

UNIVERSIDADE FEDERAL DE MATO GROSSO DO SUL
FACULDADE DE MEDICINA VETERINÁRIA E ZOOTECNIA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA ANIMAL

**ESTUDO DE ASSOCIAÇÃO GENÔMICA AMPLA PARA
CARACTERÍSTICAS DE PRODUÇÃO EM BOVINOS DA RAÇA
CANCHIM**

***GENOME WIDE ASSOCIATION STUDY FOR PRODUCTION
TRAITS IN CANCHIM BREED CATTLE***

Gustavo Garcia Santiago

CAMPO GRANDE
MATO GROSSO DO SUL - BRASIL

2015

UNIVERSIDADE FEDERAL DE MATO GROSSO DO SUL
FACULDADE DE MEDICINA VETERINÁRIA E ZOOTECNIA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA ANIMAL

**ESTUDO DE ASSOCIAÇÃO GENÔMICA AMPLA PARA
CARACTERÍSTICAS DE PRODUÇÃO EM BOVINOS DA RAÇA
CANCHIM**

***GENOME WIDE ASSOCIATION STUDY FOR PRODUCTION
TRAITS IN CANCHIM BREED CATTLE***

Gustavo Garcia Santiago

Orientador: PhD. Roberto Augusto de Almeida Torres Júnior

Co-orientadora: Dra. Fabiane Siqueira

Dissertação apresentada à Universidade Federal de
Mato Grosso do Sul, como requisito à obtenção do
título de Mestre em Ciência Animal. Área de
Concentração: Produção Animal

CAMPO GRANDE
MATO GROSSO DO SUL - BRASIL

2015

Certificado de aprovação

GUSTAVO GARCIA SANTIAGO

**Estudo de associação genômica para características de produção em bovinos da raça
Canchim**

Genome wide association study for production traits in Canchim cattle breed

Dissertação apresentada à
Universidade Federal de Mato Grosso
do Sul, como requisito à obtenção do
título de mestre em Ciência Animal.

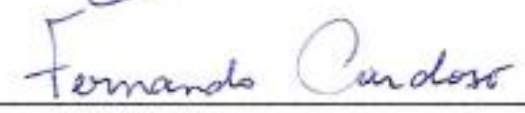
Área de concentração: Produção
Animal.

Aprovado(a) em: 27/02/2015

BANCA EXAMINADORA:



Dr. Roberto Augusto de Almeida Torres Júnior
(EMBRAPA CNPCG) – (Orientador)



Dr. Fernando Flores Cardoso
EMBRAPA CPPSUL



Dr. Gilberto Romeiro de Oliveira Menezes
EMBRAPA CNPGC

**À minha família, pelo amor, apoio incondicional, força, incentivo
e compreensão em todos os momentos de minha vida.**

Sem eles nada disto seria possível.

Dedico

AGRADECIMENTOS

Ao Dr. Roberto Augusto de Almeida Torres Júnior, pela orientação, oportunidade, exemplo profissional e por acreditar em meu trabalho.

À Dra. Fabiane Siqueira, pela orientação, ensinamentos, amizade e pelo incentivo desde o início de minha vida acadêmica.

Ao Dr. Fernando Flores Cardoso e a equipe do Laboratório de Bioinformática e Estatística Genômica (Leandro, Bruno, Rodrigo Costa, Bruna, Henry, Ândrea, Marcos, Rodrigo Azambuja, Willian, Eduardo, Luiza e Helena) por me receber em seu laboratório, pelos ensinamentos e por toda ajuda dedicada ao nosso trabalho.

Ao Dr. Gilberto Romeiro de Oliveira Menezes, a Dra. Andréa Alves do Egito, ao Dr. Gelson Luís Dias Feijó, a Dra. Anna Beatriz Robottom Ferreira e ao Dr. André Luiz Julien Ferraz pelos ensinamentos, sugestões, conselhos e incentivos.

Aos colegas do Laboratório de Genômica e Melhoramento Animal da Embrapa Gado de Corte (Isabella, Isadora, Catherine, Thalles, Paula e Franciele) pela amizade e companheirismo.

Ao Dr. José Fernando Garcia e sua equipe (Yuri, Ludmila, Anirene, Adam, Solomon, Pier) por me proporcionarem um primeiro contato fascinante com a bioinformática.

À Embrapa Gado de Corte, pela oportunidade de realizar este trabalho e por ser minha segunda casa.

À UFMS e aos professores do curso de Pós-Graduação em Ciência Animal, pelos ensinamentos.

À Associação Brasileira dos Criadores de Canchim e a equipe do programa Geneplus/Embrapa, especialmente ao médico veterinário Maury Dorta de Souza Junior, responsável técnico pelas Provas Canchim de Avaliação de Desempenho.

À Capes pela bolsa de estudos e à Embrapa pelo financiamento do projeto.

SUMÁRIO

	“Página”
1. INTRODUÇÃO.....	02
2. A RAÇA CANCHIM.....	04
3. PROVA CANCHIM DE AVALIAÇÃO DE DESEMPENHO (PCAD).....	05
4. CARACTERÍSTICAS DE CRESCIMENTO E CARCAÇA AVALIADAS NESTE TRABALHO.....	06
4.1. Peso ao sobreano.....	06
4.2. Área de olho de lombo.....	07
4.3. Espessura de gordura subcutânea.....	08
5. ESTUDOS DE ASSOCIAÇÃO AMPLA DO GENOMA (GWAS).....	08
5.1. Controle de qualidade.....	10
5.1.1. Fenótipos.....	10
5.1.2. Controle de qualidade dos SNPs.....	11
5.1.2.1. <i>Gen Train</i> e <i>Genotype Calling Score</i>.....	11
5.1.2.2. <i>Call Rate</i> do SNP.....	12
5.1.2.3. Equilíbrio de Hardy-Weinberg.....	12
5.1.2.4. Menor frequência alélica.....	12
5.1.3. Controle de qualidade das amostras.....	13
5.1.3.1. <i>Call Rate</i> da amostra.....	13
5.1.3.2. Heterozigose.....	13
5.1.3.3. Estratificação populacional.....	13
5.2. Métodos de estimação bayesiana.....	14
5.2.1. Bayes B.....	15
5.2.2. Bayes C.....	16

5.2.3. Bayes $C\pi$	17
6. IMPUTAÇÃO DE GENÓTIPOS.....	17
7. HIPÓTESES A SEREM TESTADAS.....	19
8. OBJETIVOS.....	19
8.1. Objetivos gerais.....	19
8.2. Objetivos específicos.....	19
9. REFERÊNCIAS.....	20
 CAPÍTULO I - ESTUDOS DE ASSOCIAÇÃO AMPLA DO GENOMA PARA	
ÁREA DE OLHO DE LOMBO, ESPESSURA DE GORDURA SUBCUTÂNEA E	
PESO AO SOBREANO EM BOVINOS DA RAÇA CANCHIM.....	
	30
1. INTRODUÇÃO.....	32
2. MATERIAL E MÉTODOS.....	33
2.1. Dados fenotípicos.....	33
2.2. Estimação de componentes de (co)variância e parâmetros genéticos.....	35
2.3. Dados genômicos.....	35
2.4. Controle de qualidade de dados genômicos.....	36
2.5. Estratificação populacional.....	37
2.6. Imputação de genótipos.....	37
2.7. Estudo de associação ampla do genoma.....	38
2.8. Análise de enriquecimento funcional.....	40
3. RESULTADOS.....	40
4. DISCUSSÃO.....	42
4.1. Área de olho de lombo.....	43
4.2. Espessura de gordura subcutânea.....	44
4.3. Peso ao sobreano.....	45

5. CONCLUSÃO.....	47
6. REFERÊNCIAS.....	48
7. ANEXOS.....	56

RESUMO

Características de crescimento e de carcaça afetam diretamente o valor comercial da carcaça bovina. Estudos de associação ampla do genoma (GWAS) viabilizam a identificação de locos que influenciam a expressão de características quantitativas. Neste contexto, o objetivo desse trabalho foi identificar regiões genômicas e genes associados com características de produção e qualidade de carcaça como área de olho de lombo (AOL), espessura de gordura subcutânea (EGS) e peso ao sobreano (PESO) em bovinos da raça Canchim. O GWAS foi realizado por meio dos métodos Bayes B e Bayes C, utilizando informações fenotípicas e genótipos imputados de 701 animais da raça Canchim e do grupo genético MA (progênie de touro Charolês e vaca $\frac{1}{2}$ Canchim + $\frac{1}{2}$ Zebu). Foram identificadas duas, três e cinco regiões candidatas associadas com AOL, PESO e EGS, respectivamente. A análise funcional destacou genes com funções biológicas relacionadas às características estudadas e relatos prévios de locos controladores de características quantitativas (QTLs), subjacentes ou abrangidos pelas regiões candidatas reforçam as evidências de associação encontradas por este trabalho.

Palavras chave: SNP, carcaça, crescimento, imputação, inferência Bayesiana

1. INTRODUÇÃO

No cenário agropecuário mundial, a bovinocultura brasileira ocupa lugar de destaque apresentando o maior rebanho comercial de gado bovino, com cerca de 199 milhões de cabeças (Anualpec, 2014). Em 2014, o país abateu 43,3 milhões de animais (aproximadamente 8,5 milhões de toneladas de equivalente carcaça) e manteve o título de maior exportador mundial de carne bovina (Anualpec, 2014). Embora seja o líder mundial em quantidade de carne bovina exportada, o Brasil precisa adequar sua produção aos padrões e às exigências estabelecidos pelos importadores, principalmente, no que se refere à qualidade da carne, já que está diretamente relacionada com a aceitabilidade do produto.

Segundo Luchiari Filho (2006), para atender aos nichos de mercado de alta qualidade e valor agregado é imprescindível aplicar uma estratégia que envolva desde a escolha do material genético, a adequação do sistema de produção visando à padronização da oferta de matéria prima (animais) às indústrias, principalmente quanto ao peso, rendimento de cortes cárneos e grau de acabamento.

Durante muito tempo, a composição da remuneração para pecuária de corte nacional foi baseada apenas no peso vivo e no rendimento de carcaça dos animais. Entretanto, observa-se forte tendência de mudança desse cenário, e em muitas situações a receita obtida na atividade está vinculada não somente aos aspectos quantitativos, mas também à qualidade, notadamente, ao acabamento de carcaça. Segundo Perotto et al. (2000), o valor comercial de uma carcaça bovina é determinado por um conjunto de características, destacando-se o peso, a conformação, o rendimento, a gordura de cobertura e o marmoreio. Neste sentido, o melhoramento genético de características de crescimento e de carcaça demonstra-se como uma importante ferramenta para agregar valor ao produto final.

Entre as características de crescimento, o peso ao sobreano destaca-se por permitir avaliar o potencial genético do indivíduo ainda em idade jovem, tendo em vista a menor influência dos efeitos ambientais maternos e apresenta maior correlação com o peso ao abate quando comparado ao peso a desmama e ao ano (Bergman, 1999).

Entre as características de carcaça, a área de olho de lombo (AOL) apresenta correlação com musculosidade e rendimento dos cortes de alto valor comercial, estando relacionada com a porção comestível da carcaça (Luchiari Filho, 2000). A espessura de gordura subcutânea (EGS) ou gordura de cobertura apresenta grande importância no processo de resfriamento das carcaças, isolando-as do frio, protegendo-as contra o encurtamento das fibras e, conseqüentemente, o endurecimento da carne (Meirelles et al., 2010).

No geral, as características de crescimento e de carcaça de bovinos apresentam herdabilidade de magnitude moderada a alta, evidenciando que a seleção e a utilização de raças especializadas nos programas de cruzamento, se constituem como importantes ferramentas para melhoria desses aspectos no rebanho (Forni et al., 2007).

Neste contexto, a raça Canchim destaca-se por proporcionar um ganho de peso na ordem de 15% em relação ao zebu (Tupy et al., 2006). Atualmente, é bastante utilizada em sistemas de produção de carne, principalmente pelo cruzamento de touros Canchim com vacas zebu, devido à capacidade de monta e fertilidade do touro Canchim e ao desempenho, qualidade de carne e carcaça dos seus produtos (Vianna et al., 1978; Tupy et al., 2006).

As análises do genoma, incluindo o mapeamento de QTLs (*Quantitative Trait Loci*) (Soller & Plotkin-Hazan, 1977; Soller, 1978), o emprego em larga escala de SNPs (*Single Nucleotide Polymorphism*) (Meuwissen et al., 2001) e, mais recentemente, a utilização de dados de sequenciamento (Pérez-Enciso et al., 2015) na seleção de bovinos

de corte, permitirão avanços na inclusão de características relacionadas à qualidade da carcaça e da carne nos programas de melhoramento genético.

Com o desenvolvimento de plataformas para a detecção simultânea de milhares de SNPs, tornou-se possível aumentar a confiabilidade nos resultados obtidos por meio de mapeamento de QTLs, empregando marcadores distribuídos densamente por todo o genoma (Faleiro et al., 2011). Essa metodologia tem sido chamada de estudo de associação genômica ampla (*Genome-wide association study* - GWAS).

Os GWAS visam à identificação de SNPs que estão em desequilíbrio de ligação com regiões cromossômicas que controlam a manifestação de fenótipos relevantes (Balding, 2006). Neste contexto, a identificação de genes, proteínas e vias metabólicas envolvidas na manifestação fenotípica poderá auxiliar na compreensão dos mecanismos genéticos que controlam características de crescimento e carcaça em bovinos.

2. A RAÇA CANCHIM

A raça Canchim é uma raça sintética que foi desenvolvida em 1940 na Fazenda Canchim, onde hoje se situa a Embrapa Pecuária Sudeste, em São Carlos/SP, com o objetivo de unir a precocidade, musculabilidade e rendimento de carcaça apresentados pelos animais de origem taurina, à rusticidade e adaptabilidade apresentadas pelos animais de origem zebuína (Vianna et al., 1978, Tupy et al., 2006). Essa raça foi formada por meio de diferentes esquemas de cruzamentos alternados entre animais da raça Charolês, que possui grande potencial para ganho de peso, e de diferentes raças zebuínas, como Indubrasil, Guzerá e Nelore, que se destacam pela capacidade de adaptação às condições tropicais (Vianna et al., 1978).

O fato de ter nascido dentro de uma instituição de pesquisa determinou o sucesso dessa raça, já que sua formação e posterior expansão sempre foram pautadas por

critérios técnicos de seleção. Diversos experimentos foram desenvolvidos desde o princípio para que características de interesse econômico fossem aprimoradas e fixadas na raça Canchim (Vianna et al., 1978; Santos et al., 1996; Silva et al., 2000; Tupy et al., 2006; Buzanskas et al., 2010; Meirelles et al., 2010; Buzanskas et al., 2014).

Com o passar dos anos, o Canchim foi sendo difundido pelo País, ocupando espaço na pecuária nacional. De acordo com a Associação Brasileira dos Criadores de Canchim (<http://www.abccan.com.br>), a raça é criada em nove Estados da Federação e teve cerca de 200.000 animais Canchim e 32.000 animais MA (filhos de touros Charolês e vacas 1/2 Canchim + 1/2 Zebu) nascidos e cadastrados até 2014.

Atualmente, é utilizada em sistemas de produção de carne, principalmente pelo cruzamento de touros Canchim com vacas zebu, devido à capacidade de monta e fertilidade do touro Canchim e ao desempenho, terminação, qualidade de carne e carcaça dos seus produtos (Tupy et al., 2006). Os touros Canchim podem ser utilizados em monta a campo em condições tropicais, fato que proporciona impacto econômico positivo, pois permite aumentar a produção de carne bovina sem aumento nos custos de produção, quando comparados a animais zebuínos (Tupy et al., 2006).

Por ser uma raça sintética, a utilização de diferentes sistemas de acasalamento permite empregar a seleção ocorrida nas suas raças formadoras, além da seleção na própria raça como fator muito importante para o seu desenvolvimento (Barbosa, 2000).

3. PROVA CANCHIM DE AVALIAÇÃO DE DESEMPENHO (PCAD)

A Prova Canchim de Avaliação de Desempenho (PCAD) é realizada pela Associação Brasileira dos Criadores de Canchim (ABCCAN) e Programa Geneplus/EMBRAPA com a finalidade de identificar animais com desempenho superior para utilização como reprodutores na raça ou em cruzamentos. As avaliações

são realizadas anualmente, em duas etapas com o objetivo de avaliar separadamente animais nascidos de junho a agosto dos que nasceram de setembro a novembro.

Os animais entram nas Provas com aproximadamente oito meses de idade, pesando em média 230 quilos, quando são confinados em baia coletiva por um período de 154 dias, sendo 42 dias para adaptação e 112 dias para avaliação. São avaliadas 11 características importantes para produção, qualidade de carcaça, funcionalidade e padrão racial, sendo ganho de peso diário, peso ao final das Provas, área de olho de lombo, espessura de gordura subcutânea, marmoreio, conformação frigorífica, perímetro escrotal, correção de umbigo e prepúcio, qualidade da pelagem, pigmentação de mucosas e aprumos.

As medidas de EGS (mm), AOL (cm²) e marmoreio (escores) são coletadas transversalmente no músculo *Longissimus dorsi* na região entre a 12^a e 13^a costelas com um equipamento de ultrassonografia com transdutor linear de 3,5 MHz ASP 30C/18 cm, específico para avaliação de carcaça.

A orientação técnica, supervisão, condução das avaliações e cálculos para classificação dos animais ao final das Provas são de responsabilidade da equipe de melhoramento animal do programa Geneplus/Embrapa.

4. CARACTERÍSTICAS DE CRESCIMENTO E CARCAÇA AVALIADAS

4.1. Peso ao sobreano

Características de crescimento são tradicionalmente utilizadas como critério de seleção em programas de melhoramento genético de bovinos de corte por serem de fácil mensuração, apresentarem herdabilidades com valores moderados e por estarem associadas com a produtividade (Razook et al., 2001).

Entre essas características, destacam-se os pesos mensurados em diferentes estágios de desenvolvimento do animal, principalmente o peso ao sobreano (18 meses), que permite avaliar o potencial genético do indivíduo ainda em idade jovem, tendo em vista a menor influência dos efeitos ambientais maternos e apresenta maior correlação com o peso ao abate quando comparado ao peso a desmama e ao ano (Bergman, 1999).

Diversos trabalhos demonstram que as estimativas de herdabilidade para peso ao sobreano na raça Canchim apresentam valores moderados, variando entre 0,23 a 0,54 (Mascioli et al., 1996; Castro-Pereira et al., 2007; Buzanskas et al., 2010; Meirelles et al., 2010; Borba et al., 2011). Mascioli et al. (1996), avaliando animais da raça Canchim, encontraram correlações genéticas entre pesos em diferentes idades e peso ao sobreano, o que indica que a seleção para esta característica promove mudança genética nos pesos nas demais idades.

Em outros trabalhos, as estimativas de correlações genéticas indicam que a seleção para peso ao sobreano, proporciona ganhos genéticos também em área de olho de lombo e espessura de gordura subcutânea (Meirelles et al., 2010; Yokoo et al., 2010; Caetano et al., 2013).

4.2. Área de olho de lombo (AOL)

Além das características de crescimento, é fundamental a utilização de características relacionadas à qualidade da carcaça como critérios de seleção em programas de melhoramento genético de bovinos de corte. Neste sentido, a área de olho de lombo é uma medida que apresenta correlações genéticas com medidas de qualidade de carcaça como musculosidade, rendimento de cortes cárneos, composição e peso da carcaça (Hassen et al., 1999; Luchiari Filho, 2000; Sugisawa et al., 2003; Sugisawa et al., 2006).

Esta medida é obtida por ultrassonografia e representa a área da secção transversal do músculo *Longissimus dorsi* entre as 12^a e 13^a costelas. Na literatura são relatadas estimativas de herdabilidade para AOL na raça Canchim variando entre 0,33 a 0,39 (Meirelles et al., 2009; Meirelles et al., 2010; Mokry et al., 2012).

4.3. Espessura de gordura subcutânea (EGS)

Uma carcaça de qualidade deve apresentar quantidade de gordura subcutânea suficiente para isolar a carcaça evitando o encurtamento das fibras musculares pelo frio e de preservar as características organolépticas da carne (Dolezal et al., 1982). Segundo Ferreira et al. (2006), os frigoríficos preconizam que a carcaça apresente no mínimo 3,00 mm de espessura de gordura sobre a 12^a costela. Esta medida também é obtida por ultrassonografia entre as 12^a e 13^a costelas. Na literatura são relatadas estimativas de herdabilidade variando entre 0,04 a 0,28 para EGS em bovinos da raça Canchim (Meirelles et al., 2010; Mokry et al., 2012).

5. ESTUDOS DE ASSOCIAÇÃO AMPLA DO GENÔMA (GWAS)

De acordo com Hey (2004), com exceção de pontos onde ocorre alta taxa de recombinação, o genoma é herdado em blocos de haplótipos formados por desequilíbrio de ligação (*Linkage Disequilibrium* - LD). O LD é uma medida estatística da associação entre alelos de diferentes locos, ou seja, esta medida descreve a associação não aleatória entre dois ou mais locos no mesmo cromossomo (Ardlie et al., 2002). Esta associação ocorre, principalmente, pela proximidade física entre diferentes locos e pode ser influenciada por vários processos evolutivos e acontecimentos históricos da população (Ardlie et al., 2002; Khatkar et al., 2008).

As primeiras análises envolvendo marcadores moleculares possibilitaram grande avanço no estudo de características quantitativas, principalmente na elaboração de mapas genéticos. Esta abordagem ficou conhecida como mapeamento de QTLs e resultou no descobrimento de regiões genômicas associadas à várias características de interesse econômico (Soller & Plotkin-Hazan 1977; Soller 1978).

O conceito de GWAS foi primeiramente utilizado pela comunidade científica envolvida com a genética epidemiológica humana, tendo como objeto de estudo a identificação de variações genéticas e a associação dessas variações com determinadas doenças. Uma vez que tais variações são identificadas, é possível utilizá-las para localizar e entender como os genes contribuem para a manifestação de determinadas doenças e desenvolver diferentes estratégias de prevenção e tratamento (*National Human Genome Research Institute*, 2011).

Portanto, GWAS são estudos de marcadores moleculares do tipo SNPs que estão em desequilíbrio de ligação com mutações causais e que visam determinar os efeitos genéticos sobre os fenótipos dos indivíduos de uma população (Balding, 2006).

Após a conclusão do Projeto Genoma Bovino (*The Bovine HapMap Consortium*, 2009; *The Bovine Genome Sequencing and Analysis Consortium*, 2009) e o desenvolvimento de plataformas de genotipagem de SNPs em alta densidade ao longo de todo genoma (Matukumalli et al., 2009), tornou-se possível averiguar grande parte da variação genética que influencia a manifestação de características de interesse econômico em bovinos por meio dos GWAS.

Essa abordagem exige alguns elementos essenciais para ser realizada com sucesso. Primeiramente, é necessária a utilização de um grande tamanho amostral e que os indivíduos forneçam informações genéticas que efetivamente representem o fenótipo estudado (Meuwissen et al., 2001; Ziegler et al., 2008; Cantor et al., 2010).

A utilização de um tamanho amostral grande torna essa metodologia bastante dispendiosa. Como alternativa, a imputação de genótipos se destaca por permitir prever genótipos de marcadores faltantes em uma amostra com base em haplótipos presentes na população ou em indivíduos aparentados (Li et al., 2009).

Devido ao grande número de SNPs e amostras utilizadas nos GWAS, pequenos erros de determinação dos genótipos podem resultar em níveis inaceitáveis de erros do tipo I (falso-positivo) e II (falso-negativo) (Carmo, 2012).

Dessa forma, a qualidade dos dados é um dos fatores decisivos que definem a validade dos resultados e das conclusões. Até mesmo uma baixa taxa de erro nos genótipos pode resultar em um nível inaceitável de associações falsas. Então, o delineamento do controle de qualidade dos dados é baseado em atributos que objetivam refletir a utilidade e a integridade dos genótipos. Assim sendo, a seguir são apresentados alguns procedimentos de controle de qualidade que podem ser aplicados para reduzir as taxas de erro.

5.1. Controle de qualidade

5.1.1. Fenótipos

Para facilitar a visualização e interpretação dos dados, é adequado criar gráficos da distribuição de um fenótipo em um histograma ou um gráfico de barra, de acordo com a natureza do fenótipo (numérico ou categórico). Em uma simples observação dos dados pode-se identificar os indivíduos com fenótipos extremos. Estes podem ser indivíduos verdadeiramente incomuns, mas eles também podem indicar erros na coleta de informações. Portanto, é prudente averiguar se esses valores são fenótipos verdadeiros ou erros cometidos durante a coleta dos dados. Problemas sistemáticos na coleta de informações fenotípicas como desvios nas medições ao longo do tempo ou

entre lotes são de grande preocupação, pois indica uma importante fonte de erro (Broman & Sen, 2009).

5.1.2. Controle de qualidade dos SNPs

5.1.2.1. *Gen Train* e *Genotype Calling Score*

Na tecnologia Illumina®, a designação do genótipo dos SNPs é determinada pela intensidade de fluorescência emitida pela reação físico-química que ocorre no sítio de hibridização específico de cada SNP. Para tanto, um *software* chamado GenomeStudio®, contém uma série de algoritmos que normalizam as intensidades observadas nos dois canais de cores emitidas, além das interferências de “*background*” e as possíveis sobreposições entre os canais. Dependendo das formas e das distâncias entre os aglomerados, é atribuída uma pontuação estatística (*Gen Train Score*), que será utilizada na formação de três possíveis grupos que representam os genótipos hipotéticos AA, AB e BB (Ritchie et al., 2011).

Na determinação dos genótipos é estabelecida uma pontuação “*Genotype Calling Score*” (GC *score*) que consiste na determinação do valor de confiança de cada genótipo. Essa pontuação dependerá da distância do SNP do centro do grupo no qual ele foi designado. Este parâmetro pode ser usado como padrão de qualidade filtrando genótipos atribuídos com baixa confiabilidade. Os valores de GC variam entre 0 e 1, sendo que quanto mais próximo a 1, maior é a confiança ao genótipo atribuído (Ritchie et al., 2011).

5.1.2.2. *Call Rate* do SNP

Os marcadores que apresentarem valores abaixo do limiar estabelecido para GC *score* são determinados como “*missing*”. O *Call Rate* do SNP é definido como a razão

entre o número de genótipos com *GC Score* superior ao limiar estabelecido e o número total de amostras genotipadas para determinado marcador (Turner et al., 2011) e é um indicador da qualidade da genotipagem.

5.1.2.3 Equilíbrio de Hardy-Weinberg

O equilíbrio de Hardy-Weinberg preconiza que os genótipos de um loco bi alélico AA, AB e BB ocorram com frequências relativas de p^2 , $2pq$ e q^2 respectivamente, onde p refere-se a frequência do alelo A e $q = 1 - p$ (Hardy, 1908; Weinberg, 1908). Desvios das proporções de Hardy-Weinberg podem ser indicativos de possíveis erros de genotipagem, estratificação da população, endogamia e marcas de seleção (Salanti et al., 2005; Turner et al., 2011) o que tem levado este parâmetro a ser bastante utilizado no controle de qualidade dos SNPs.

5.1.2.4. Menor frequência alélica

A “*Minor Allele Frequency*” (MAF) refere-se à frequência à qual o alelo menos comum ocorre numa dada população. Geralmente, alelos com frequência inferior a 2% na população são excluídos das análises posteriores, a fim de se aumentar o poder de detecção da associação genótipo-fenótipo. Este parâmetro é utilizado para diferenciar polimorfismos comuns de variantes raras e erros de genotipagem (Carmo, 2012). Dickson et al. (2011) afirmaram que as variantes raras podem dar origem a taxas inaceitáveis de falso-positivo em GWAS.

5.1.3. Controle de qualidade das amostras

5.1.3.1. *Call Rate* da amostra

É muito comum observar variação na qualidade e na quantidade de DNA de um grupo amostral grande. A baixa qualidade ou baixa concentração de DNA podem ocasionar falhas na determinação dos genótipos, reduzindo o *GC score* e aumentando o numero de genótipos *missing*.

A proporção total de genótipos não designados de uma amostra consiste no *Call Rate* das amostras. Amostras com baixa eficiência de genotipagem ou baixo *Call Rate* devem ser removidas das análises posteriores. O limiar recomendado varia entre 95% a 99% para *Call Rate* da amostra (Tuner et al., 2011).

O *Call Rate* deve ser determinado buscando o equilíbrio, minimizando o número de amostras retiradas e maximizando a eficiência da genotipagem.

5.1.3.2. Heterozigose

Amostras com elevada heterozigose podem ser indicação de que duas amostras foram colocadas erroneamente uma sobre a outra durante a montagem das placas, contaminação de DNA em alguma etapa laboratorial ou simplesmente indivíduos muito diferentes da maior parte do grupo amostral. A remoção de amostras a ± 3 desvios padrão é uma abordagem razoável para contornar esse problema (Laurie et al., 2010; Gondro et al., 2013).

5.1.3.3. Estratificação populacional

Estratificação da população refere-se a desigualdades de origem não aleatória nas frequências alélicas e na distribuição do fenótipo entre as diferentes subpopulações amostradas. A estratificação da população ocorre geralmente quando as amostras são

provenientes de populações ancestrais distintas e uma das principais consequências da estratificação é aumento do número de SNPs associados ao fenótipo de forma espúria (Purcell et al., 2007; Liu et al., 2013).

Em animais de criação, a estratificação da população ocorre principalmente por consequência da seleção, inseminação artificial, amostragem em diferentes grupos genéticos ou endogamia. Uma estratégia para evitar os erros induzidos por estratificação da população é a de assegurar que as amostras do estudo são provenientes de uma população relativamente homogênea e a escolha de métodos estatísticos que consideram este fator (Price et al., 2006; Li & Yu, 2008; Tse et al., 2009; Ma et al., 2013; Liu et al., 2013).

5.2. Métodos de estimação bayesiana

Para todos os métodos que serão tratados a seguir, os efeitos dos marcadores são estimados baseando-se no modelo (Kizilkaya et al., 2009; Habier et al., 2011):

$$y = \mu + X\beta + \sum_{i=1}^k z_i a_i \delta_i + e$$

Onde y é o vetor de observações fenotípicas; μ , a média geral; β , o vetor de parâmetros de perturbação; X , a matriz de incidência de β ; k , o número de marcadores; z_i , o vetor de genótipos do SNP i , codificados como o número de cópias de determinado alelo: 0, 1 e 2; a_i , o vetor de efeito de substituição alélica do SNP i ; δ_i , refere-se a determinação de π e indica se o SNP i será incluído ($\delta_i = 1$) ou excluído ($\delta_i = 0$) da análise e e é o vetor de efeito residual.

Para a realização das regressões bayesianas, distribuições *a priori* devem ser especificadas para os parâmetros de perturbação (β), efeitos genéticos aditivos (a) e para o resíduo (e). De acordo com Fernando e Garrick (2013), são adotadas as

seguintes distribuições *a priori*: discreta para β ; para vetor de resíduos e será atribuída distribuição *a priori* normal com média zero, condicional a variância residual (σ_e^2) que por sua vez é tratada como desconhecida com distribuição *a priori* qui-quadrado inversa e escalonada:

$$e|\sigma_e^2 \sim N(0, \sigma_e^2)$$

$$\sigma_e^2 \sim \chi^{-2}(\nu_e, S_e^2)$$

Onde, segundo Habier et al. (2011):

$$E(\sigma_e^2) = \frac{S_e^2 \nu_e}{\nu_e - 2} \text{ e } S_e^2 = \frac{E(\sigma_e^2)(\nu_e - 2)}{\nu_e}$$

Em que ν_e é o número de graus de liberdade e S_e^2 é o parâmetro de escala da distribuição.

Todas as metodologias tratadas a seguir, diferem somente na distribuição de probabilidade *a priori* para a , σ_a^2 e na atribuição do valor de π . Esses métodos são denominados Bayes B (Meuwissen et al., 2001), Bayes C (Kizilkaya et al., 2009) e Bayes C π (Habier et al., 2011).

5.2.1. Bayes B

O método proposto por Meuwissen et al. (2001) é ajustado por meio de abordagem bayesiana com estrutura hierárquica em dois níveis, assumindo-se que os efeitos dos marcadores apresentam uma mistura de distribuição de probabilidade *a priori*, onde um grande número de marcadores, com probabilidade π , apresentam um ponto de massa em zero e variância genética nula. Já para o restante dos marcadores com probabilidade $1 - \pi$, os efeitos dos marcadores são adotados como amostras de uma distribuição normal com média zero e variância genética específica para cada loco, $\sigma_{ai}^2 > 0$, com distribuição *a priori* qui-quadrado inversa e escalonada:

$$a_i | \sigma_{ai}^2 \begin{cases} 0 & \text{com probabilidade } \pi \\ \sim N(0, \sigma_{ai}^2) & \text{com probabilidade } (1 - \pi) \end{cases}$$

$$\sigma_{ai}^2 | \pi \begin{cases} 0 & \text{com probabilidade } \pi \\ \sim \chi^{-2}(\nu_a, S_a^2) & \text{com probabilidade } (1 - \pi) \end{cases}$$

Em que ν_a é o número de graus de liberdade e S_a^2 é o parâmetro de escala da distribuição. Esta mistura de distribuições é baseada na teoria de que muitos locos não apresentam efeito sobre as características e apenas poucos locos apresentam efeito e variância genética, portanto estão segregando (Meuwissen et al., 2001). A definição do valor *a priori* de π apresenta um desafio, pois requer conhecimento prévio sobre a característica estudada para que a determinação de seu valor seja correspondente com a proporção de QTLs que controlam a manifestação do fenótipo (Resende et al., 2011).

Gianola et al. (2009) sugeriram uma reformulação dos hiperparâmetros (σ_{ai}^2), em Bayes B. Segundo estes autores, o método Bayes B não permite o aprendizado bayesiano sobre essa variância, pois a *posteriori* é completamente condicional da variância locos específica e apresenta apenas um grau de liberdade adicional em comparação com a *priori*. Ao designar-se *a priori* $\sigma_{ai}^2 = 0$, não necessariamente conduzirá a $a_i = 0$, portanto, os autores sugerem que o estado zero seja especificado apenas para os efeitos e não para as variâncias e que para π poderia ser atribuída uma distribuição *a priori* Beta.

5.2.2. Bayes C

No método Bayes C descrito por Kizilkaya et al. (2009) e Habier et al. (2011), os efeitos dos marcadores *a priori* são adotados como amostras de uma distribuição normal com média zero e uma mesma variância genética é atribuída para todos os marcadores, dada por uma distribuição *a priori* qui-quadrado inversa e escalonada. A definição do valor de π é subjetiva assim como em Bayes B:

$$a \sim N(0, \sigma_a^2)$$

$$a|\sigma_a^2 \begin{cases} 0 & \text{com probabilidade } \pi \\ \sim N(0, \sigma_a^2) & \text{com probabilidade } (1 - \pi) \end{cases}$$

$$\sigma_a^2 \sim \chi^{-2}(\nu_a, S_a^2)$$

380

381 5.2.3. Bayes C π

382 Habier et al. (2011) propuseram uma alternativa para Bayes B ao amostrar
383 iterativamente o valor de π sob distribuição *a priori* uniforme (0, 1), conjuntamente
384 com os outros parâmetros desconhecidos do modelo. Em Bayes C π todos os marcadores
385 apresentam igual distribuição *a priori* para variância genética dada por uma distribuição
386 qui-quadrado inversa e escalonada:

$$\pi \sim \text{uniforme}(0, 1)$$

$$a|\sigma_a^2 \begin{cases} 0 & \text{com probabilidade } \pi \\ \sim N(0, \sigma_a^2) & \text{com probabilidade } (1 - \pi) \end{cases}$$

$$\sigma_a^2 \sim \chi^{-2}(\nu_a, S_a^2)$$

387 Os parâmetros ν_a e S_a^2 são semelhantes aos utilizados em Bayes B.

388

389 6. IMPUTAÇÃO DE GENÓTIPOS

390 Para detectar associação entre marcadores com QTLs é necessário que uma
391 amostra representativa da população seja genotipada (Miształ, 2011). Devido ao alto
392 custo da obtenção de genótipos em larga escala, é de grande relevância o estudo de
393 alternativas economicamente mais eficientes (Carvalho et al., 2014). Entre tais
394 alternativas, a imputação de genótipos se destaca por permitir prever genótipos de

marcadores faltantes em uma amostra com base em haplótipos presentes na população ou em indivíduos aparentados (Li et al., 2009).

Quando *chips* de diferentes densidades de SNPs são utilizados em GWAS, a imputação de genótipos é uma abordagem oportuna para combinar de forma eficiente todos os dados dos marcadores disponíveis. Mesmo em experimentos onde são utilizados marcadores em mesma densidade para todos os indivíduos, a imputação é útil para incrementar o *Call Rate* de amostras com baixa qualidade (Ma et al., 2013).

A utilização de genótipos imputados em GWAS foi inicialmente realizada a partir de dados obtidos com estudos realizados em humanos (Browning, 2008; Howie et al., 2012; Chen et al., 2012). Recentemente, diversos estudos de GWAS em bovinos de corte têm sido conduzidos com genótipos imputados (Barendse, 2011; Chen et al., 2012; Santana et al., 2014a, Santana et al., 2014b).

Porém, para obtenção de resultados confiáveis, a acurácia de imputação deve ser alta o suficiente para que as informações geradas forneçam confiabilidade semelhante à obtida se a amostra fosse genotipada com *chip* de maior densidade (Mulder et al., 2012). Entre os fatores que influenciam a acurácia de imputação, destacam-se a diferença na densidade de marcadores entre população de referência e população de validação (Mulder et al., 2012); o desequilíbrio de ligação (Druet et al., 2010); a diferença do número de indivíduos entre população de referência e população de validação (Hickey et al., 2012) e, principalmente, as diferenças entre diferentes metodologias (Druet et al., 2010; Hickey et al., 2012).

A acurácia da imputação pode ser avaliada de diferentes formas a nível de genótipos ou alelos: (1) proporção de genótipos ou alelos imputados corretamente (Weigel et al., 2010) ou incorretamente (Druet et al., 2010), (2) coeficiente de

correlação (r) e de determinação (r^2) entre genótipos reais e imputados (Mulder et al., 2010), (3) taxa de erro alélico (Khatkar et al., 2012).

7. HIPÓTESES A SEREM TESTADAS

As suposições teóricas que este experimento pretende testar são as seguintes:

1. É possível realizar imputação de genótipos a partir de painel de média densidade para alta densidade, com alta acurácia em bovinos da raça Canchim.
2. O estudo de associação ampla do genoma é capaz de identificar locos e genes ao longo dos cromossomos autossômicos associados com peso ao sobreano (PESO), área de olho de lombo (AOL) e espessura de gordura subcutânea (EGS) em bovinos da raça Canchim.

8. OBJETIVOS

8.1. Objetivo geral

Contribuir para o melhoramento genético da raça Canchim por meio do estudo de marcadores moleculares associados com características de produção e qualidade de carcaça em bovinos de corte.

8.2. Objetivos específicos

Avaliar a acurácia de imputação de dois diferentes cenários, sendo o primeiro a partir do Illumina *BovineSNP50 v2 BeadChip*® (50k) e o segundo a partir do *GeneSeek Genomic Profiler HD*® (80k), para o *BovineHD BeadChip*® (777k).

Identificar regiões genômicas associadas com a manifestação de fenótipos destinados à produção e qualidade de carcaça como peso ao sobreano, área de olho de lombo e espessura de gordura subcutânea, por meio dos métodos Bayes B e Bayes C.

444 Explorar as regiões encontradas, em busca de genes e QTLs com função
445 biológica potencialmente relacionada com as três características em estudo.

9. REFERÊNCIAS

- Anualpec. (2014) *Anuário da Pecuária Brasileira*. Informa Economics FNP, São Paulo, SP, Brasil.
- Ardlie, K. G.; Kruglyak, L.; Seielstad, M. (2002) Patterns of linkage disequilibrium in the human genome. *Nature Reviews Genetics*, 3, 299-309.
- Balding, D. J. (2006) A tutorial on statistical methods for population association studies. *Nature Reviews Genetics*, 7, 781-791.
- Barbosa, P. F. (2000) O Canchim na Embrapa Pecuária Sudeste. In: Anais da Convenção Nacional da Raça Canchim, São Carlos, SP (Brazil). Copyright EMBRAPA-CPPSE, São Paulo, Brasil.
- Barendse, W. (2011) Haplotype Analysis Improved Evidence for Candidate Genes for Intramuscular Fat Percentage from a Genome Wide Association Study of Cattle. *PLoS ONE*, 6, 12.
- Bergmann, J. A. G. (1999) Índices Zootécnicos para Produção de Bovinos de Carne. Escola de Veterinária da UFMG. Belo Horizonte, MG. Disponível em <http://professor.ucg.br/siteDocente/admin/arquivosUpload/8949/material/indices%20zo%20ot%20bov%20corte%20jagb.pdf> (Acesso em 20/02/2015).
- Borba, L. H. F.; Rey, F. S. B.; Silva, L. O. C.; Boligon, A. A.; Alencar, M. M. (2011) Parâmetros genéticos para características de crescimento e reprodução de bovinos da raça Canchim. *Pesquisa Agropecuária Brasileira*, 46, 1570-1578.
- Broman, K. W.; Sen, S. (2009) Data cheking. In: Broman, K. W.; Sen, S. A Guide to QTL Mapping with R/qtl. Springer-Verlag, New York, USA.
- Browning, S. R. (2008) Missing data imputation and haplotype phase inference for genome-wide association studies. *Human Genetics*, 124, 439-450.

470 Buzanskas, M. E.; Grossi, D. A.; Baldi, F.; Barrozo, D.; Silva, L. O. C.; Torres Júnior,
 471 R. A. A.; Munari, D. P.; Alencar, M. M. (2010) Genetic associations between
 472 stayability and reproductive and growth traits in Canchim beef cattle. *Livestock Science*,
 473 132, 107-112.

474 Buzanskas, M. E.; Grossi, D. A.; Ventura, R. V.; Schenkel, F. S.; Sargolzaei, M.;
 475 Meirelles, S. L. C.; Mokry, F. B.; Higa, R. H.; Mudadu, M. A.; da Silva, M. V. G. B.;
 476 Niciura, S. C. M.; Torres Júnior, R. A. A.; Alencar, M. M.; Regitano, L. C. A.; Munari,
 477 D. P. (2014) Genome-Wide Association for Growth Traits in Canchim Beef Cattle.
 478 *PLoS ONE*, 9, 4, e94802. doi:10.1371/journal.pone.0094802.

479 Caetano, S. L.; Savegnago, R. P.; Boligon, A. A.; Ramos, S. B.; Chud, T. C. S.; Lôbo,
 480 R. B.; Munari, D. P. (2013) Estimates of genetic parameters for carcass, growth and
 481 reproductive traits in Nelore cattle. *Livestock Science*, 155, 1-7.

482 Cantor, R. M.; Lange, K.; Sinsheimer, J. S. (2010) Prioritizing GWAS Results: A
 483 Review of Statistical Methods and Recommendations for Their Application. *The*
 484 *American Journal of Human Genetics*, 86, 6-22.

485 Carmo, A. S. (2012) Estudo de associação genômica e análise de enriquecimento
 486 genético da criotolerância espermática em bovinos da raça Nelore. Tese Doutorado.
 487 Universidade Estadual Paulista, Jaboticabal, Brasil.

488 Carvalheiro, R.; Boison, S. A.; Neves, H. H. R.; Sargolzaei, M.; Schenkel, F. S.;
 489 Utsunomiya, Y. T.; O'Brien, A. M. P.; Sölkner, J.; McEwan, J. C.; Van Tassell, C. P.;
 490 Sonstegard, T. S.; Garcia, J. F. (2014) Accuracy of genotype imputation in Nelore
 491 cattle. *Genetics Selection Evolution*, 1, 46-69.

492 Castro-Pereira, V. M.; Alencar, M. M.; Barbosa, R. T. (2007) Estimativas de parâmetros
 493 genéticos e de ganhos direto e indireto à seleção para características reprodutivas e de

494 crescimento em um rebanho da raça Canchim. *Revista Brasileira de Zootecnia*, 36,
 495 1029-1036.

496 Chen, M. H.; Huang, J.; Chen, W. M.; Larson, M. G.; Fox, C. S.; Vasan, R. S.;
 497 Seshadri, S.; O'Donnell, C. J.; Yang, Q. (2012) Using Family-Based Imputation in
 498 Genome-Wide Association Studies with Large Complex Pedigrees: The Framingham
 499 Heart Study. *PLoS ONE*, 7, 12.

500 Dickson, S. P.; Wang, K.; Krantz, I.; Hakonarson, H.; Goldstein, D. B. (2011) Rare
 501 variants create synthetic genome-wide associations. *PLoS Biology* , 8, 1.

502 Dolezal, H. G.; Smith, G. C.; Savell, J. W.; Carpenter, Z. L. (1982) Comparison of
 503 subcutaneous fat thickness, marbling and quality grade for predicting palatability of
 504 beef. *Journal of Food Science*, 47, 397-401.

505 Druet, T.; Schrooten, C; De Roos, A. P. (2010) Imputation of genotypes from different
 506 single nucleotide polymorphism panels in dairy cattle. *Journal of Dairy Science*, 93,
 507 5443-5454.

508 Faleiro, F. G.; Andrade, S. R. M.; Reis Junior, F. B. (2011) *Biotecnologia: estado da*
 509 *arte e aplicações na agropecuária*. Embrapa Cerrados, Planaltina, DF, Brasil.

510 Fernando, R; Garrick, D. (2013) Bayesian Methods Applied to GWAS. In: Gondro, C.;
 511 van der Werf, J.; Hayes, B. *Genome-Wide Association Studies and Genomic Prediction*,
 512 Springer, pp. 237-274.

513 Ferreira, J. J.; Brondani, I. L.; Leite, D. T.; Restle, J.; Filho, D. C. A.; Missio, R. L.;
 514 Heck, I; Segabinazzi, L. R. (2006) Características da carcaça de tourinhos Charolês e
 515 mestiços Charolês x Nelore terminados em confinamento. *Ciência Rural*, 36, 191-196.

516 Forni, S.; Federici, J. F.; Albuquerque, L. G. (2007) Tendências genéticas para escores
 517 visuais de conformação, precocidade e musculatura à desmama de bovinos Nelore.
 518 *Revista Brasileira de Zootecnia*, 36, 572-577.

519 Gianola, D.; De Los Campos, G.; Hill, W. G.; Manfredi, E.; Fernando, R. (2009)
 520 Additive genetic variability and the Bayesian alphabet. *Genetics*, 183, 347-363.
 521 Gondro, C.; Lee, S. H.; Lee, H. K.; Porto-Neto, L. R. (2013) Quality Control for
 522 Genome-Wide Association Studies. In: Gondro, C.; van der Werf, J.; Hayes, B.
 523 *Genome-Wide Association Studies and Genomic Prediction*, Springer, pp. 129-147.
 524 Habier, D.; Fernando, R. L.; Kizilkaya, K.; Garrick, D. J. (2011) Extension of the
 525 bayesian alphabet for genomic selection. *BMC Bioinformatics*, 12, 186.
 526 Hardy, G. H. (1908) Mendelian proportions in a mixed population. *Science*, 28, 49-50.
 527 Hassen, A.; Wilson, D. E.; Rouse, G. H. (1999) Evaluation of carcass, live, and real-
 528 time ultrasound measures in feedlot cattle: I. Assessment of sex and breed effects.
 529 *Journal of Animal Science*, 77, 273-282.
 530 Hey, J. (2004) What's So Hot about Recombination Hotspots? *PLoS Biology*, 2, 6.
 531 Hickey, J. M.; Crossa, J.; Babu, R.; Campos, G. (2012) Factors affecting the accuracy of
 532 genotype imputation in populations from several maize breeding programs. *Crop*
 533 *Science*, 52, 654-663.
 534 Howie, B.; Fuchsberger, C.; Stephens, M.; Marchini, J.; Abecasis, G. R. (2012) Fast and
 535 accurate genotype imputation in genome-wide association studies through pre-phasing. *Nature*
 536 *Genetics*, 44, 955-959.
 537 Khatkar, M. S.; Moser, G.; Hayes, B. J.; Raadsma, H. W. (2012) Strategies and utility
 538 of imputed SNP genotypes for genomic analysis in dairy cattle. *BMC Genomics*, 13,
 539 538.
 540 Khatkar, M. S.; Nicholas, F. W.; Collins, A. R.; Zenger, K. R.; Cavanagh, J. A.; Barris,
 541 W.; Schnabel, R. D.; Taylor, J. F.; Raadsma, H. W. (2008) Extent of genome-wide
 542 linkage disequilibrium in Australian Holstein- Friesian cattle based on a high-density
 543 SNP panel. *BMC Genomics*, 9, 187.

544 Kizilkaya, K.; Fernando, R. L.; Garrick, D. J. (2009) Genomic prediction of simulated
 545 multibreed and purebred performance using observed fifty thousand single nucleotide
 546 polymorphism genotypes. *Journal of Animal Science*, 88, 544-551.

547 Laurie, C. C.; Doheny, K. F.; Mirel, D. B.; Pugh, E. W.; Bierut, L. J.; Bhangale,
 548 T.; Boehm, F.; Caporaso, N. E.; Cornelis, M. C.; Edenberg, H. J.; Gabriel, S. B.; Harris,
 549 E. L.; Hu, F. B.; Jacobs, K. B.; Kraft, P.; Landi, M. T.; Lumley, T.; Manolio, T.
 550 A.; McHugh, C.; Painter, I.; Paschall, J.; Rice, J. P.; Rice, K. M.; Zheng, X.; Weir, B. S.
 551 (2010) Quality control and quality assurance in genotypic data for genome-wide
 552 association studies. *Genetic Epidemiology*, 34 (suppl. 6): 591-602.

553 Li, Q.; Yu, K. (2008) Improved correction for population stratification in genomewide
 554 association studies by identifying hidden population structures. *Genetics Epidemiology*,
 555 32, 215-226.

556 Li, Y.; Willer, C.; Sanna, S.; Abecasis, G. (2009) Genotype Imputation. *Annual Review*
 557 *of Genomics and Human Genetics*, 10, 387-406.

558 Liu, L.; Zhang, D.; Liu, H.; Arendt, C. (2013) Robust methods for population
 559 stratification in genome wide association studies, *BMC Bioinformatics*, 14, 132.

560 Luchiari Filho, A. (2000) *Pecuária da carne bovina*. A. São Paulo, Brasil, 134p.

561 Luchiari Filho, A. (2006) Produção de carne bovina no Brasil: qualidade, quantidade ou
 562 ambas? In: Simpósio sobre desafios e novas tecnologias na bovinocultura de corte.
 563 Brasília (Brasil).

564 Ma, P.; Brøndum, R. F.; Zhang, Q.; Lund, M. S.; Su, G. (2013) Comparison of
 565 different methods for imputing genome-wide marker genotypes in Swedish and Finnish
 566 Red Cattle. *Journal of Dairy Science*, 96, 4666-4677.

567 Mascioli, A. S.; Alencar, M. M.; Barbosa, P. F.; Novaes, A. P.; Oliveira, M. C. S.
 568 (1996) Estimativas de parâmetros genéticos e proposição de critérios de seleção para
 569 pesos na raça Canchim. *Revista Brasileira de Zootecnia*, 24, 72-82.

570 Matukumalli, L. K.; Lawley, C. T.; Schnabel, R. D. (2009) Development and
 571 Characterization of a High Density SNP Genotyping Assay for Cattle. *PLoS ONE*, 4, 4.

572 Meuwissen, T. H. E.; Goddard, M. E.; Hayes, B. J. (2001) Prediction of total genetic
 573 value using genome-wide dense marker maps. *Genetics*, 157, 1819-1829.

574 Meirelles, S. L.; Alencar, M. M.; Oliveira, H. N.; Regitano, L. C. A. (2009) Estimativas
 575 de parâmetros genéticos em função do sexo para características de carcaça de animais
 576 da raça Canchim. In: Anais da Reunião Anual da Sociedade Brasileira de Zootecnia,
 577 Maringá (Brasil), 2009. SBZ:UEM, Maringá, Brasil.

578 Meirelles, S. L.; Alencar, M. M.; Oliveira, H. N.; Regitano, L. C. A. (2010) Efeitos de
 579 ambiente e estimativas de parâmetros genéticos para características de carcaça em
 580 bovinos da raça Canchim criados em pastagem. *Revista Brasileira de Zootecnia*, 39,
 581 1437-1442.

582 Misztal, I. (2011) FAQ for genomic selection. *Journal of Animal Breeding and*
 583 *Genetics*, 128, 245-246.

584 Mokry, F. B.; Meirelles, S. L. C.; Tullio, R. R.; Alencar, M. M.; Regitano, L. C. A.
 585 (2012) Interação genótipo x ambiente em características avaliadas por ultrassom em
 586 bovinos da raça Canchim. In: Anais do Simpósio Brasileiro de Melhoramento Animal.
 587 João Pessoa (Brasil).

588 Mulder, H. A.; Meuwissen, T. H. E.; Calus, M. P. L.; Veerkamp, R. F. (2010). The
 589 effect of missing marker genotypes on the accuracy of gene-assisted breeding value
 590 estimation: A comparison of methods. *Animal*, 4, 9-19.

591 Mulder, H. A.; Calus, M. P. L.; Druet, T.; Schrooten, C. (2012) Imputation of genotypes
 592 with low-density chips and its effect on reliability of direct genomic values in Dutch
 593 Holstein cattle. *Journal of Dairy Science*, 95, 876-889.

594 National Human Genome Research Institute. (2011) A Catalog of Published Genome
 595 Wide Association Studies. Disponível em: <http://www.genome.gov/GWASudies/>.
 596 Acesso em: 09 de julho, 2013.

597 Pérez-Enciso, M.; Rincón, J. C.; Legarra, A. (2015) Sequence- vs. chip-assisted
 598 genomic selection: accurate biological information is advised. *Genetics Selection
 599 Evolution*, 47, 43.

600 Perotto, D.; Abrahão, J. J. S.; Moletta, J. L. (2000) Características Quantitativas de
 601 Carcaça de Bovinos Zebu e de Cruzamentos Bos taurus x Zebu. *Revista Brasileira de
 602 Zootecnia*, 29(supl. 1): 2019-2029.

603 Price, A. L.; Patterson, N. J.; Plenge, R. M.; Weinblatt, M. E.; Shadick, N. A.; Reich,
 604 D. (2006) Principal components analysis corrects for stratification in genome-wide
 605 association studies. *Nature Genetics*, 38, 904-909.

606 Purcell, S.; Neale, B.; Todd-Brown, K.; Thomas, L.; Ferreira, M. A. R.; Bender, D.;
 607 Maller, J.; Sklar, P.; Bakker, P. I. W.; Daly, M. J.; Sham, P. C. (2007) PLINK: A Tool
 608 Set for Whole-Genome Association and Population-Based Linkage Analyses. *The
 609 American Journal of Human Genetics*, 81, 559-575.

610 Razook, A. G.; Figueiredo, L. A.; Nardon, R. F.; Cyrillo, J. N. S. G.; Ruggieri, A. C.
 611 (2001) Efeitos de raça e da seleção para peso pós-desmame sobre características de
 612 confinamento e de carcaça da 15a progênie dos rebanhos Zebu e Caracu de Sertãozinho
 613 (SP). *Revista Brasileira de Zootecnia*, 30, 115-124.

614 Resende, M. D. V.; Silva, F. F.; Viana, J. M. S.; Peternelli, L. A.; Resende Jr, M. F. R.;
 615 del Valle, P. M. (2011) Métodos estatísticos na seleção genômica ampla. In: Série
 616 Documentos 219, Embrapa Florestas, Colombo, Paraná, Brasil.
 617 Ritchie, M. E.; Liu, R.; Carvalho, B. S. (2011) Comparing genotyping algorithms for
 618 Illumina's Infinium whole-genome SNP BeadChips. *BMC Bioinformatics*, 12, 68.
 619 Salanti, G.; Amountza, G.; Ntzani, E.; Ioannidis, J. P. A. (2005) Hardy-Weinberg
 620 equilibrium in genetic association studies: an empirical evaluation of reporting,
 621 deviations, and power. *European Journal of Human Genetics*, 13, 840-848.
 622 Santana, M. H. A.; Utsunomiya, Y. T.; Neves, H. H. R.; Gomes, R. C.; Garcia, J. F.;
 623 Fukumasu, H.; Silva, S. L.; Leme, P. R.; Coutinho, L. L.; Eler, J. P.; Ferraz, J. B. S.
 624 (2014a) Genome-wide association study for feedlot average daily gain in Nellore cattle
 625 (*Bos indicus*). *Journal of Breeding and Genetics*, 131, 210-216.
 626 Santana, M. H. A.; Utsunomiya, Y. T.; Neves, H. H. R.; Gomes, R. C.; Garcia, J. F.;
 627 Fukumasu, H.; Silva, S. L.; Oliveira Junior, G. A.; Alexandre, P. A.; Leme, P. R.;
 628 Brassaloti, R. A.; Coutinho, L. L.; Lopes, T. G.; Meirelles, F. V.; eler, J. P.; Ferraz, J. B.
 629 S. (2014b) Genome-wide association analysis of feed intake and residual feed intake in
 630 Nellore cattle. *BMC Genetics*, 15, 21.
 631 Santos, A. M.; Alencar, M. M.; Barbosa, P. F.; Novaes, A. P.; Oliveira, M. C. S. (1996)
 632 Estimativas de parâmetros genéticos e proposição de critérios de seleção para pesos na
 633 raça Canchim. *Revista Brasileira de Zootecnia*, v. 25, n. 1, p. 72-82
 634 Silva, A. M.; Alencar, M. M.; Freitas A. R.; Barbosa, R. T.; Barbosa, P. F.; Oliveira, M.
 635 C. S.; Corrêa L. A.; Novaes, A. P.; Tullio, R. R. (2000) Herdabilidades e correlações
 636 genéticas para peso e perímetro escrotal de machos e características reprodutivas e de
 637 crescimento de fêmeas, na raça Canchim. *Rev. bras. zootec.*, 29(6):2223-2230, sup. 2.

638 Soller, M. (1978) The use of loci associated with quantitative effects in dairy cattle
639 improvement. *Animal Production*, 27, 133-139.

640 Soller, M.; Plotkin-Hazan, J. (1977) The use marker alleles for the introgression of
641 linked quantitative alleles. *TAG Theoretical and Applied Genetics* 51: 133-137.

642 Suguisawa, L.; Mattos, W. R. S.; Oliveira, H. N.; Silveira, A. C.; Arrigoni, M. B.;
643 Haddad, C. M.; Chardulo, L. A. L.; Martins, C. L. (2003) Ultrasonography as a
644 predicting tool for carcass traits of young bulls. *Scientia Agrícola*, 60, 79-784.

645 Suguisawa, L.; Mattos, W. R. S.; Oliveira, H. N.; Silveira, A. C.; Arrigoni, M. B.;
646 Souza, A. A. (2006) Correlações simples entre as medidas de ultrassom e a composição
647 da carcaça de bovinos jovens. *Revista Brasileira de Zootecnia*, 35, 169-176.

648 The Bovine Genome Sequencing and Analysis Consortium. (2009) The Genome
649 Sequence of Taurine Cattle: A window to ruminant biology and evolution. *Science*, 324,
650 522-528.

651 The Bovine Hapmap Consortium. (2009) Genome-Wide survey of SNP variation
652 uncovers the genetic structure of cattle breeds. *Science*, 324, 528-532.

653 Tse, K. P.; Su, W. H.; Chang, K. P.; Tsang, N. M.; Yu, C. J. (2009) Genome-wide
654 association study reveals multiple nasopharyngeal carcinoma-associated loci within the
655 HLA region at chromosome 6p21.3. *The American Journal of Human Genetics*, 85,
656 194-203.

657 Tupy, O.; Primavesi, O.; Barbosa, P. F. (2006) Avaliação dos impactos econômicos
658 sociais e ambientais de tecnologias da Embrapa Pecuária Sudeste: Utilização de touros
659 da raça canchim em cruzamentos terminal com fêmeas da raça nelore. In: Série
660 Documentos 54, Embrapa Pecuária Sudeste, São Carlos, São Paulo, Brasil.

661 Turner, S.; Armstrong, L. L.; Bradford, Y.; Carlson, C. S.; Crawford, D. C.; Crenshaw,
662 A. T.; de Andrade, M.; Doheny, K. F.; Haines, J. L.; Hayes, G.; Jarvik, G.; Jiang, L.;

663 Kullo, I. J.; Li, R.; Ling, H.; Manolio, T. A.; Matsumoto, M.; McCarty, C. A.;
 664 McDavid, A. N.; Mirel, D. B.; Paschall, J. E.; Pugh, E. W.; Rasmussen, L. V.; Wilke, R.
 665 A.; Zuvich, R. L.; Ritchie, M. D. (2011) Quality Control Procedures for Genome-Wide
 666 Association Studies. *Current Protocols in Human Genetics*, 1, 19.
 667 Vianna, A. T.; Gomes, F. P.; Santiago, M. (1978) *Formação do Gado Canchim pelo*
 668 *Cruzamento Charolês-Zebu*. Nobel, São Paulo, SP, Brasil.
 669 Weigel, K. A.; Van Tassell, C. P.; O'Connell, J. R.; VanRaden, P. M.; Wiggans, G. R.
 670 (2010) Prediction of unobserved single nucleotide polymorphism genotypes of Jersey
 671 cattle using reference panels and population-based imputation algorithms. *Journal of*
 672 *Dairy Science*, 93, 2229-2238.
 673 Weinberg W (1908) On the demonstration of heredity in man. In: Boyer SH, editor,
 674 Papers on human genetics. Englewood Cliffs, NJ: Prentice Hall. Translated, 1963.
 675 Yokoo, M. J.; Lôbo, R. B.; Araujo, F. R. C.; Bezerra, L. A. F.; Sainz, R. D.;
 676 Albuquerque, L. G. (2010) Genetic associations between carcass traits measured by
 677 realtime ultrasound and scrotal circumference and growth traits in Nellore cattle.
 678 *Journal of Animal Science*, 88, 52-58.
 679 Ziegler, A.; Köning, I. R; Thompson, J. R. (2008) Biostatistical Aspects of Genome-
 680 Wide Association Studies. *Biometrical Journal*, 50, 8-28.

CAPÍTULO I

ESTUDOS DE ASSOCIAÇÃO AMPLA DO GENOMA PARA ÁREA DE OLHO DE LOMBO, ESPESSURA DE GORDURA SUBCUTÂNEA E PESO AO SOBREANO EM BOVINOS DA RAÇA CANCHIM

RESUMO

O valor comercial da carcaça bovina é determinado por um conjunto de características, onde se destacam o peso, o rendimento, a gordura de cobertura e o marmoreio. Portanto, o melhoramento genético de características de crescimento e de carcaça demonstra-se como uma importante ferramenta para agregar valor ao produto final. Estudos de associação ampla do genoma (*Genome-wide Association Study* - GWAS) viabilizam a identificação de locos que controlam a expressão fenotípica de características quantitativas (*Quantitative Trait Loci* - QTLs). Neste contexto, o objetivo desse trabalho foi identificar regiões genômicas e genes associados com características de crescimento e qualidade de carcaça como área de olho de lombo (AOL), espessura de gordura subcutânea (EGS) e peso ao sobreano (PESO) em bovinos da raça Canchim. Para aumentar o tamanho amostral e a densidade de marcadores, foi realizada imputação de genótipos e nas análises subsequentes foram utilizados somente marcadores imputados com acurácia superior a 95%. O GWAS foi realizado por meio dos métodos Bayes B e Bayes C, utilizando informações fenotípicas e genótipos imputados de 701 animais da raça Canchim e do grupo genético MA (progênie de touro Charolês e vaca $\frac{1}{2}$ Canchim + $\frac{1}{2}$ Zebu). Foram identificadas no total, duas, três e cinco regiões genômicas associadas com AOL, PESO e EGS, respectivamente. Os métodos Bayes B e Bayes C permitiram identificar regiões cromossômicas associadas com AOL, EGS e PESO em bovinos da raça Canchim e a análise funcional destacou genes com

705 funções biológicas relacionadas à essas características. A identificação de QTLs
706 previamente descritos na literatura reforçam as evidências de que as regiões encontradas
707 por este trabalho estão efetivamente associadas com as características estudadas.

708

709 **Palavras chave:** SNP, carcaça, crescimento, imputação, inferência Bayesiana

1. INTRODUÇÃO

O melhoramento de características de crescimento e de carcaça demonstra-se de grande importância para criadores e indústria, dado que o valor comercial de uma carcaça bovina é determinado por um conjunto de aspectos, destacando-se o peso, o rendimento, a gordura de cobertura e o marmoreio (Perotto et al., 2000).

Entre as características de crescimento, o peso ao sobreano (PESO) destaca-se por permitir avaliar o potencial genético do indivíduo ainda em idade jovem, tendo em vista a menor influência dos efeitos ambientais maternos e maior correlação com o peso ao abate quando comparado ao peso a desmama e ao ano (Bergman, 1999).

Entre as características de carcaça, a área de olho de lombo (AOL) apresenta correlação com musculosidade e rendimento de cortes cárneos, estando relacionada com a porção comestível da carcaça (Luchiari Filho, 2000). A espessura de gordura subcutânea (EGS) apresenta grande importância no processo de resfriamento das carcaças, isolando-as do frio, protegendo-as contra o encurtamento das fibras e, consequentemente, o endurecimento da carne (Meirelles et al., 2010).

No geral, as características de crescimento e de carcaça de bovinos apresentam herdabilidade de magnitude moderada a alta, evidenciando que a seleção e a utilização de raças especializadas em programas de cruzamento, se constituem como importantes ferramentas para melhoria desses aspectos no rebanho (Forni et al., 2007).

Neste contexto, a raça Canchim destaca-se por proporcionar um ganho de peso na ordem de 15% em relação ao zebu (Tupy et al., 2006). Atualmente, é bastante utilizada em sistemas de produção de carne, principalmente pelo cruzamento de touros Canchim com vacas zebu, devido à capacidade de monta e fertilidade do touro Canchim e ao desempenho, qualidade de carne e carcaça dos seus produtos (Vianna et al., 1978; Tupy et al., 2006).

As análises do genoma, incluindo o mapeamento de QTLs (*Quantitative Trait Loci*) (Soller & Plotkin-Hazan 1977; Soller 1978), o emprego em larga escala de SNPs (*Single Nucleotide Polymorphism*) (Meuwissen et al., 2001) e, mais recentemente, a utilização de dados de sequenciamento (Pérez-Enciso et al., 2015) na seleção de bovinos de corte, permitirão avanços na inclusão de características relacionadas à qualidade da carcaça e da carne nos programas de melhoramento genético.

Com o desenvolvimento de plataformas para a detecção simultânea de milhares de SNPs, tornou-se possível aumentar a confiabilidade nos resultados obtidos por meio de mapeamento de QTLs, empregando marcadores distribuídos densamente por todo o genoma (Faleiro et al., 2011). Essa metodologia tem sido chamada de estudo de associação genômica ampla (*Genome-wide Association Study* - GWAS) e visa à identificação de SNPs que estão em desequilíbrio de ligação com regiões cromossômicas que controlam a manifestação de fenótipos relevantes (Balding, 2006).

Neste sentido, o objetivo deste trabalho foi identificar regiões genômicas e genes associados com características de produção e qualidade de carcaça como AOL, EGS e PESO em bovinos da raça Canchim por meio de GWAS, utilizando os métodos Bayes B e Bayes C.

2. MATERIAL E MÉTODOS

2.1. Dados fenotípicos

Os dados fenotípicos utilizados foram provenientes de projetos desenvolvidos na Embrapa Pecuária Sudeste e na Embrapa Gado de Corte. Dos projetos da Embrapa Pecuária Sudeste foram obtidas informações fenotípicas de 1.652 animais, sendo 777 machos e 875 fêmeas, onde 1.231 eram da raça Canchim e 421 do grupo genético MA. Os animais foram provenientes de sete fazendas localizadas nos estados de Goiás e São

Paulo e tiveram informações fenotípicas coletadas entre 2005 e 2010. Em São Paulo, as fazendas encontram-se nos municípios de Águas de Santa Bárbara, Angatuba, Capão Bonito, Sandovalina, São Carlos e São Miguel Arcanjo. No estado de Goiás, a fazenda encontra-se no município de Jussara.

Dos projetos da Embrapa Gado de Corte, foram coletadas informações fenotípicas e amostras biológicas de 976 bezerros participantes de sete edições das PCADs (Prova Canchim de Avaliação de Desempenho) promovidas pela Associação Brasileira dos Criadores de Canchim (ABCCAN) e Programa Geneplus/Embrapa. A primeira Prova foi realizada na cidade de Campo Grande - MS em 2011. A partir de 2012, a PCAD passou a ser realizada anualmente em duas etapas com o objetivo de avaliar separadamente animais nascidos de junho a agosto dos que nasceram de setembro a novembro. Neste ano, as etapas foram realizadas nas cidades de Campo Grande - MS e Angatuba - SP e nos dois anos seguintes as Provas foram realizadas nos municípios de Jussara - GO e Angatuba – SP. Todos os animais participantes das PCADs eram machos, sendo 854 indivíduos da raça Canchim e 122 do grupo genético MA (progênie de touro Charolês e vaca $\frac{1}{2}$ Canchim + $\frac{1}{2}$ Zebu e apresenta em média 65,6% Charolês e 34,4% Zebu), oriundos de 23 fazendas.

Nas PCADs, foram coletadas informações fenotípicas de peso ao sobreano (PESO), área de olho de lombo (AOL) e espessura de gordura subcutânea (EGS). As medidas de EGS (mm), AOL (cm²) foram coletadas transversalmente no músculo *Longissimus dorsi* na região entre a 12^a e 13^a costelas com um equipamento de ultrassonografia com transdutor linear de 3,5 MHz ASP 30C/18 cm, específico para avaliação de carcaça.

2.2. Estimação de componentes de (co)variância e parâmetros genéticos

Os componentes de (co)variância utilizados *a priori* para as análises de GWAS foram estimados por meio da máxima verossimilhança restrita sob modelo animal, em análise multi característica utilizando o *software* REMLF90 (Misztal et al., 2002).

Como efeito fixo, foi considerado apenas o grupo contemporâneo, que foi formado por ano de nascimento, estação de nascimento, fazenda, Prova, sexo e grupo genético (Canchim ou MA). Para esta análise, foram excluídos animais que não apresentavam informações completas para as três características avaliadas, os que apresentavam fenótipos com ± 3 desvios padrão a partir da média do seu grupo contemporâneo e grupos contemporâneos com menos de cinco indivíduos.

Como covariáveis foram considerados os efeitos de idade linear e quadrático. O modelo utilizado em conotação matricial pode ser representado como:

$$y = \mu + X\beta + Zg + e$$

Onde, y é o vetor de observações; μ , a média geral; β , o vetor de efeitos fixos; X , a matriz de incidência dos efeitos fixos; g , o vetor de efeitos genéticos aditivos diretos; Z , a matriz de incidência dos efeitos genéticos aditivos diretos; e , o efeito residual.

2.3. Dados genômicos

Para todos os animais avaliados, foram colhidas amostras de sangue em tubos a vácuo (Vacutainer®) de 4,5 mL contendo K3 EDTA, que em seguida foram processadas para obtenção dos leucócitos. Para os touros, pais de alguns bezerros representativos das PCADs, foram obtidas amostras de sêmen. As extrações de DNA foram realizadas utilizando os protocolos descritos por Regitano & Coutinho (2001). As genotipagens foram feitas por empresas prestadoras de serviços e todos os animais foram genotipados por meio da tecnologia Illumina®.

Na Embrapa Pecuária Sudeste, foram genotipados 400 animais por meio do *BovineHD BeadChip*® (HD), sendo 194 machos e 206 fêmeas. Destes, 288 da raça Canchim e 112 do grupo genético MA. Este conjunto de dados foi chamado de HD_CPPSE.

Na Embrapa Gado de Corte, foram genotipados 689 animais em três diferentes densidades. O primeiro conjunto de dados foi composto por 500 indivíduos genotipados por meio do Illumina *BovineSNP50 v2 BeadChip*® e foi chamado de 50k_CNPGC. O segundo conjunto de dados foi composto por 176 indivíduos genotipados por meio do *GeneSeek Genomic Profiler HD*® e foi chamado de 80k_CNPGC. O terceiro conjunto de dados foi composto por 13 indivíduos genotipados por meio do *BovineHD BeadChip*® e foi chamado de HD_CNPGC. Destes animais, 593 eram da raça Canchim e 96 do grupo genético MA.

Também foram genotipados 31 touros pais de bezerros participantes das PCADs 2011 e 2012, que foram escolhidos com o objetivo de relacionar os diferentes conjuntos de dados, visando obter maior eficiência nas análises subsequentes. Destes touros, 23 eram da raça Canchim, 1 do grupo genético MA e 7 da raça Charolesa (pais de animais MA). Este conjunto de dados foi chamado de HD_CNPGC_SIRE.

2.4. Controle de qualidade dos dados genômicos

Para o controle de qualidade dos dados genômicos, foram considerados somente SNPs com coordenadas genômicas conhecidas e contidos nos cromossomos autossômicos e, nesta etapa, cada conjunto de dados foi analisado separadamente. Os seguintes critérios e limiares de exclusão foram utilizados: genótipos com *Genotype Calling Score* inferior a 50%; amostras e SNPs com *Call Rate* inferior a 95%; frequência alélica mínima (MAF) de 5%; amostras que apresentaram heterozigose >

75%; valor de P do teste qui-quadrado para equilíbrio de Hardy-Weinberg < 0.000001 ; para SNPs situados em mesma coordenada genômica excluiu-se o de menor MAF; para pares de SNPs com correlação superior a 0,98, excluiu-se o de menor MAF. Essa etapa foi realizada por meio de scripts no *software* estatístico R v.3.1.3, utilizando o pacote *snpStats* v1.16.0 (Clayton, 2014).

2.5. Estratificação populacional

Utilizando apenas SNPs em comum para todos os *data sets* foi calculada e representada graficamente a estrutura da população com base nos resultados de uma análise de componentes principais (PCA) (Figura 1) e o parentesco entre indivíduos foi calculado usando a equação clássica de VanRaden (2007).

2.6. Imputação de genótipos

Com o objetivo de aumentar o tamanho amostral e densidade de marcadores, foi realizada imputação de genótipos para combinar as informações disponíveis nos diferentes conjuntos de dados por meio do *software* FImpute v.2.2. (Sargolzaei et al., 2014).

Foram utilizados dois cenários, sendo que no primeiro, a população de imputação foi composta pelos animais do 50k_CNPGC. Já no segundo, a população de imputação foi composta pelos animais do 80k_CNPGC. A população de validação e de referência utilizadas em ambos os cenários foram as mesmas.

A população de validação foi composta por 100 animais, sendo os 13 animais do HD_CNPGC e 87 animais mais jovens do HD_CPPSE. Estes indivíduos tiveram seus genótipos mascarados para menores densidades de acordo com cada cenário e os genótipos imputados foram comparados com os originais.

A população de referência foi composta por 332 indivíduos, sendo 301 animais provenientes do HD_CPPSE e os 31 touros do HD_CNPGC_SIRE

Após a imputação, foi efetuado um novo controle de qualidade dos dados genômicos, onde foram excluídos SNPs com MAF < 5%, valor de P do teste qui-quadrado para equilíbrio de Hardy-Weinberg < 0.000001 e entre pares de SNPs com correlação superior a 0,98 excluiu-se o de menor MAF.

2.7. Estudo de associação ampla do genoma

Para as análises de GWAS foram utilizados informações de 701 animais, sendo 462 indivíduos provenientes do 50k_CNPGC, 169 do 80k_CNPGC, 57 do HD_CPPSE e 13 do HD_CNPGC.

Foram utilizados SNPs com percentagem de genótipos imputados corretamente maior que 95%. As análises foram realizadas em sistema operacional Linux, sob enfoque Bayesiano, por meio das metodologias Bayes C (Kizilkaya et al., 2009; Habier et al., 2011) e Bayes B (Meuwissen et al., 2001), implementadas no *software* GenSel v1.14 (Fernando & Garrick, 2014).

Para a obtenção da distribuição *a posteriori* dos efeitos genéticos dos marcadores $p(a_i, \sigma_{a_i}^2 | y)$, adotou-se o procedimento de simulação estocástica Monte Carlo via Cadeias de Markov, utilizando o algoritmo Metropolis-Hastings, que permite amostrar a distribuição *posteriori* conjunta de duas ou mais variáveis aleatórias, utilizando uma distribuição candidata como distribuição auxiliar para atualizar o algoritmo durante o processo iterativo (Tierney, 1994; Chib & Greenberg, 1995). O modelo utilizado para todos os métodos foi descrito por Kizilkaya et al. (2010):

$$y = \mu + X\beta + \sum_{i=1}^k z_i a_i \delta_i + e$$

Onde $\mathbf{y}$ é o vetor de observações fenotípicas; $\boldsymbol{\mu}$, a média geral; $\boldsymbol{\beta}$, o vetor de parâmetros de perturbação; $\mathbf{X}$, a matriz de incidência de $\boldsymbol{\beta}$; k , o número de marcadores; $\mathbf{z}_i$, o vetor de genótipos do SNP i , codificados como -10, 0 e 10 para os respectivos genótipos AA, AB e BB; $\mathbf{a}_i$, o vetor de efeitos de substituição alélica do SNP i ; $\boldsymbol{\delta}_i$, indica se o SNP i será incluído ($\delta_i = 1$) ou excluído ($\delta_i = 0$) da análise e $\mathbf{e}$ é o vetor de efeito residual. Foi adotado como parâmetro de perturbação $\boldsymbol{\beta}$ apenas o grupo contemporâneo, que foi formado por ano de nascimento, estação de nascimento, fazenda, Prova, sexo e grupo genético. Como covariáveis, foram considerados os efeitos de idade linear e quadrático.

Utilizaram-se os mesmos critérios na obtenção das cadeias de Markov para todas as características avaliadas. Foi realizado um total de 165.000 iterações, sendo que as 15.000 primeiras foram descartadas no *burn-in*. Para garantir a independência entre as observações, os resultados foram salvos a cada 10 iterações. Dessa forma, a estimação sobre as densidades de probabilidade *a posteriori* dos parâmetros foi realizada a partir de 15.000 amostragens. As distribuições de π a priori utilizadas em Bayes B e Bayes C foram obtidas por meio da metodologia Bayes C π (Habier et al., 2011), utilizando o valor de π a priori = 0,5.

Janelas compostas por vários marcadores capturam melhor parte da variabilidade de determinada região cromossômica (Saatchi et al., 2012), portanto, a associação entre as regiões genômicas e as características fenotípicas foi avaliada utilizando-se 2.521 janelas de 1Mb (1 milhão de pares de bases) não sobrepostas, incluindo número variável de SNPs e com diferentes magnitudes de efeito. A proporção da variância genética aditiva esperada para uma única janela foi calculada de acordo com Onteru et al. (2013), onde $100\% / 2.521$, que resultou em 0,039%. Portanto, janelas que explicaram pelo menos 0,19% da variância genética, que é cinco vezes maior do que o esperado (0,039%

* 5 = 0,19%), foram consideradas como regiões associadas à QTLs. Os efeitos dos marcadores foram padronizados de acordo com Sollero et al. (2014), onde $(a_i)^2 / \sigma_{ai}^2$.

2.8. Análise de enriquecimento funcional

As regiões encontradas foram exploradas quanto a sua localização para a prospecção de genes utilizando-se do banco de dados *Ensembl* (Cunningham et al., 2015). Os genes encontrados foram submetidos à ferramenta *Database for Annotation, Visualization, and Integrated Discovery* – DAVID (Dennis Jr, et al., 2003) e tiveram sua funcionalidade verificada nas categorias: componente celular, processo biológico e função molecular, por meio do banco de dados *QuickGO* (Binns, et al., 2009). A verificação de vias metabólicas foi realizada por meio do banco de dados *Kyoto Encyclopedia of Genes and Genomes* - KEGG (Kanehisa & Goto, 2000). A busca por QTLs previamente descritos foi feita utilizando o banco de dados *AnimalQTLdb* (Hu et al., 2007). Quando não foram encontradas informações funcionais para *Bos taurus*, foram realizadas buscas em genes ortólogos no genoma de outras espécies de mamíferos (*Homo sapiens*, *Sus scrofa*, *Mus musculus*, etc.).

3. RESULTADOS

A estatística descritiva para os fenótipos AOL, EGS e PESO encontram-se na Tabela 1. Para as estimativas de componentes de (co)variâncias e parâmetros genéticos foram excluídos 448 animais por meio dos critérios de exclusão adotados para os fenótipos. Os resultados obtidos para cada característica por meio da máxima verossimilhança restrita, por Bayes B e Bayes C encontram-se na Tabela 2.

Pelos critérios adotados no controle de qualidade, a partir dos 500 animais genotipados por meio do *BovineSNP50 v2 BeadChip®*, foram removidos 23.985

(43,92%) SNPs e 38 (7,6%) amostras, resultando em 30.624 (56,08%) SNPs e 462 animais (92,4%) utilizados nas análises subsequentes. A partir dos 176 animais genotipados por meio do *GeneSeek Genomic Profiler HD*®, foram removidos 27.548 (35,83%) SNPs e 7 (3,97%) amostras, resultando em 49.331 (64,17%) SNPs e 169 animais (96,03%) utilizados nas análises subsequentes. A partir dos animais genotipados por meio *BovineHD BeadChip*® (400 HD_CPPSE, 13 HD_CNPGC e 31 HD_CNPGC_SIRE), foram removidos 191.630 SNPs (24,63%) e 12 animais (2,70%), resultando em 586.332 SNPs (75,37%) e 432 animais (97,30%).

Como ilustrado na Figura 1, a estrutura desta população foi visualizada com base nos dois principais auto vetores, utilizado o componente principal 1 (PC1) e o componente principal 2 (PC2). A Figura 2 ilustra as relações genéticas dentro das populações estudadas.

Para a imputação, em ambos os cenários, a população de referência foi composta por 332 animais e a população de validação por 100 amostras que foram utilizadas para avaliar a eficiência da imputação por meio da percentagem de genótipos imputados corretamente. No primeiro cenário, a população de imputação foi composta por 468 animais genotipados por meio do *BovineSNP50 v2 BeadChip*®. A média de genótipos imputados corretamente foi de 95%. No segundo cenário, a população de imputação foi composta por 169 animais genotipados por meio do *GeneSeek Genomic Profiler HD*. A média de genótipos imputados corretamente foi de 97%. O número de SNPs em comum nos dois cenários e que foram imputados com acurácia superior a 95% totalizaram 281.004 marcadores, que foram utilizados nas análises de GWAS.

Para AOL, o valor de π (proporção de SNPs que não apresentam efeito sobre a característica) obtido por Bayes $C\pi$ foi de 0,99. Foram encontradas por meio de Bayes B, duas janelas localizadas nos cromossomos 3 e 13 que explicaram 0,24% e 0,20% da

variância genética, respectivamente. Não foram encontradas regiões associadas com AOL por meio de Bayes C.

Para EGS, o valor de π obtido por Bayes C π foi de 0,99. Foram encontradas por meio de Bayes B, quatro janelas localizadas nos cromossomos 1, 6, 7 e 14 que explicaram 0,32%; 0,35%; 0,22% e 0,77% da variância genética, respectivamente. Por meio de Bayes C foi encontrada uma janela localizada no cromossomo 14 que explicou 0,20% da variância genética.

Para PESO, o valor de π obtido por Bayes C π foi de 0,97. Foram encontradas por meio de Bayes B três janelas, sendo duas no cromossomo 6 e uma no cromossomo 4, que explicaram 6,27%; 0,19% e 0,87% da variância genética, respectivamente. Não foram encontradas regiões associadas com PESO por meio de Bayes C.

As estimativas obtidas para Bayes B e Bayes C para cada janela de SNPs associada com as características estudadas, localização e número de SNPs abrangidos encontram-se na Tabela 3. Os efeitos dos marcadores foram apresentados em gráficos do tipo "*Manhattan plot*" (Figuras 3, 4, 5, 6, 7 e 8).

4. DISCUSSÃO

A proporção estimada da variância fenotípica explicada pelos marcadores ou “herdabilidade genômica”, para todas as características foram de moderadas a altas (Tabela 2). Estes resultados estão de acordo com trabalhos realizados na raça Canchim, que obtiveram estimativas de herdabilidade para peso ao sobreano variando entre 0,23 a 0,54 (Mascioli et al., 1996; Castro-Pereira et al., 2007; Buzanskas et al., 2010; Meirelles et al., 2010; Borba et al., 2011); para AOL variando entre 0,33 a 0,39 (Meirelles et al., 2009; Meirelles et al., 2010; Mokry et al., 2012) e para EGS variando entre 0,04 a 0,28 (Meirelles et al., 2010; Mokry et al., 2012).

Os resultados da PCA e do método de VanRaden (2007) demonstram ausência de subgrupos populacionais distintos. Nas abordagens Bayesianas utilizadas, a estratificação populacional é implicitamente contabilizada ajustando todos os marcadores simultaneamente (Zeng 2011). Portanto, os resultados da análise de associação estão livres deste tipo de viés.

Diversos genes foram identificados nas regiões associadas com as características analisadas (Tabela 3). Na análise funcional destacam-se os genes candidatos a seguir.

4.1. Área de olho de lombo

No cromossomo 3, foi identificada a janela 407 onde foi encontrado o gene *SFPQ* (*Splicing factor, proline- and glutamine-rich*, também conhecido como *PSF*), que codifica uma proteína multifuncional, que atua principalmente como fator de transcrição de diversos genes e quando associada a outras proteínas formam os spliceossomos que moderam o “*splicing*” alternativo de pré-mRNAs (Shav-Tal & Zipori, 2002, Urban et al., 2002). Como fator de transcrição, o *SFPQ* liga-se na sequência 5'-CTGAGTC-3' em sítios de ligação do tipo *IGFRE* (*insulin-like growth factor response element*) que ocorre em genes cuja transcrição é estimulada pelo *IGF-I* (*insulin-like growth factor-I*), que por sua vez, é um hormônio produzido no fígado e atua no eixo regulador do crescimento (Boguszewski, 2001).

Dois tipos de fatores de transcrição ligam-se especificamente nestes sítios, o *SFPQ*, apresentando efeito supressor (Urban et al., 2002) e o *Sp1*, que apresenta efeito ativador (Urban et al., 1996).

A presença de sítios de ligação do tipo *IGFRE* foi relatada no gene *δ 1-crystallin* (Alemany et al., 1992) em frangos. Urban et al. (2002) demonstraram a presença de um sítio de ligação *IGFRE* no gene *P450scc* (*P-450 cholesterol side-chain cleavage enzyme*

gene) que atua no processo de esteroidogênese. Também há relatos de que alterações na expressão do gene da elastina em tecido muscular liso de ratos ocorram pela perda do sítio *IGFRE*, dessa forma, suprimindo a expressão gênica (Wolfe, et al., 1993).

No cromossomo 13, foi identificada a janela 1416, onde estão contidos os genes *OR4S2* (*Olfactory Receptor, Family 4, Subfamily S, Member 2*), *TMX4* (*Thioredoxin-Related Transmembrane Protein 4*) e *PLCB1* (*Phospholipase C β 1*) (Tabela 3). O gene *PLCB1* (*Phospholipase C β 1*) expressa uma enzima fundamental na ação mitogênica do *IGF-I* (Manzioli et al., 1997; Xu et al., 2001).

4.2. Espessura de gordura subcutânea

No cromossomo 6, foi identificada a janela 703, onde localiza-se o gene *KCNIP4* (*Kv channel-interacting protein 4*) também conhecido como *KChIP4*, que traduz uma proteína componente de uma família de pequenas subunidades que formam os canais iônicos permeáveis de íons de potássio (*Kv channel*) na parede plasmática de neurônios. Estes canais atuam na excitabilidade neuronal (Holmqvist et al., 2002). Buzanskas et al., (2014) em um estudo de GWAS encontraram um SNP (rs135591504) no gene *KCNIP4* associado a peso a desmama em bovinos da raça Canchim.

No cromossomo 7, foi identificada a janela 820, onde estão contidos os genes *ARL10* (*ADP-Ribosylation Factor-Like 10*), *NOP16* (*NOP16 Nucleolar Protein*), *HIGD2A* (*HIG1 Hypoxia Inducible Domain Family, Member 2A*), *CLTB* (*Clathrin, Light Chain B*), *FAF2* (*Fas Associated Factor Family Member 2*), *RNF44* (*Ring Finger Protein 44*), *CDHR2* (*Cadherin-Related Family Member 2*), *GPRIN1* (*G Protein Regulated Inducer Of Neurite Outgrowth 1*), *SNCB* (*Synuclein, Beta*), *EIF4E1B* (*Eukaryotic Translation Initiation Factor 4E Family Member 1B*), *TSPAN17* (*Tetraspanin 17*), *UNC5A* (*Unc-5 Homolog A (C. Elegans)*), *HK3* (*Hexokinase 3 (White*

Cell)), *UIMC1* (*Ubiquitin Interaction Motif Containing 1*), *ZNF346* (*Zinc Finger Protein 346*) e *FGFR4* (*Fibroblast Growth Factor Receptor 4*) (Tabela 3).

O gene *FAF2* também conhecido como *Ubx8*, atua como um sensor de ácidos graxos de cadeia longa e na síntese de triglicerídeos em mamíferos (Lee et al., 2010). Outro gene candidato identificado nesta mesma janela é o gene *FGFR4*, que de acordo com Huang et al., (2007) desempenha papel essencial na homeostase de lipídios e da glucose. Camundongos com alterações no gene *FGFR4* em dieta normal exibiram características de síndrome metabólica que incluem o aumento de massa do tecido adiposo branco, hiperlipidemia, intolerância à glucose, resistência à insulina, além de hipercolesterolemia (Huang et al., 2007).

4.3. Peso ao sobreano

No cromossomo 6, foram identificadas duas janelas associadas a PESO: na janela 699 estão contidos os genes *ABCG2* (*ATP-binding cassette sub-family G member 2*), *PKD2* (*Polycystin-2*), *SPP1* (*Osteopontin*), *MEPE* (*Matrix extracellular phosphoglycoprotein*), *IBSP* (*Bone sialoprotein 2*), *LAP3* (*Cytosol aminopeptidase*), *MED28* (*Mediator of RNA polymerase II transcription subunit 28*), *FAM184B* (*Family With Sequence Similarity 184, Member B*), *DCAF16* (*DDB1- and CUL4-associated factor 16*), *HCAP-G* (*non-SMC condensin I complex, subunit G (NCAPG)*) e *LCORL* (*Ligand Dependent Nuclear Receptor Corepressor-Like*) (Tabela 3). Trabalhos com bovinos encontraram QTLs na mesma região associados à produção, gordura e proteína no leite (Cohen-Zinder et al., 2005; Olsen et al., 2007); proporção total de ossatura, peso ao nascimento, comprimento ao nascimento (Gutierrez-Gil, et al., 2009), distocia e morte fetal (Olsen et al., 2009).

O loco que abrange os genes *HCAP-G (NCAPG)* e *LCORL* tem sido sugerido como candidato para peso da carcaça (Setoguchi et al., 2009), facilidade de parto (Bongiorni et al., 2012), estatura a idade adulta e sobrevivência do bezerro ao nascimento (Sahana et al., 2015), consumo alimentar, ganho de peso, peso de carcaça quente, área de olho de lombo e menor quantidade de gordura subcutânea (Lindholm-Perry et al., 2011) em bovinos de corte.

Hayes et al., (2008) encontraram uma assinatura de seleção indicando seleção para produção leiteira e produção de carne em animais das raças Holandesa, Angus, Hereford e Limousin. Estes resultados sugerem que QTLs no cromossomo 6 associados com características de produção e composição do leite também podem afetar características de crescimento em bovinos.

Na janela 707, estão contidos os genes *ECSOD (Superoxide dismutase [Cu-Zn])*, *LGI2 (Leucine-Rich Repeat LGI Family, Member 2)*, *SEPSECS (O-phosphoseryl-tRNA(Sec) selenium transferase)*, *ZCCHC4 (Zinc finger CCHC domain-containing protein 4)*, *ANAPC4 (Anaphase Promoting Complex Subunit 4)*, *SLC34A2 (Sodium-dependent phosphate transport protein 2B)*, *SEL1L3 (Sel-1 Suppressor Of Lin-12-Like 3)* e *SMIM20 (Small Integral Membrane Protein 20)* (Tabela 3). Esta região está contida em um QTL previamente identificado associado a proporção total de ossatura em animais cruzados Charolês x Holandês (Gutierrez-Gil, et al., 2009).

Na janela 538, estão contidos os genes *NCAPG2 (Non-SMC Condensin II Complex, Subunit G2)*, *ESYT2 (Extended Synaptotagmin-Like Protein 2)*, *WDR60 (WD Repeat Domain 60)* e *VIPR2 (Vasoactive Intestinal Peptide Receptor 2)* (Tabela 3). O gene *VIPR* também conhecido como *VPAC*, que codifica um receptor do peptídeo intestinal vasoativo (*VIP*) que por sua vez é sintetizado na glândula pituitária. Trabalhos identificaram associação do gene *VIPR2* à obesidade em humanos (Liu et al.,

2010) e gordura no leite em bovinos da raça Simental (Capomaccio et al., 2015). Evangelou et al. (2012) em análise de *vias metabólicas* encontraram associação significativa entre a via metabólica do *VIP* e índice de massa corporal e obesidade.

5. CONCLUSÃO

Para estes conjuntos de dados, foi possível realizar imputação de genótipos com alta acurácia. O cenário a partir do *GeneSeek Genomic Profiler HD®* apresentou maior acurácia quando comparado ao cenário a partir do *BovineSNP50 v2 BeadChip®*.

Os métodos Bayes B e Bayes C permitiram identificar regiões cromossômicas e genes associados com área de olho de lombo, espessura de gordura subcutânea e peso ao sobreano em bovinos da raça Canchim. A identificação de QTLs e genes previamente descritos na literatura reforçam as evidências de que as regiões encontradas por este trabalho estão efetivamente associadas com as características estudadas.

6. REFERÊNCIAS

- Aleman J, Klement JF, De Pablo F. DNA binding factors which interact with the Sp1 site of the chicken δ 1-crystallin promoter are developmentally regulated. *Biochem Biophys Res Commun*. 1992; 183: 659-665. doi:10.1016/0006-291X(92)90533-Q
- Balding DJ. A tutorial on statistical methods for population association studies. *Nat Rev Genet*. 2006; 7: 781-791. doi:10.1038/nrg1916
- Bergmann JAG. Índices Zootécnicos para Produção de Bovinos de Carne. 1999. Disponível:<http://professor.ucg.br/siteDocente/admin/arquivosUpload/8949/material/indices%20zoot%20bov%20corte%20jagb.pdf>. Acessado em 20 Fev 2015.
- Binns D, Dimmer E, Huntley R, Barrell D, O'Donovan C, Apweiler R. QuickGO: a web-based tool for Gene Ontology searching. *Bioinformatics*. 2009; 25: 3045-3046. doi: 10.1093/bioinformatics/btp536
- Boguszewski CL. Genética Molecular do Eixo GH-IGF1. *Arquivos Brasileiros de Endocrinologia & Metabologia*. 2001; 45. doi: <http://dx.doi.org/10.1590/S0004-27302001000100003>
- Bongiorni S; Mancini G, Chillemi G, Pariset L, Valentini A. Identification of a Short Region on Chromosome 6 Affecting Direct Calving Ease in Piedmontese Cattle Breed. *PLoS ONE*. 2012; 7: e50137. doi:10.1371/journal.pone.0050137
- Borba LHF, Rey FSB, Silva LOC, Boligon AA, Alencar MM. Parâmetros genéticos para características de crescimento e reprodução de bovinos da raça Canchim. *Pesquisa Agropecuária Brasileira*. 2011; 46: 1570-1578.
- Buzanskas ME, Grossi DA, Baldi F, Barrozo D, Silva LOC, Torres Júnior RAA, et al. Genetic associations between stayability and reproductive and growth traits in Canchim beef cattle. *Livest Sci*. 2010; 132: 107-112. doi: 10.1016/j.livsci.2010.05.008

1116 Buzanskas ME, Grossi DA, Ventura RV, Schenkel FS, Sargolzaei M, Meirelles SLC, et
 1117 al. Genome-Wide Association for Growth Traits in Canchim Beef Cattle. PLoS ONE.
 1118 2014; 4: e94802. doi:10.1371/journal.pone.0094802
 1119 Castro-Pereira VM, Alencar MM, Barbosa RT. Estimativas de parâmetros genéticos e
 1120 de ganhos direto e indireto à seleção para características reprodutivas e de crescimento
 1121 em um rebanho da raça Canchim. Revista Brasileira de Zootecnia. 2007; 36: 1029-1036.
 1122 doi: 10.1590/S1516-35982007000500007
 1123 Capomaccio S, Milanesi M, Bomba L, Cappelli K, Nicolazzi EL, Williams JL, et al.
 1124 Searching new signals for production traits through gene-based association analysis in
 1125 three Italian cattle breeds. Anim Genet. 2015; doi: 10.1111/age.12303.
 1126 Chib S, Greenberg E. Understanding the Metropolis-Hastings Algorithm. Am Stat.
 1127 1995; 49: 327-335. doi:10.1080/00031305.1995.10476177
 1128 Clayton D. snpStats: SnpMatrix and XSnpmatrix classes and methods. R package
 1129 version 1.16.0. 2014.
 1130 Cohen-Zinder M, Seroussi E; Larkin DM, Loo JJ, Wind AE, Lee JH, et al.
 1131 Identification of a missense mutation in the bovine ABCG2 gene with a major effect on
 1132 the QTL on chromosome 6 affecting milk yield and composition in Holstein cattle.
 1133 Genome Res. 2005; 15: 936-944.
 1134 Copland JA, Pardini AW, Wood TG, Yin D, Green A, Bodenbun YH, et al. IGF-1
 1135 controls GLUT3 expression in muscle via the transcriptional factor Sp1. Biochim
 1136 Biophys Acta. 2007; 1769: 631-640.
 1137 Cunningham F, Amode MR, Barrell D, Beal K, Billis K, Brent S, et al. Ensembl 2015.
 1138 Nucleic Acids Res. 2015; 43: D662–D669. doi: 10.1093/nar/gku1010

1139 Dennis Jr G, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA.
 1140 DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome*
 1141 *Biol.* 2003; 4: 3.
 1142 Evangelou M, Rendon A, Ouwehand WH, Wernisch L, Dudbridge F. Comparison of
 1143 Methods for Competitive Tests of Pathway Analysis. *PLoS ONE*. 2012; 7: e41018.
 1144 doi:10.1371/journal.pone.0041018
 1145 Faleiro FG, Andrade SRM, Reis Junior FB. *Biotecnologia: estado da arte e aplicações*
 1146 *na agropecuária*. 1º edição. Planaltina: Embrapa Cerrados; 2011.
 1147 Fernando R, Garrick D. *GenSel - User Manual for a Portfolio of Genomic Selection*
 1148 *Related Analyses*. Maio 2009. Disponível:
 1149 <http://www.biomedcentral.com/content/supplementary/1471-2105-12-186-s1.pdf>.
 1150 Acessado em 7 abril de 2014.
 1151 Forni S, Federici JF, Albuquerque LG. Tendências genéticas para escores visuais de
 1152 conformação, precocidade e musculatura à desmama de bovinos Nelore. *Revista*
 1153 *Brasileira de Zootecnia*. 2007; 36: 572-577. doi: 0.1590/S1516-35982007000300008
 1154 Gutierrez-Gil B, Williams JL, Homer D, Burton D, Haley CS, Wiener P. Search for
 1155 quantitative trait loci affecting growth and carcass traits in a cross population of beef
 1156 and dairy cattle. *J Anim Sci*. 2009; 87: 24-36. doi: 10.2527/jas.2008-0922
 1157 Habier D, Fernando RL, Kizilkaya K, Garrick DJ. Extension of the bayesian alphabet
 1158 for genomic selection. *BMC Bioinformatics*. 2011; 12: 186. doi: 10.1186/1471-2105-
 1159 12-186
 1160 Hayes BJ, Lien S, Nilsen H, Olsen HG, Berg P, Maceachern S, et al. The origin of
 1161 selection signatures on bovine chromosome 6. *Anim Genet*. 2008; 39: 105-11. doi:
 1162 10.1111/j.1365-2052.2007.01683.x

1163 Holmqvist MH, Cao J; Hernandez-Pineda R, Jacobson MD, Carroll KI, Sung MA, et al.
 1164 Elimination of fast inactivation in Kv4 A-type potassium channels by an auxiliary
 1165 subunit domain. *Proc Natl Acad Sci U S A*. 2002; 99: 1035-1040.
 1166 Hu Z, Fritz ER, Reecy JM. AnimalQTLdb: a livestock QTL database tool set for
 1167 positional QTL information mining and beyond. *Nucleic Acids Res*. 2007; 35: D604-
 1168 D609.
 1169 Huang X, Yang C, Luo Y, Jin C, Wang F, McKeehan WL. FGFR4 Prevents
 1170 Hyperlipidemia and Insulin Resistance but Underlies High-Fat Diet-Induced Fatty
 1171 Liver. *Diabetes*. 2007; 56: 2501-2510.
 1172 Kizilkaya K, Fernando RL, Garrick DJ. Genomic prediction of simulated multibreed
 1173 and purebred performance using observed fifty thousand single nucleotide
 1174 polymorphism genotypes. *J Anim Sci*. 2009; 88: 544-551. doi: 10.2527/jas.2009-2064
 1175 Lee JN, Kim H, Yao H, Chen Y, Weng K, Ye J. Identification of Ubx8 protein as a
 1176 sensor for unsaturated fatty acids and regulator of triglyceride synthesis. *Proc Natl Acad*
 1177 *Sci U S A*. 2010; 107: 21424-21429. doi: 10.1073/pnas.1011859107
 1178 Liu YJ, Guo YF, Zhang LS, Pei YF, Yu N, Yu P, et al. Biological Pathway-Based
 1179 Genome-Wide Association Analysis Identified the Vasoactive Intestinal Peptide (VIP)
 1180 Pathway Important for Obesity. *Obesity*. 2010; 18: 2339-2346. doi:
 1181 10.1038/oby.2010.83
 1182 Lindholm-Perry AK, Sexten AK, Kuehn LA, Smith TP, King DA, Shackelford SD, et
 1183 al. Association, effects and validation of polymorphisms within the NCAPG - LCORL
 1184 locus located on BTA6 with feed intake, gain, meat and carcass traits in beef cattle.
 1185 *BMC Genet*. 2011; 12: 103. doi: 10.1186/1471-2156-12-103
 1186 Luchiari Filho, A. *Pecuária da carne bovina*. 1º edição. São Paulo: Luchiari Filho, A.;
 1187 2000.

1188 Ma P, Brøndum RF, Zhang Q, Lund MS, Su G. Comparison of different methods for
 1189 imputing genome-wide marker genotypes in Swedish and Finnish Red Cattle. *J Dairy*
 1190 *Sci.* 2013; 96: 4666-4677. doi: 10.3168/jds.2012-6316

1191 Manzoli L, Billi AM, Rubbini S, Bavelloni A, Faenza I, Gilmour RS, et al. Essential
 1192 role for nuclear phospholipase C b1 in insulin-like growth factor I-induced mitogenesis.
 1193 *Cancer Research.* 1997; 57: 2137-2139.

1194 Mascioli AS, Alencar MM, Barbosa PF, Novaes AP, Oliveira MCS. Estimativas de
 1195 parâmetros genéticos e proposição de critérios de seleção para pesos na raça Canchim.
 1196 *Revista Brasileira de Zootecnia.* 1996; 24: 72-82.

1197 Meirelles SL, Alencar MM, Oliveira HN, Regitano LCA. Estimativas de parâmetros
 1198 genéticos em função do sexo para características de carcaça de animais da raça
 1199 Canchim. 21 Jul 2009. Disponível:
 1200 <http://www.alice.cnptia.embrapa.br/handle/doc/83265>. Acessado em 20 Fev 2015.

1201 Meirelles SLC, Alencar MM, Oliveira HN, Regitano LCA. Efeitos de ambiente e
 1202 estimativas de parâmetros genéticos para características de carcaça em bovinos da raça
 1203 Canchim criados em pastagem. *Revista Brasileira de Zootecnia.* 2010; 39: 1437-1442.

1204 Meuwissen THE, Goddard ME, Hayes BJ. Prediction of total genetic value using
 1205 genome-wide dense marker maps. *Genetics.* 2001; 157: 1819-1829.

1206 Misztal I, Tsuruta S, Strabel T, Auvray B, Druet T, Lee DH. BLUPF90 and related
 1207 programs (BGF90). 7th World Congress on Genetics Applied to Livestock Production.
 1208 23 Agosto 2002. Disponível: [http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=28-](http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=28-07.pdf)
 1209 [07.pdf](http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=28-07.pdf). Acessado em 07 abril 2014.

1210 Mokry FB, Higa RH, Mudadu MA, Lima AO, Meirelles SLC, da Silva MVGB, et al.
 1211 Genome-wide association study for backfat thickness in Canchim beef cattle using
 1212 Random Forest approach. *BMC Genet.* 2013; 14: 47. doi: 10.1186/1471-2156-14-47

1213 Olsen HG, Hayes BJ, Kent MP, Nome T; Svendsen M, Lien S. A genome wide
 1214 association study for QTL affecting direct and maternal effects of stillbirth and dystocia
 1215 in cattle. *Anim Genet.* 2009; 4: 273-280 doi: 10.1111/j.1365-2052.2009.01998.x

1216 Olsen HG, Nilsen H; Hayes BJ, Berg PR, Svendsen M, Lien S, Meuwissen T. Genetic
 1217 support for a quantitative trait nucleotide in the ABCG2 gene affecting milk
 1218 composition of dairy cattle. *BMC Genet.* 2007; 8: 32.

1219 Onteru SK, Gorbach DM, Young JM, Garrick DJ, Dekkers JCM, Rothschild MF.
 1220 Whole genome association studies of residual feed intake and related traits in the pig.
 1221 *PLoS ONE.* 2012; 8: e61756. doi:10.1371/journal.pone.0061756

1222 Perotto D, Abrahão JJS, Moletta JL. Características Quantitativas de Carcaça de
 1223 Bovinos Zebu e de Cruzamentos Bos taurus x Zebu. *Revista Brasileira de Zootecnia.*
 1224 2000; 29: 2019-2029.

1225 Pérez-Enciso M, Rincón JC, Legarra A. Sequence- vs. chip-assisted genomic selection:
 1226 accurate biological information is advised. *Genet Sel Evol.* 2015; 47: 43. doi:
 1227 10.1186/s12711-015-0117-5

1228 Regitano LCA, Coutinho LL. *Biologia molecular aplicada à produção animal.* 1º
 1229 edição. Brasília: Embrapa Informação Tecnológica; 2001.

1230 Sahana G, Höglund JK, Guldbrandtsen B, Lund MS. Loci associated with adult stature
 1231 also affect calf birth survival in cattle. *BMC Genet.* 2015; 16: 47. doi: 10.1186/s12863-
 1232 015-0202-3

1233 Saatchi M, Garrick DJ, Fernando JL, Boddicker N. Comparison of Different Bayesian
 1234 Methods for QTL Mapping in Hereford Beef Cattle Using 1 Mb Windows. *Plant and*
 1235 *Animal Genome Conference XX.* 18 Jan 2012. Disponível:
 1236 <https://pag.confex.com/pag/xx/webprogram/Paper3217.html>. Acessado em: 22 maio
 1237 2015.

1238 Sargolzaei M, Chesnais JP, Schenkel FS. A new approach for efficient genotype
 1239 imputation using information from relatives. BMC Genomics. 2014; 15: 478. doi:
 1240 10.1186/1471-2164-15-478
 1241 Setoguchi K, Furuta M, Hirano T, Nagao T, Watanabe T, Sugimoto Y, Takasuga A.
 1242 Cross-breed comparisons identified a critical 591-kb region for bovine carcass weight
 1243 QTL (CW-2) on chromosome 6 and the Ile-442-Met substitution in NCAPG as a
 1244 positional candidate. BMC Genet. 2009; 10: 43. doi: 10.1186/1471-2156-10-43
 1245 Shav-Tal Y, Zipori D. PSF and p54(nrb)/NonO—multi-functional nuclear proteins.
 1246 FEBS Lett. 2002; 531: 109-114.
 1247 Soller M. The use of loci associated with quantitative effects in dairy cattle
 1248 improvement. Anim Prod. 1978; 27: 133-139.
 1249 Soller M, Plotkin-Hazan J. The use marker alleles for the introgression of linked
 1250 quantitative alleles. Theor Appl Genet. 1977; 51: 133-137. doi: 10.1007/BF00273825
 1251 Sollero BP, Gomes CCG, Roso VM, Higa RH, Yokoo MJ, Cardoso LL, et al.
 1252 Goodness of Fit Comparisons among Five Bayesian Models in Genome-Wide
 1253 Association of Tick Resistance in Brazilian Hereford and Braford Beef Cattle. 10th
 1254 World Congress of Genetics Applied to Livestock Production. 4 Dez 2014. Disponível:
 1255 <http://www.alice.cnptia.embrapa.br/alice/handle/doc/1003693>. Acessado em: 15 maio
 1256 2015.
 1257 Tierney L. Markov Chains for Exploring Posterior Distributions. Ann Stat. 1994; 22:
 1258 1701-1728.
 1259 Tupy O, Primavesi O, Barbosa PF. Avaliação dos impactos econômicos sociais e
 1260 ambientais de tecnologias da Embrapa Pecuária Sudeste: Utilização de touros da raça
 1261 canchim em cruzamentos terminal com fêmeas da raça nelore. Série Documentos 54.
 1262 Dez 2006. Disponível:

1263 <http://ainfo.cnptia.embrapa.br/digital/bitstream/CPPSE/16765/1/documentos-54.pdf>.
 1264 Acessado em 10 Fev 2015.

1265 Urban RJ, Bodenbunrg YH, Wood TG. NH₂ terminus of PTB-associated splicing factor binds to
 1266 the porcine P450scc IGF-I response elemento. *Am J Physiol Endocrinol Metab*.
 1267 2002; 283: 423-427. doi: 10.1152/ajpendo.00057

1268 VanRaden PM. Efficient Methods to Compute Genomic Predictions. *J Dairy Sci*. 2007;
 1269 91: 4414-4423. doi: 10.3168/jds.2007-0980

1270 Vianna AT, Gomes FP, Santiago M. Formação do Gado Canchim pelo Cruzamento
 1271 Charolês-Zebu. 1º edição. São Paulo: Nobel; 1978.

1272 Xu A, Suh PG, Marmy-Conus N, Pearson RB, Seok OY, Cocco L, et al.
 1273 Phosphorylation of Nuclear Phospholipase C b1 by Extracellular Signal-Regulated
 1274 Kinase Mediates the Mitogenic Action of Insulin-Like Growth Factor I. *Mol Cell Biol*.
 1275 2001; 21: 2981-2990.

1276 Wolfe BL, Rich CB, Goud HD, Terpstra AJ, Bashir M, Rosenbloom J, et al. Insulin-like
 1277 growth factor-I regulates transcription of the elastin gene. *The Journal of Biological*
 1278 *Chemistry*. 1993; 268: 12418-12426.

1279 Zeng J. Genomic selection of purebred animals for crossbred performance under
 1280 dominance. Dissertação de mestrado submetida à Iowa State University. 2011.
 1281 Disponível em: <http://lib.dr.iastate.edu/cgi/viewcontent.cgi?article=1472&context=etd>.

1282 Tabela 1. Estatística descritiva para área de olho de lombo (AOL), espessura de gordura
1283 subcutânea (EGS) e peso ao sobreano (PESO) em bovinos da raça Canchim.

Característica	Nº de observações	Média	Desvio Padrão	Valor mínimo	Valor máximo
AOL	1893	77,22	9,15	38,93	108,89
EGS	1893	3,25	0,97	0,11	8,89
PESO	1893	415,60	48,61	290,42	596,29

1284 Tabela 2. Estimativa dos parâmetros genéticos e componentes de (co)variância para área
1285 de olho de lombo (AOL), espessura de gordura subcutânea (EGS) e peso ao sobreano
1286 (PESO) em bovinos da raça Canchim.

Característica	Método	Herdabilidade	Variância genética	Variância residual	Variância total
AOL	REML*	0,38	18,14	29,10	47,24
	Bayes B	0,32	23,13	47,24	70,38
	Bayes C	0,58	43,82	31,38	75,20
EGS	REML*	0,52	0,2009	0,1841	0,3850
	Bayes B	0,28	0,2085	0,5211	0,7297
	Bayes C	0,40	0,3070	0,4450	0,7520
PESO	REML*	0,43	560,8	720,7	1281,5
	Bayes B	0,47	669,8	752,2	1422
	Bayes C	0,63	934,9	546,4	1481,4

1287 * REML: Máxima Verossimilhança Restrita.

1288 Tabela 3. Regiões e genes associados com área de olho de lombo (AOL), espessura de gordura subcutânea (EGS) e peso ao sobreano (PESO) em
 1289 bovinos da raça Canchim.

Característica	Teste	Janela	N° de SNPs	Cromossomo	Localização (incio-final) em Mb*	Genes	% Var	PPA* (p>0)
AOL	Bayes B	407	59	3	111,00-111,98	SFPQ, ZMYM1, ZMYM6, ZMYM6NB, DLGAP3, SMIM12, GJA4, GJB3, GJB4, GJB5	0,24	0,516
AOL	Bayes B	1416	68	13	0,12-0,99	OR4S2, TMX4, PLCB1	0,20	0,540
EGS	Bayes B	1553	169	14	53,00-53,99	CSMD3	0,77	0,890
EGS	Bayes B	703	135	6	42,02-42,99	KCNIP4	0,35	0,784
EGS	Bayes B	1	111	1	1,00-1,99	CRYZL1, DONSON, SON, GART, DNAJC28, TMEM50B, IFNGR2, IFNAR1, IL10RB, IFNAR2, OLIG1	0,32	0,690
EGS	Bayes B	820	189	7	39,00-39,99	ARL10, NOP16, HIGD2A, CLTB, FAF2, RNF44, CDHR2, GPRIN1, SNCB, EIF4E1B, TSPAN17, UNC5A, HK3, UIMC1, ZNF346, FGFR4	0,22	0,870
EGS	Bayes C	1553	169	14	53,00-53,99	CSMD3	0,20	0,892
PESO	Bayes B	699	125	6	38,00-38,99	ABCG2, PKD2, SPP1, MEPE, IBSP, LAP3, MED28, FAM184B, DCAF16, HCAP-G, LCORL	6,27	0,996
PESO	Bayes B	538	97	4	120,01-120,61	NCAPG2, ESYT2, WDR60, VIPR2	0,87	0,962
PESO	Bayes B	707	211	6	46,00-46,99	ECSOD, LGI2, SEPSECS, ZCCHC4, ANAPC4, SLC34A2, SEL1L3, SMIM20	0,19	0,998

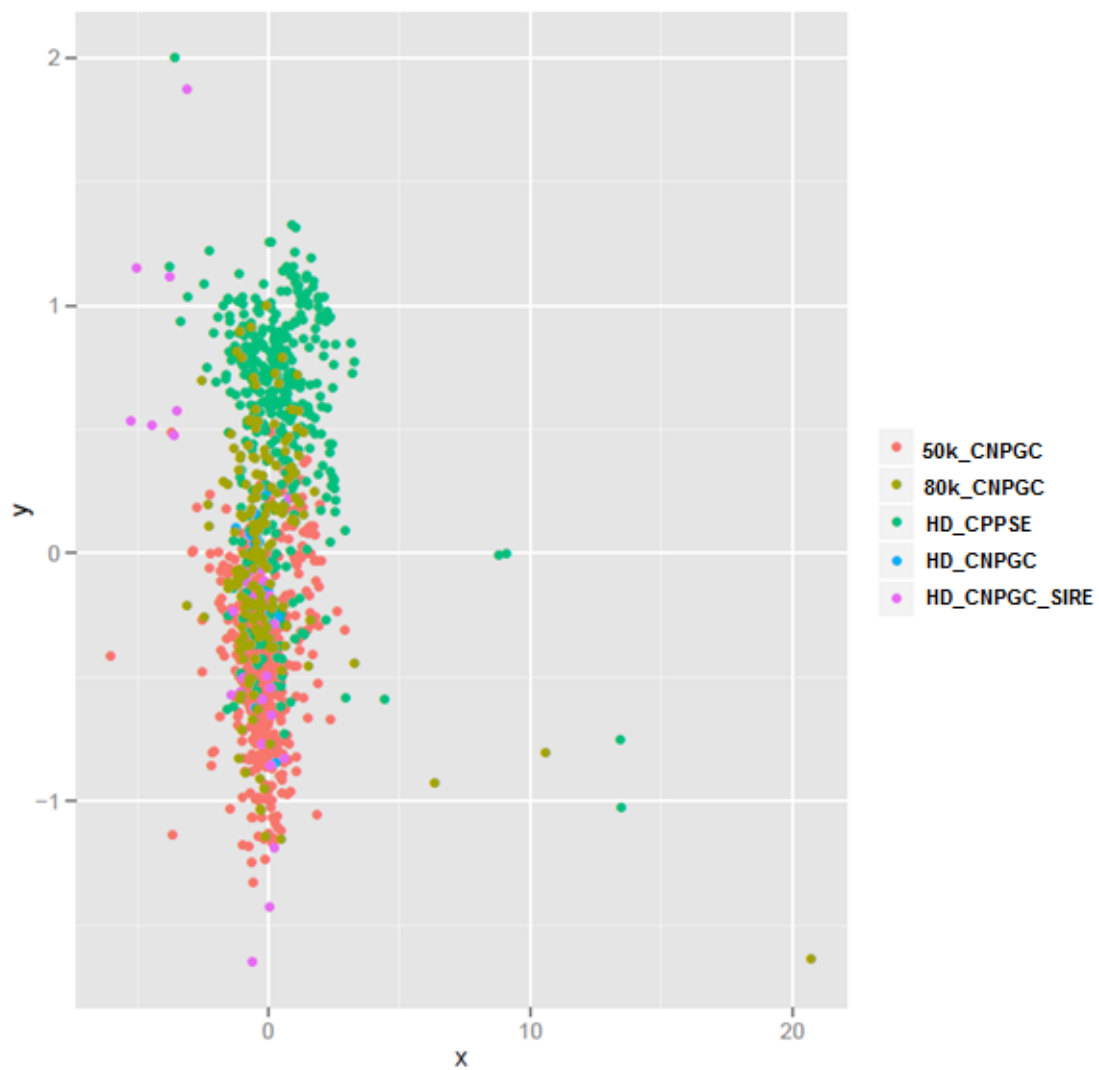
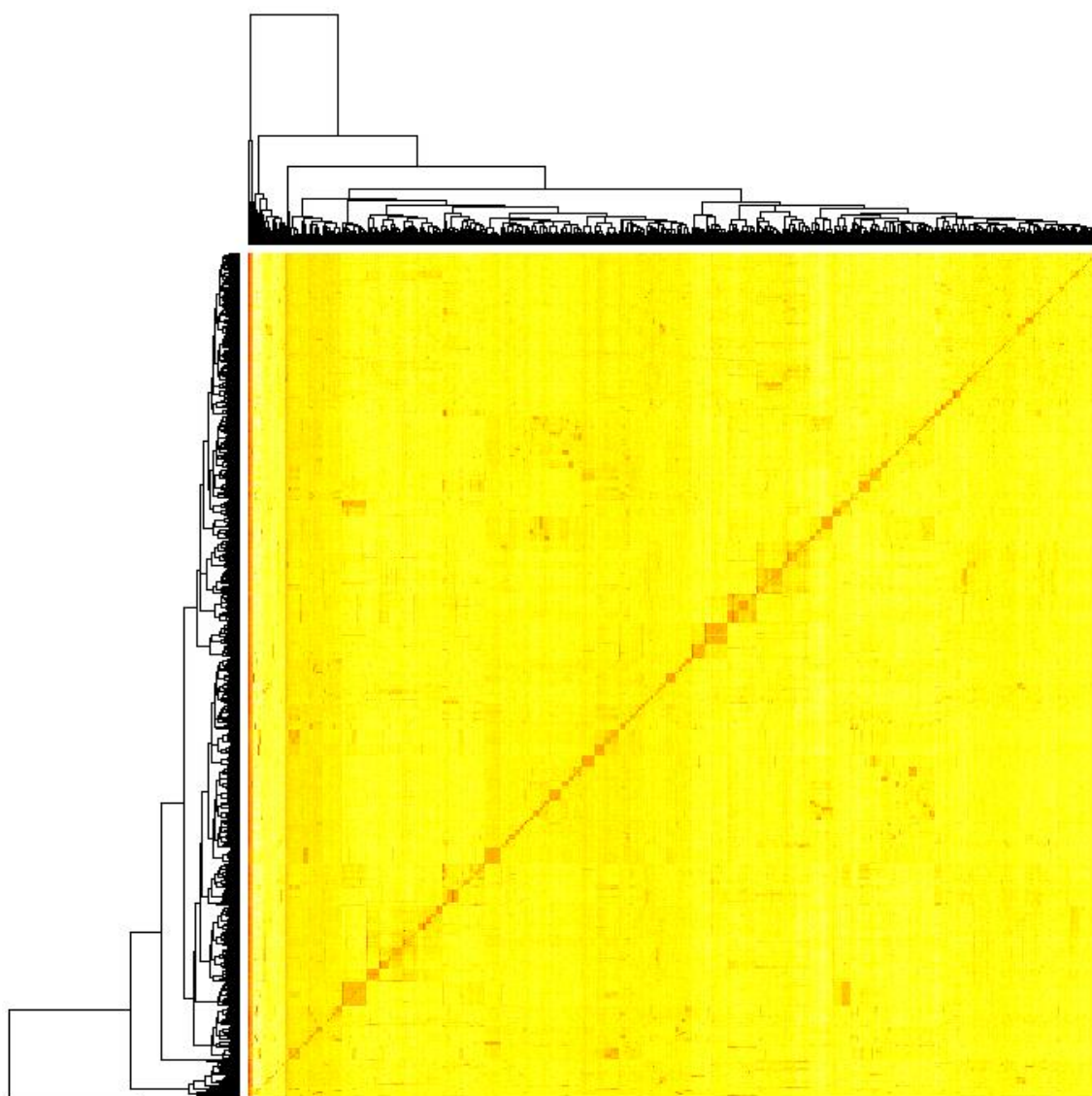
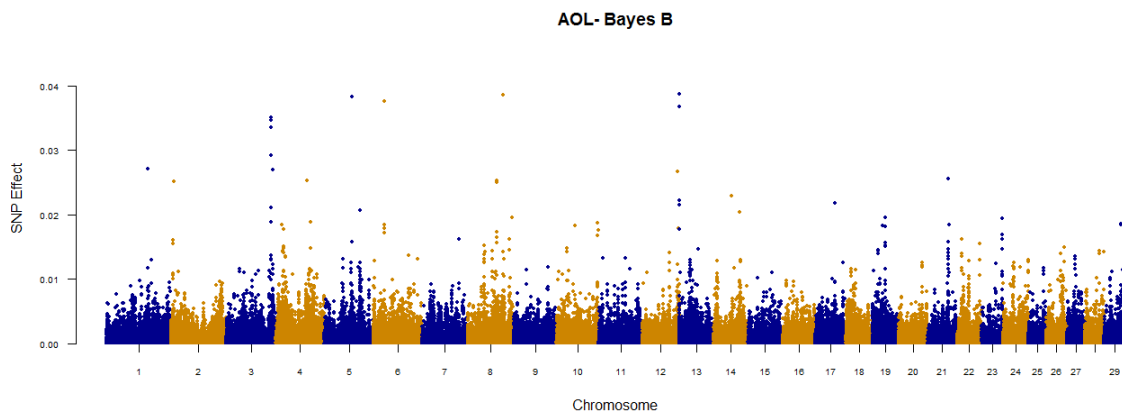


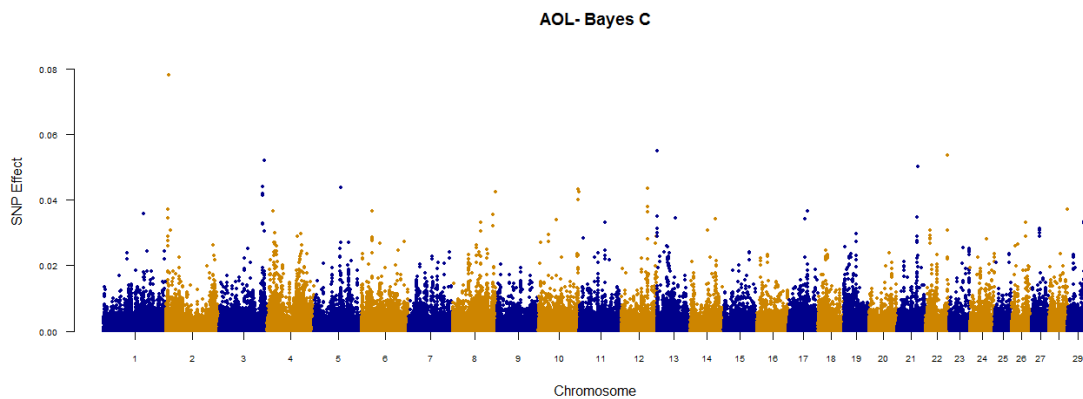
Figura 1. “Scatter plot” da estrutura populacional com base nos resultados de análise de componentes principais (PCA). Os eixos X e Y indicam os componentes principais 1 e 2, respectivamente.



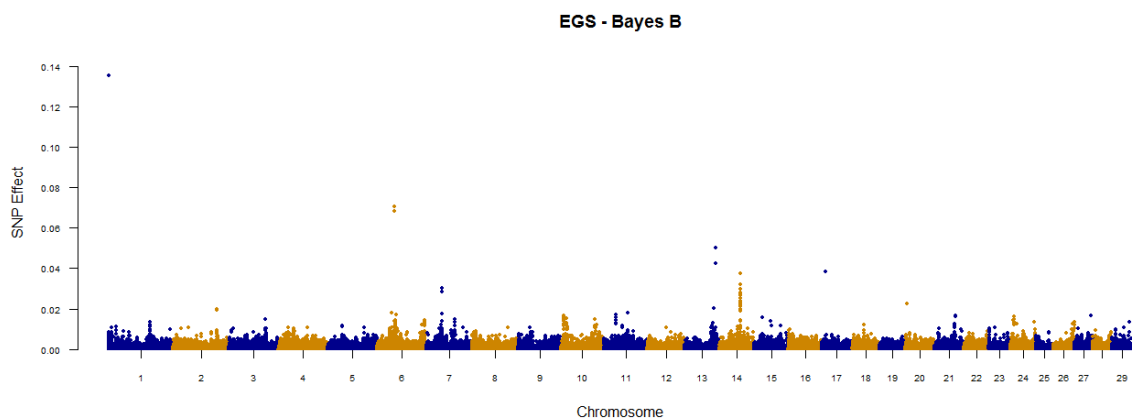
1290 **Figura 2.** “Heatmap” do parentesco entre os indivíduos das populações estudadas.



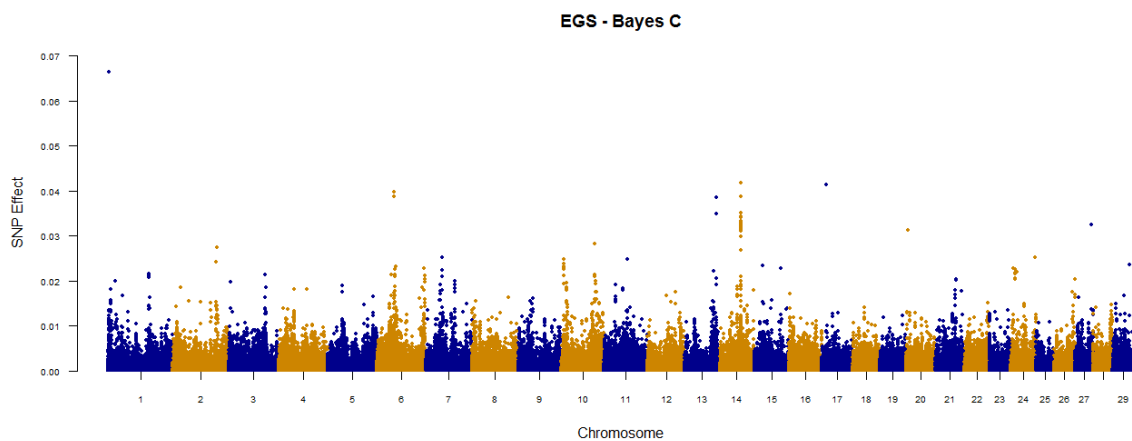
1291 **Figura 3.** “*Manhattan plot*” dos efeitos padronizados do teste de associação ampla do
 1292 genoma por meio de Bayes B com $\pi = 0,99$, para área de olho de lombo (AOL) em
 1293 bovinos da raça Canchim.



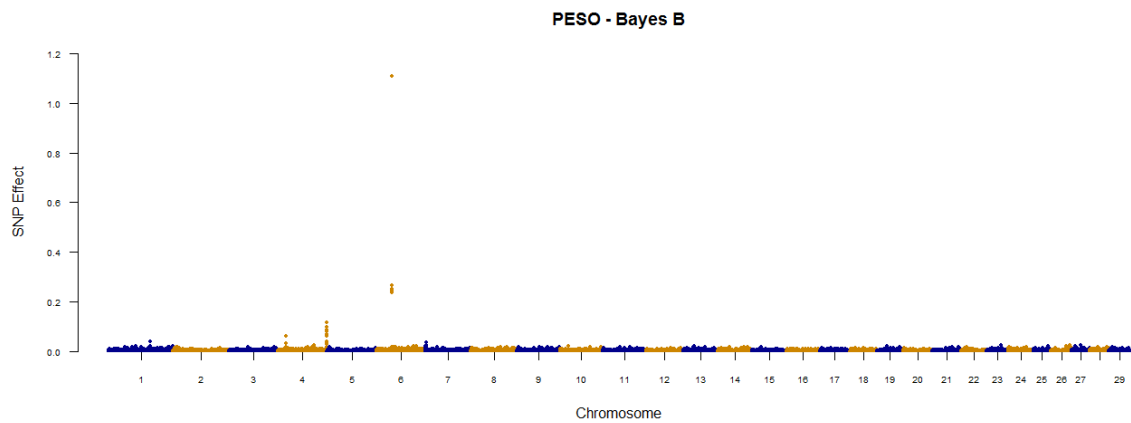
1294 **Figura 4.** “*Manhattan plot*” dos efeitos padronizados do teste de associação ampla do
 1295 genoma por meio de Bayes C com $\pi = 0,99$, para área de olho de lombo (AOL) em
 1296 bovinos da raça Canchim.



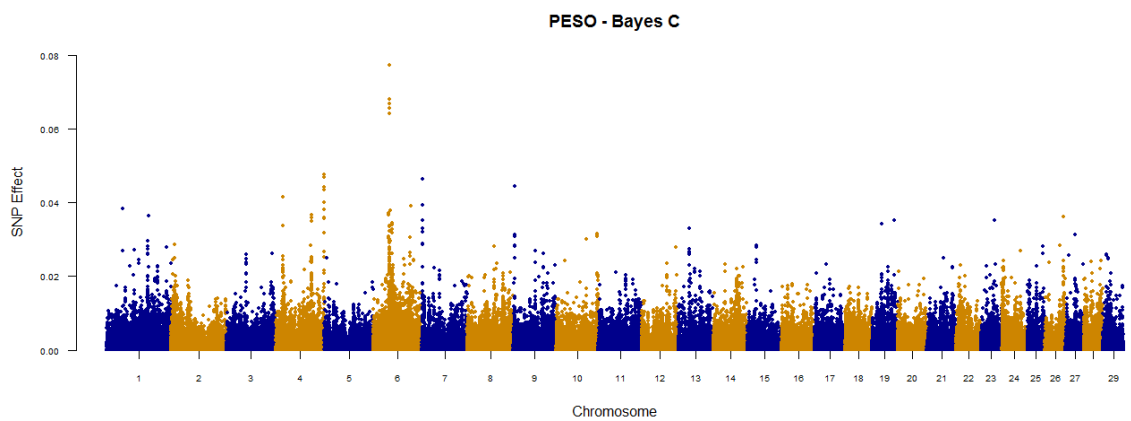
1297 **Figura 5.** “*Manhattan plot*” dos efeitos padronizados do teste de associação ampla do
 1298 genoma por meio de Bayes B com $\pi = 0,99$, para espessura de gordura subcutânea
 1299 (EGS) em bovinos da raça Canchim.



1300 **Figura 6.** “*Manhattan plot*” dos efeitos padronizados do teste de associação ampla do
 1301 genoma por meio de Bayes C com $\pi = 0,99$, para espessura de gordura subcutânea
 1302 (EGS) em bovinos da raça Canchim.



1303 **Figura 7.** “*Manhattan plot*” dos efeitos padronizados do teste de associação ampla do
 1304 genoma por meio de Bayes B com $\pi = 0,97$, para peso ao sobreano (PESO) em bovinos
 1305 da raça Canchim.



1306 **Figura 8.** “*Manhattan plot*” dos efeitos padronizados do teste de associação ampla do
 1307 genoma por meio de Bayes C com $\pi = 0,97$, para peso ao sobreano (PESO) em bovinos
 1308 da raça Canchim.