de nucleotídeo único (SNPs), explorando a região do genoma em que este SNP está inserido e com isso buscar genes ou QTLs (*Quantitative Traits Loci*) relacionados à expressão fenotípica da característica em estudo (UTSUNOMYA et al., 2013). Avanços importantes têm sido alcançados na genotipagem de animais utilizando *chips* de SNPs, com plataformas já disponíveis no mercado para a detecção simultânea de milhares desses marcadores, em várias espécies domésticas.

O único painel disponível no mercado para a genotipagem de alta densidade de búfalos tem possibilitado avanços nos estudos de localização e associação de genes que afetam a produção do leite de búfalas (EL-HALAWANY et al., 2017). No entanto, a seleção de SNPs incluídos neste chip é baseada em dados bubalinos sequenciados, porém a posição dos SNPs e os genes anotados utilizam o genoma bovino como referência, uma vez que não há disponível, até o momento, referência de genoma completo público para esta espécie animal (DE CAMARGO et al., 2015).

Os estudos que possibilitam a identificação de genes e/ou regiões cromossômicas que estejam relacionados à expressão fenotípica das características de produção de leite de búfalas e seus constituintes, podem auxiliar na construção de bibliotecas genômicas próprias para a espécie e assim, contribuir para o melhor entendimento do controle genético destas características. Além disso, estes tipos de estudos podem tornar possível o desenvolvimento de um painel específico para bubalinos, e de menor dimensão, o que tornaria esta tecnologia genômica mais acessível para o produtor.

Uma vez que, apenas alguns indivíduos podem ser genotipados devido ao elevado custo, Wang et al. (2012) propuseram o método *single-step* GWAS, no qual consiste na adaptação da metodologia *single-step* GBLUP (MISZTAL et al., 2009). Esta metodologia reúne em um único passo todas as informações disponíveis (fenótipo, genótipo e pedigree), além dos dados fenotípicos de animais que não foram genotipados. Com isso, a utilização de um modelo que faz uso de todas as informações disponíveis pode trazer ganhos, uma vez que uma amostra mais representativa da população está sendo analisada.

O objetivo deste trabalho foi identificar regiões do genoma bubalino possivelmente relacionadas às características de produção e qualidade do leite de búfalas no Brasil, utilizando informações de indivíduos genotipados e não genotipados.

3.2 MATERIAL E MÉTODOS

3.2.1 Informações fenotípicas

Foram utilizados dados da produção de leite truncada aos 305 dias (PL) e das porcentagens de gordura (%G) e proteína (%P) do leite de búfalas provenientes de doze rebanhos brasileiros de búfalos da raça Murrah, localizados nos Estados de São Paulo e Rio Grande do Norte. O banco de informações continha 10.507 lactações de 3.355 búfalas controladas mensalmente entre os anos de 1995 e 2014. Um total de 15.495 animais estava presente na matriz de parentesco.

Foram consideradas informações de controles obtidos a partir do quinto dia de produção, sendo mantidas apenas as lactações com duração superior a 90 dias. As %G e %P foram obtidas através da média dos registros mensais por lactação. Os grupos de contemporâneos (GC) foram formados de acordo com o rebanho, ano e estação de parto (Outubro-Março e Abril-Setembro), com a restrição de que cada GC deveria conter no mínimo quatro animais, assim como registro das características entre ± 3 desvios padrão da média do grupo. As estatísticas descritivas dos dados estão presentes na tabela 1.

Tabela 1. Estatísticas descritivas para a produção de leite (PL) e para as porcentagens de gordura (%G) e proteína (%P).

Característica	N	Média	Desvio padrão	GC
PL (Kg)	10.507	2012,80	697,54	169
%G	4.545	6,63	1,01	49
%P	4.542	4,29	0,29	49

N = número, GC = Grupos de contemporâneos

Um modelo animal de repetibilidade foi utilizado para todas as características. Os componentes de variância (Tabela 2) foram estimados pelo método de Máxima Verossimilhança Restrita, em análises unicaracterísticas, utilizando-se o programa REMLF90 (MISZTAL, 2005). Foram considerados como efeitos fixos os grupos de contemporâneos, como covariável a idade da búfala ao parto (linear e quadrático) e

como efeito aleatório os efeitos genético aditivo, ambiente permanente e residual. A notação matricial do modelo animal utilizado pode ser assim apresentada:

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

Em que y é o vetor da característica observada, X é a matriz de incidência dos efeitos fixos, β é o vetor dos efeitos fixos, Z é a matriz de incidência dos efeitos aleatórios genéticos aditivos, a é o vetor dos efeitos aleatórios genéticos aditivos, W é a matriz de incidência do efeito aleatório de ambiente permanente, p é o vetor do efeito aleatório de ambiente permanente e e é o vetor dos efeitos aleatórios residuais.

Tabela 2. Variância genética aditiva (σ_a^2), ambiente permanente (σ_{pe}^2), residual (σ_e^2), herdabilidade (h^2) e repetibilidade (r) para as características produção de leite (PL) e as porcentagens de gordura (%G) e proteína (%P).

Característica	σ_a^2	σ_{pe}^2	σ_e^2	h^2	r
PL	75.670	87.550	184.900	0,22	0,47
%G	0,1666	0,1817	0,4447	0,21	0,44
%P	0,0153	0,0177	0,0389	0,21	0,46

3.2.2 Genotipagem e controle de qualidade

Os animais foram genotipados utilizando o painel 90 K Axiom® Buffalo Genotyping (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 p-value < 10^{-6} . Os SNPs coincidentes também foram eliminados. Em relação à amostra foi considerado call rate < 0,90. Após os critérios adotados, o número de animais genotipados foi 322 (289 fêmeas e 33 machos) e um total de 45.330 SNPs foi considerado para o presente estudo.

3.2.3 Análise de associação genômica ampla utilizando informações de animais genotipados e não genotipados (ssGWAS)

Os valores genéticos genômicos (GEBVs) foram obtidos através do programa BLUPF90 (MISZTAL et al., 2002), aplicando-se um modelo animal unicaracterística considerando como efeitos fixos os grupos de contemporâneos, como covariável a

idade da búfala ao parto (linear e quadrático) e como efeito aleatório os efeitos genético aditivo, ambiente permanente e residual. O modelo geral utilizado para as características pode ser descrito da seguinte forma matricial:

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

Em que y é o vetor das observações para cada característica, X é a matriz de incidência dos efeitos fixos, β é o vetor dos efeitos fixos, Z é a matriz de incidência dos efeitos aleatórios genéticos aditivos, a é o vetor dos efeitos aleatórios genéticos aditivos, assumindo-se $a \sim N(0, H\sigma_a^2)$, onde H é a matriz que combina as matrizes de parentesco baseada no pedigree e a matriz de parentesco genômica e σ_a^2 é a variância genética aditiva, W é a matriz de incidência do efeito aleatório de ambiente permanente, p é o vetor do efeito aleatório de ambiente permanente e e é o vetor dos efeitos aleatórios residuais.

A combinação da matriz de parentesco tradicional (\mathbf{A}) com a matriz de parentesco genômico (\mathbf{G}) deu origem a uma nova matriz chamada de \mathbf{H}^{-1} (Aguilar et al., 2010):

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

Onde \mathbf{A}^{-1} , \mathbf{G}^{-1} e \mathbf{A}_{22}^{-1} são, respectivamente, a inversa da matriz de parentesco tradicional, a inversa da matriz genômica e a inversa da matriz de parentesco baseada no pedigree para animais genotipados.

A matriz genômica foi obtida de acordo com Vanraden (2008):

$$\mathbf{G} = \frac{\mathbf{Z}\mathbf{Z}'}{2 \sum_{j=1}^m p_j(1-p)}$$

Em que \mathbf{Z} é a subtração de $\mathbf{M} - \mathbf{P}$, no qual \mathbf{M} é a matriz de genótipos sendo as colunas representada pelos marcadores e linhas os animais, e \mathbf{P} é a matriz de frequência do segundo alelo p_j , expresso como $2p_j$.

Os efeitos dos SNPs foram obtidos a partir dos valores genéticos genômicos (GEBVs) dos animais genotipados de forma iterativa utilizando o programa postGSf90, pertencente à família BLUPF90 (MISZTAL et al., 2002). Seguido pelos programas preGSf90 e BLUPF90, o postGSf90 calcula o efeito dos SNPs, como descrito por Wang et al. (2012), utilizando diferentes pesos na matriz de relacionamento genômico

e portanto, permite o uso de diferentes ponderadores para os SNPs. A equação para calcular os efeitos dos SNPs pode ser demonstrada em forma matricial como:

$$\hat{\mathbf{u}} = \mathbf{DZ}'[\mathbf{ZDZ}']^{-1}\hat{\mathbf{a}}_g$$

Em que $\hat{\mathbf{u}}$ é o vetor de efeito de cada SNP, \mathbf{D} é a matriz diagonal contendo ponderadores para o efeito dos SNPs, \mathbf{Z} é a matriz de genótipos, $\hat{\mathbf{a}}_g$ é o vetor dos valores genéticos preditos para os animais genotipados.

No presente estudo foram realizadas duas iterações para estimar os efeitos dos SNPs. Na primeira, assumiram-se ponderadores iguais a um, os quais em seguida foram calculados em função do quadrado dos efeitos dos SNPs e das frequências alélicas, e em seguida utilizados na segunda iteração. De acordo com Wang et al. (2014) a realização de ponderação é importante para evidenciar as regiões de maior efeito sobre uma característica, ou seja, aumentando-se os pesos atribuído aos SNPs de maior efeito e diminuindo-se os pesos atribuído aos SNPs de menor efeito.

A variância de cada SNP foi calculada multiplicando-se o efeito ao quadrado do SNP i (\hat{u}_i^2) por $2p_iq_i$, onde p_i é a frequência do segundo alelo do SNP i e q_i é $(1 - p_i)$ (ZHANG et al., 2010). A porcentagem de variância genética explicada por cada SNP foi calculada de acordo com Wang et al. (2014):

$$\frac{Var(a_i)}{\sigma_a^2} \times 100\%$$

Onde a_i é o valor genético de cada região com apenas um SNP e σ_a^2 é a variância genética total.

3.2.4 Atribuição dos SNPs aos genes

Foi utilizado um chip específico para búfalos nas análises do ssGWAS. Embora a seleção dos SNPs incluídos no chip seja baseada em dados de sequenciamento dos búfalos, a posição dos SNPs e os genes anotados utilizam como referência o genoma bovino, uma vez que não há disponível, até o momento, referência de genoma completo público para esta espécie (DE CAMARGO et al., 2015). Por isso, foi utilizado o UMD 3.1 *bovine genome sequence assembly* (ZIMIN et al., 2009) para atribuir o SNP ao gene.

Os SNPs foram atribuídos aos genes se sua localização estivesse contida na sequência genômica do referido gene ou se sua localização estivesse dentro de 20

kbp da 5' e 3' finais do primeiro e último éxon, respectivamente. Esta distância de 20 kbp foi utilizada com intuito de capturar regiões reguladoras e funcionais, que estivessem próximas aos genes, mas não necessariamente contidas neles (PEÑAGARICANO et al., 2013).

3.3 RESULTADOS E DISCUSSÃO

Os resultados do presente trabalho indicaram SNPs mais relevantes, quanto à proporção da variância genética explicada, para as três características estudadas (Figura 1). Os dez SNPs de maiores efeitos para a PL, %G e %P explicaram 7,48, 9,94 e 6,56% da proporção da variância genética, respectivamente. Tais SNPs estão localizados nos cromossomos 6, 7, 16, 17, 20 e 29 para PL, 1, 5, 6, 17, 18 e 26 para %G e 1, 2, 6, 8, 9, 15, 26 e 27 para %P (Tabela 3).

Os SNPs que mais explicaram a variância genética aditiva para as características não são os mesmos dos reportados por lamartino et al. (2013), os quais trabalharam com a produção de leite de búfalas na Itália, e de Camargo et al. (2015), trabalhando com o mesmo banco de dados. Atribui-se a esses resultados divergentes, o fato de que as metodologias estatísticas utilizadas não foram as mesmas e com isso, diferentes regiões no genoma e novos genes candidatos foram relacionadas com as características analisadas no presente estudo.

Para a característica de produção de leite, três genes da mesma família foram identificados (*ADAMTS12*, *ADAMTS8* e *ADAMTS15*). Esses genes participam do processo de angiogênese (KUMAR et al., 2012). A relação da angiogênese com a produção de leite vem do fato do leite ser um derivado de sangue, portanto o aumento da vascularização dos alvéolos do úbere está potencialmente relacionado a uma maior produção de leite.

O gene *KIRREL3* também associado à produção de leite, não tem uma relação direta com a característica, todavia já foi associado a características de peso e carcaça em bovinos (SAATCHI et al., 2014), podendo ser um possível efeito de pleiotropia. Sua função está inicialmente descrita na formação de sinapses neurais, exercendo função biológica que afeta o metabolismo de maneira geral.

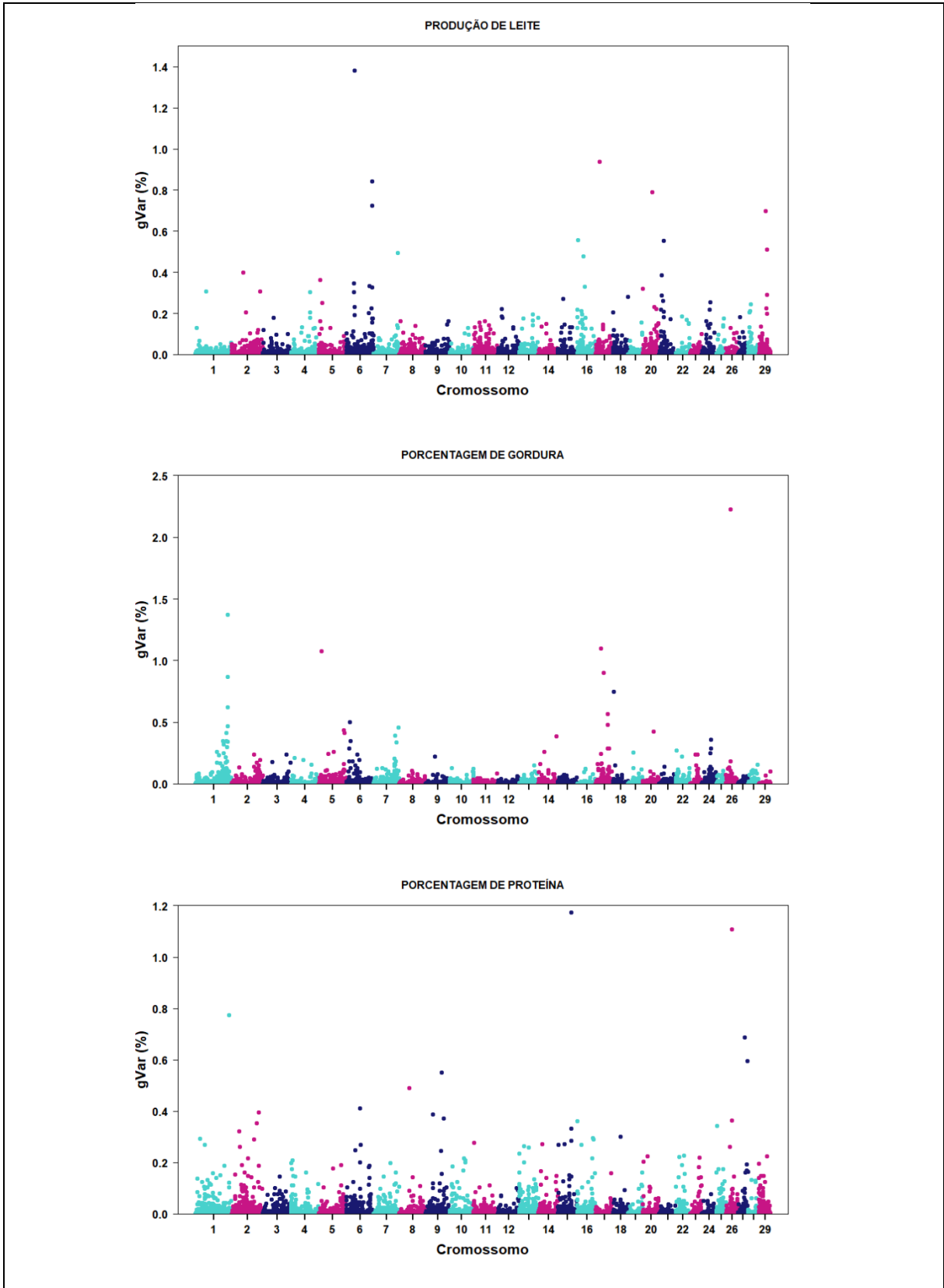


Figura 1. Manhattan plot com a proporção da variância genética explicada por cada SNP para as características estudadas (Chromosome: UMD 3.1 *Bos taurus*).

Tabela 3. Cromossomo (Chr), posição, identificação do gene e proporção da variância genética explicada (%Var) dos top 10 SNPs para as características estudadas.

Chr*	Posição* (pb)	Gene	%Var
Produção de leite			
6	36639835	LOC104972721, LOC104972720	1,38
17	18186434	LOC104974583, MAML3	0,93
6	113636698	RAB28, NKX3-2, BOD1L1	0,84
20	40314400	LOC104975280, TRNAT-UGU, ADAMTS12, LOC104975279, LOC104975281, TARS, LOC104975282	0,79
6	113316913	RAB28, NKX3-2, BOD1L1	0,72
29	30515360	KIRREL3	0,69
16	3875685	FAM72A, SRGAP2	0,56
21	19898813	MRPL46, MRPS11, DET1, MIR1179, MIR7-1, AEN	0,55
29	37119604	ZBTB44, ADAMTS8, ADAMTS15	0,51
7	104954506	LOC101906022, NUDT12, LOC101905838	0,49
Porcentagem de gordura			
26	20766959	ABCC2, DNMBP, CPN1	2,22
1	141426779	IGSF5, PCP4	1,37
17	22180650	TRNAF-AAA, LOC104974590, LOC101908032	1,09
5	12789423	TMTC2, LOC104972347	1,07
17	33099831	LOC101901944	0,90
1	141549126	PCP4	0,86
18	4308813	LOC104974726, LOC104974747, LOC101905587	0,74
1	141349261	B3GALT5, LOC101905982, IGSF5	0,62
17	52383298	ABCC4, LOC104974630	0,56
6	17495052	COL25A1	0,50
Porcentagem de proteína			
15	59208716	BDNF, LOC104974283	1,17
26	27650119	LOC104975989	1,11
1	146701480	AGPAT3, TRAPPC10	0,77
27	30794748	UNC5D	0,69
27	43266954	LOC100300804	0,60
9	72343818	EYA4	0,55
8	43849701	LOC104969341, DMRT2, DMRT3, DMRT1	0,49
6	61515887	APBB2	0,41
2	118172022	PID1, DNER	0,39
9	32997296	LOC101907108, CEP85L, TRNAV-CAC, LOC782839, LOC104969521, LOC104969522	0,39

*Chromosome: UMD 3.1 *Bos taurus*

Para a característica de porcentagem de gordura, foi identificado o gene *ABCC2* que atua no transporte hepatobiliar e por consequência no metabolismo de lipídeos, visto que a bile atua como agente emulsificador de gordura no trato gastrointestinal. A expressão do gene foi aumentada no fígado de humanos submetidos a dietas com alta quantidade de gordura (GHONEIM et al., 2015) e, portanto, possui ação biológica estreita com a característica em estudo.

Outro gene também identificado para porcentagem de gordura é o *ABCC4*. Esse gene foi associado anteriormente com número de serviços por concepção em búfalas (DE CAMARGO et al., 2015), apresentando efeitos tanto em características produtivas como reprodutivas. Em bovinos da raça Angus, o gene foi relacionado com conversão alimentar residual (CHEN et al., 2012). A ingestão de alimentos afeta o metabolismo de maneira holística e está diretamente associada com a deposição e utilização de gordura, visto que o excesso de energia ingerido gera a formação de tecido adiposo e impacta na gordura no leite. Em bovinos da raça Nelore, foram identificados CNVs no gene associados com características de marmoreio na carne (FEITOSA et al., 2014).

Para a característica de porcentagem de proteína, o gene *BDNF* foi identificado. Ele já havia sido associado anteriormente com a mesma característica em bovinos (ZIELKE et al., 2011), revelando que alguns genes podem atuar em mesma ordem de importância em espécies diferentes.

Outros três genes *AGPAT3*, *DNER*, e *MNRT2*, associados com a mesma característica, foram relacionados com outras características em bovinos. O *AGPAT3* está relacionado à produção de ácidos graxos (LI et al., 2014), o *DNER* com marmoreio de carne (CLARK et al., 2011) e o *MNRT2* com ganho de peso residual (SANTANA et al., 2015). Este fato demonstra que os genes desempenham papéis múltiplos no organismo e que as associações, antes de qualquer coisa, indicam a existência de variação genética em várias espécies/raças.

3.4 CONCLUSÃO

Os resultados do ssGWAS indicaram regiões genômicas relacionadas à produção de leite e às porcentagens de gordura e proteína do leite de búfalas. O estudo também identificou que tais regiões cromossômicas, que mais explicaram a variância genética aditiva, abrigam genes candidatos com funções biológicas que podem estar relacionadas diretamente às características estudadas.

3.5 REFERÊNCIAS BIBLIOGRÁFICAS

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**, n. 93, p. 743-752, 2010.

CHEN, Y. et al. Using gene expression information obtained by quantitative real-time PCR to evaluate Angus bulls divergently selected for feed efficiency. **Animal Production Science**, v. 52, n. 11, p. 1058–1067, 2012.

CLARK, D. L. et al. Muscle gene expression associated with increased marbling in beef cattle. **Animal biotechnology**, v. 22, n. 2, p. 51-63, 2011.

DE CAMARGO, G. M. F. . M. F. et al. Prospecting major genes in dairy buffaloes. **BMC Genomics**, v. 16, n. 1, p. 872, 2015.

EL-HALAWANY, N. et al. Genome-wide association study for milk production in Egyptian buffalo. **Livestock Science**, v. 198, p. 10-16, 2017.

FEITOSA, L. B. F. et al. Genome wide association study between copy number variation regions with marbling score in Nelore cattle. In: 10th World Congress of Genetics Applied to Livestock Production, 10, 2014, Vancouver.

GHONEIM, R. H. et al. Effect of a high-fat diet on the hepatic expression of nuclear receptors and their target genes: relevance to drug disposition. **The British journal of nutrition**, v. 113, n. 3, p. 507–16, 2015.

IAMARTINO, D. et al. The buffalo genome and the application of genomics in animal management and improvement. **Buffalo Bulletin**, v. 32, n. SPEC. ISSUE 1, p. 151–158, 2013.

INSTITUTO FUNDAÇÃO BRASILEIRA DE GEOGRAFIA E ESTATÍSTICA (IBGE). Efetivos dos rebanhos em 31.12 e variação anual segundo as categorias. Disponível em: <ftp://ftp.ibge.gov.br/Producao_Pecuaria/Producao_da_Pecuaria_Municipal/2011/tabelas_pdf/tab01.pdf> Acesso em: 25 nov. 2016.

KUMAR, S.; RAO, N.; GE, R. Emerging roles of ADAMTSS in angiogenesis and cancer. **Cancers**, v. 4, n. 4, p. 1252–1299, 2012.

LI, C. et al. Genome wide association study identifies 20 novel promising genes associated with milk fatty acid traits in Chinese Holstein. **PLoS ONE**, v. 9, n. 5, 2014.

MEIGNANALAKSHMI, S.; MAHALINGA NAINAR, A. PCR-RFLP analysis of beta-lactoglobulin gene in Murrah buffaloes. **Tamilnadu J. Veterinary & Animal Sciences**, v. 5, n. 5, p. 194-197, 2009.

MISZTAL, I. [2005]. REMLF90 manual. Disponível em: <<ftp://nce.ads.uga.edu/pub/ignacy/blupf90/>>. Acesso em: 11/10/2016.

MISZTAL, I., et al. "BLUPF90 family of programs (2002)." *University of Georgia*. <http://nce.ads.uga.edu/~ignacy/numpub/blupf90/>, Acesso em 30 de nov. 2016.

MISZTAL, I.; LEGARRA, A.; AGUILAR, I. Computer procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. **Journal of Dairy Science**, n.42, p. 4648-4655, 2009.

OTAVIANO, A.R. et al. Kappa-casein gene study with molecular markers in female buffaloes (*Bubalus bubalis*). **Genetics and Molecular Biology**, v. 28, n. 2, p. 237-241, 2005.

PEÑAGARICANO, F. et al. Inferring quantitative trait pathways associated with bull fertility from a genome-wide association study. **Frontiers in Genetics**, v. 3, n. JAN, p. 1–7, 2013.

RAMESHA, K.P. et al. Alpha-Lactalbumin Gene Polymorphism: A preliminary study on two breeds of the river Buffalo (*Bubalus bubalis*). **AsPac J. Mol. Biol. Biotechnol.**, v. 16, n. 2, p. 47-52, 2008.

RIAZ, M.N. et al. Molecular marker assisted study of kappa-casein gene in Nili-Ravi (Buffalo) breed of Pakistan. **Pakistan Vet. J.**, v. 28, n. 3, p. 103-106, 2008.

SAATCHI, M. et al. Large-effect pleiotropic or closely linked QTL segregate within and across ten US cattle breeds. **BMC genomics**, v. 15, n. 1, p. 442, 2014.

SANTANA, M. H. A. et al. Genome-wide association with residual body weight gain in *Bos indicus* cattle. **Genetics and Molecular Research**, v. 14, n. 2, p. 5229–5233, 2015.

TONHATI, H. et al. Programa de mejoramiento genético de búfalos. *Revista Argentina de Producción Animal*, v. 28, n. 1, p. 53-67, 2008.

UTSUNOMIYA, Y. T. et al. Genome-wide mapping of loci explaining variance in scrotal circumference in Nelore cattle. **Plos One**, v. 9, n. 2, 2013.

VANRADEN, P.M. Efficient Methods to Compute Genomic Predictions. **Journal of Dairy Science**, 91, p. 4414–4423, 2008.

WANG, H. et al. Genome-wide association mapping including phenotypes from relatives without genotypes. **Genetics Research**, v.94, n.2, p.73-83, 2012.

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, 2014.

ZHANG, Z. Best linear unbiased prediction of genomic breeding values using a trait-specific marker-derived relationship matrix. **PLoS One**, v. 5, 2010.

ZIELKE, L. G. et al. BDNF contributes to the genetic variance of milk fat yield in German Holstein cattle. **Frontiers in Genetics**, v. 2, n. APR, p. 1–8, 2011.

ZIMIN, A. V. et al. A whole-genome assembly of the domestic cow, *Bos Taurus*. **Genome Biol.**, v. 10, 2009.

CAPÍTULO 4 - Considerações Finais

Os búfalos chegaram ao Brasil pela Ilha de Marajó a pouco mais de um século e adaptaram-se tão bem as diversas condições climáticas do país que, atualmente, podem ser encontrados em quase toda a sua extensão territorial, se destacando como importantes fontes de carne, leite e trabalho.

Após anos de permanência no Brasil, os búfalos alcançaram recente destaque no mercado de produtos lácteos em virtude da qualidade do seu leite. Esses produtos apresentam sabor singular e encontraram lugar especial na culinária brasileira. A crescente demanda por produtos da bubalinocultura, dando ênfase para o queijo tipo *mozzarella*, tem contribuído significativamente para o crescimento dessa cadeia produtiva.

A atividade tem se tornado mais atrativa para os produtores, pois a rentabilidade é melhor quando comparada à bovinocultura de leite. Além disso, a adaptação desses animais aos mais variados climas brasileiros, possibilitou a escolha dos criadores pela atividade da criação de búfalos, ao invés da criação de bovinos.

Todos esses fatores têm contribuído com o considerável crescimento do rebanho bubalino brasileiro. Entretanto, apenas o crescimento não é suficiente para atender essa demanda em alta e para a consolidação dessa atividade pecuária. Os rebanhos apresentam baixos índices produtivos. A melhoria da produtividade desses rebanhos se faz necessária para alavancar o setor. Para isto é indispensável o aprimoramento nas técnicas de manejo e melhoramento genético dos animais. Como na maioria das atividades pecuárias, com o desenvolvimento e aplicação de técnicas ajustadas no setor é possível incrementar os índices produtivos praticados atualmente.

Dentro desse contexto, a predição genômica pode contribuir de forma positiva para a seleção de animais superiores e, portanto, aumentar os índices produtivos da cadeia da bubalinocultura leiteira. Tais informações ainda são bastante escassas no Brasil, uma vez que se trata de uma atividade relativamente recente no país. Além do mais, como a bubalinocultura é uma atividade mais presente em países em desenvolvimento e poucas são as tecnologias próprias para a espécie, também são

poucas as informações na literatura mundial a respeito da seleção genômica de búfalos.

Esse fato vem reforçar a importância desse trabalho, que teve como principal objetivo avaliar os modelos bayesianos quanto as suas capacidades de predição dos valores genéticos genômicos para as características produção e qualidade do leite de búfalas e identificar regiões cromossômicas relacionadas às mesmas.

Os resultados demonstraram informações relevantes quanto a aplicação da seleção genômica de búfalas leiteiras. Pode-se verificar que os seguintes métodos bayesianos: Bayes A, Bayes B, Bayes C e Bayes LASSO podem ser utilizados para a predição dos efeitos dos marcadores das características produção de leite e as porcentagens de gordura e proteína no leite de búfalas, obtendo-se praticamente os mesmos valores de acurácia de predição. Tais resultados similares entre os modelos se deve, provavelmente, ao reduzido número de animais avaliados e à natureza complexa das características estudadas.

O estudo de associação genômica ampla apontou regiões cromossômicas relacionadas à produção e qualidade do leite de búfalas. Com os resultados do ssGWAS, foi possível identificar também que tais regiões genômicas abrigam genes candidatos com funções biológicas que podem estar relacionadas às características estudadas. As regiões e genes identificados neste estudo, contribuem para o melhor entendimento do controle genético destas características e possibilitam a investigação de mutações causais em estudos posteriores de bubalinos leiteiros.

Um incremento nas acurácias de predição é esperado com a adição de mais animais genotipados nos estudos de predição genômica das características de produção e qualidade do leite bubalino. Com isso, dado a importância da seleção assistida por marcadores moleculares, fazem-se necessários mais estudos com a espécie bubalina, para que, assim como na bovinocultura leiteira, as ferramentas genômicas possam alavancar a produção dos rebanhos bubalinos no Brasil.