

UNIVERSIDADE FEDERAL DE PELOTAS
Faculdade de Agronomia Eliseu Maciel
Programa de Pós-Graduação em Zootecnia



Tese

Seleção e associação genômica ampla para características de crescimento e
escores visuais em bovinos da raça Hereford e Braford

Gabriel Soares Campos

Pelotas, 2017

Gabriel Soares Campos

Seleção e associação genômica ampla para características de crescimento e escores visuais em bovinos da raça Hereford e Braford

Tese apresentada ao Programa de Pós-Graduação em Zootecnia da Faculdade de Agronomia Eliseu Maciel da Universidade Federal de Pelotas, como requisito parcial à obtenção do título de Doutor em Zootecnia.

Orientador: Fernando Flores Cardoso

Coorientadora: Arione Augusti Boligon

Pelotas, 2017

C198s Campos, Gabriel Soares

Uso de marcadores moleculares no estudo de associação e seleção genômica ampla para características de crescimento e escores visuais em bovinos das raças Hereford e Braford. / Gabriel Soares Campos ; Fernando Flores Cardoso, orientador ; Arione Augusti Boligon, coorientadora. — Pelotas, 2017.

146 f.

Tese (Doutorado) — Programa de Pós-Graduação em Zootecnia, Faculdade de Agronomia Eliseu Maciel, Universidade Federal de Pelotas, 2017.

1. Bovinos de corte. 2. Modelos bayesianos. 3. Nucleotídeo de polimorfismo único. 4. Predição genômica. 5. Validação cruzada. I. Cardoso, Fernando Flores, orient. II. Boligon, Arione Augusti, coorient. III. Título.

CDD : 636.1

Gabriel Soares Campos

Seleção e associação genômica ampla para características de crescimento e escores visuais em bovinos da raça Hereford e Braford

Tese apresentada como requisito parcial, para obtenção do grau de Doutor em Ciências (Produção Animal), pelo Programa de Pós-Graduação em Zootecnia, Faculdade de Agronomia Eliseu Maciel, Universidade Federal de Pelotas.

Data da defesa: 23/02/2017

Banca Examinadora

Pesq. PhD. Fernando Flores Cardoso (Orientador)
Doutor em Animal Science pela Michigan State University

Prof. D. Sc. José Braccini Neto
Doutor em Zootecnia pela Universidade Federal de Viçosa

Pesq. D. Sc. Marcos Jun-Iti Yokoo
Doutor em Genética e Melhoramento Animal pela Universidade Estadual Paulista
"Júlio de Mesquita Filho"

Prof. D. Sc. Nelson José Laurino Dionello
Doutor em Biotecnologia Agrícola pela Universidade Federal de Pelotas

Pesq. D. Sc. Leandro Lunardini Cardoso
Doutor em Zootecnia pela Universidade Federal do Rio Grande do Sul

AGRADECIMENTOS

Ao Professor e Pesquisador Fernando Flores Cardoso, por ter sido um grande orientador, pelos ensinamentos em pesquisa, pelo incentivo profissional, pela oportunidade e em especial, pela amizade.

À Professora, co-orientadora e grande amiga Arione Augusti Boligon, pela orientação, incentivo, ajuda e pelos momentos de atenção gastos comigo.

Aos pesquisadores da Embrapa Pecuária Sul, Marcos Yokoo e Bruna Sollero, pela ajuda, orientação e amizade.

À Universidade Federal de Pelotas pela oportunidade de realizar o curso de Pós-Graduação.

À Fundação Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela concessão da bolsa de estudos.

Aos colegas e amigos do Laboratório de Bioinformática e Estatística Genômica da Embrapa Pecuária Sul (Labegen) pela amizade, companheirismo, alegria, ajuda e os ótimos momentos de trabalho e confraternização.

Ao meu pai e colega Leonardo Talavera Campos, pela troca de ideias, pelo apoio e principalmente por ter colocado todo seu conhecimento ao meu alcance. À minha mãe, Maria Soares Campos, por todos os momentos de conforto, incentivo e amizade nos momentos difíceis. De forma geral, por sempre me mostrarem o caminho certo e servirem de exemplo para mim.

À minha irmã, Isabel Soares Campos, pela amizade e companheirismo.

À minha namorada, Patrícia Lana Schmidt, pelo amor, carinho, compreensão, apoio e participação em vários momentos, me ajudando e incentivando nos períodos difíceis, sem a tua presença seria tudo mais difícil.

Aos verdadeiros amigos, obrigado pela grande amizade e por entenderem a ausência.

A todos, que de alguma forma contribuíram para realização deste trabalho.

“Seja você quem for, seja qual for a posição social que você tenha na vida, a mais alta ou a mais baixa, tenha sempre como meta muita força, muita determinação e sempre faça tudo com muito amor e com muita fé em Deus, que um dia você chega lá. De alguma maneira você chega lá. ”

Ayrton Senna

RESUMO

CAMPOS, Gabriel Sares. **Seleção e associação genômica ampla para características de crescimento e escores visuais em bovinos das raças Hereford e Braford**. 2017. 146f. Tese (Doutorado) – Programa de Pós Graduação em Zootecnia. Universidade Federal de Pelotas, Pelotas.

Esta tese foi estruturada em três capítulos. No primeiro, foram estimados os parâmetros genéticos usando modelos lineares e de limiar para as características de escores visuais e também utilizada a validação cruzada e regressão multinomial para validação dos modelos. Não houve diferença na estimação dos parâmetros quando os escores possuem distribuição normal, como para conformação, precocidade, musculatura e tamanho. Os valores de herdabilidade (h^2) variaram de 0,18 a 0,26 com o modelo linear e de 0,19 a 0,29 com o de limiar. No entanto, quando o escore não possui distribuição normal, como umbigo, houve vantagens na utilização do modelo de limiar, com valor de h^2 de 0,42, e de 0,22 com modelo linear. O segundo estudo, teve como objetivo avaliar as acurácias predições genômicas utilizando diferentes métodos, para características de crescimento e escores visuais obtidas na desmama e ao sobreano em bovinos das raças Hereford e Braford. Foram utilizados dados de fenótipos 126.290 animais pertencentes ao programa de melhoramento Conexão Delta G, e um conjunto de 3.592 animais genotipados. Os métodos GBLUP, BayesB e BayesC foram testados e maiores acurácias foram obtidas com os métodos bayesianos. Para as características de crescimento, maior ganho em acurácia em relação ao método tradicional (BLUP) foi com a metodologia BayesB para peso ao nascer (PN), de 23,8%, e para os escores visuais foi para o tamanho ao sobreano (PS), de 29,8% com os métodos BayesB e BayesC. Para as abordagens combinando todas as fontes de informação, maiores ganhos foram obtidos com a metodologia de passo único ssGBLUP. Entre todas as características, para as medidas a desmama o ganho médio foi de 40,7% e para as de sobreano, de 36,7%. Menores acurácias de predição foram observadas nos grupos onde continham somente animais da raça Hereford, indicando que o conjunto de treinamento composto pela sua maioria de animais da raça Braford não irá estimar predições acurada para os animais Hereford no conjunto de validação. O terceiro estudo, teve como objetivo a realização de um estudo de associação genômica ampla (GWAS), utilizando metodologia Bayesiana para identificação de regiões genômicas e SNPs mais representativos associados com as características de crescimento. Foram as janelas mais representativas e os SNPs que explicaram mais do que 20% da variância genética estimada para as características estudadas. Após, essa seleção, os SNPs mais informativos quanto aos parâmetros, frequência do modelo (MF), estatística do tipo t (TL), desequilíbrio de ligação (DL) e frequência do alelo menor (MAF), para serem utilizados em um painel de baixa densidade. As acurácias do painel reduzido foram estimadas a partir da validação cruzada, usando métodos de agrupamento k-médias e aleatório. Maiores estimativas de acurácia foram obtidas para as características medidas a desmama. Maiores ganhos em acurácia podem ser obtidos se um maior número de animais for

genotipados e fenotipados. Estes painéis podem ser úteis para estudos futuros relacionados ao mapeamento fino para a descoberta de variantes causais e são uma alternativa interessante para redução dos custos de genotipagem e implementação da seleção genômica.

Palavras-chave: bovinos de corte, modelos bayesianos, nucleotídeo de polimorfismo único, predição genômica, validação cruzada

ABSTRACT

CAMPOS, Gabriel Soares. **Selection and genome wide association for growth traits and visual scores in cattle Hereford and Braford breeds.** 2017. 146f. Tese (Doutorado) – Programa de Pós Graduação em Zootecnia. Universidade Federal de Pelotas, Pelotas.

This thesis was structured in three chapters. In the first one, the genetic parameters were estimated using linear and threshold models for the traits of visual scores and also the cross validation and multinomial regression were used for validation of the models. There was no difference in the parameter estimation when the scores had normal distribution, such as for conformation, precocity, musculature and size. Heritability values (h^2) ranged from 0.18 to 0.26 with the linear model and from 0.19 to 0.29 with the threshold. However, when the score had no normal distribution, such as navel, there were advantages in using the threshold model, with a h^2 value of 0.42 and a linear model of 0.22. The second study aimed to evaluate the accuracy genomic predictions using different methods, for growth traits and visual scores obtained at weaning and yearling in cattle of the Hereford and Braford breeds. Phenotype data 126,290 animals belonging to the Delta G Connection breeding program and a set of 3,552 genotyped animals were used. The GBLUP, BayesB and BayesC methods were tested and higher accuracy were obtained with Bayesian methods. For the growth traits, greater gain in accuracy compared to the traditional method (BLUP) was with the BayesB methodology for birth weight (BW) of 23.8%, and for the visual scores it was for size at the yearling (SY), of 29.8% with the BayesB and BayesC methods. For the approaches combining all sources of information, greater gains were obtained with the single-step ssGBLUP methodology. Among all the characteristics, for weaning measures, the average gain was 40.7% for the weaning measures and 36.7% for yearling. Lower prediction accuracy was observed in the groups containing only Hereford cattle, indicating that the training set composed of the majority of Braford animals will not estimate accurate predictions for the Hereford in the validation set. The third study aimed to perform a genome wide association study (GWAS) using Bayesian methodology to identify the most representative genomic regions and SNPs associated with growth traits. It was selected the most representative windows and the SNPs that explained more than 20% of the genetic variance estimated for the traits studied. After this selection, the most informative SNPs regarding parameters, model frequency (MF), t-like (TL), linkage disequilibrium (DL) and minor allele frequency (MAF) were used in a panel of low density. Reduced panel accuracy was estimated from cross-validation, using k-means and random clustering methods. Higher accuracy estimates were obtained for weaning characteristics. Greater gains in accuracy can be obtained if more animals are genotyped and phenotyped. These panels may be useful for future studies related to fine mapping for the discovery of causal variants and are an interesting alternative for reducing the costs of genotyping and implementation of genomic selection.

Keywords: bayesian models, cross validation, beef cattle, genomic prediction, single polymorphism nucleotide

LISTA DE FIGURAS

Figura 1. Distribuição das observações dos escores visuais de umbigo, conformação, precocidade, musculatura e tamanho.	45
Figura 2. Estimativa da densidade a posteriori do lado esquerdo e traço da herdabilidade estimada do lado direito para umbigo (U), conformação (C), precocidade (P), musculatura (M) e tamanho (T) para o modelo linear.....	51
Figura 3. Estimativa da densidade a posteriori do lado esquerdo e traço da herdabilidade estimada do lado direito para umbigo (U), conformação (C), precocidade (P), musculatura (M) e tamanho (T) para o modelo de limiar.....	52
Figura 4. Acurácia dos valores genômicos diretos (DGV) preditos a partir da validação cruzada usando agrupamento k-médias com o método GBLUP.....	79
Figura 5. Acurácia dos valores genômicos diretos (DGV) preditos a partir da validação cruzada usando agrupamento k-médias com o método BayesC.	80
Figura 6. Acurácia dos valores genômicos diretos (DGV) preditos a partir da validação cruzada usando agrupamento k-médias com o método BayesB.	80
Figura 7. Acurácia dos valores genéticos aprimorados pela genômica (GEBV) preditos a partir da validação cruzada usando agrupamento k-médias com o método Biv_GBLUP(a).....	86
Figura 8. Acurácia dos valores genéticos aprimorados pela genômica (GEBV) preditos a partir da validação cruzada usando agrupamento k-médias com o método IS_GBLUP.....	87
Figura 9. Acurácia dos valores genéticos aprimorados pela genômica (GEBV) preditos a partir da validação cruzada usando agrupamento k-médias com o método Biv_BayesB(a).....	87
Figura 10. Acurácia dos valores genéticos aprimorados pela genômica (GEBV) preditos a partir da validação cruzada usando agrupamento k-médias com o método IS_BayesB.....	88
Figura 11. Acurácia dos valores genéticos aprimorados pela genômica (GEBV) preditos a partir da validação cruzada usando agrupamento k-médias com o método ssGBLUP.....	88
Figura 12. Gráfico de manhattan plot mostrando os resultados para o estudo de associação genômica ampla realizado pelo método Bayes B ($\pi=0.99$) para peso ao	

nascer. O eixo Y representa a proporção da variância genética explicada por janelas de 1Mb e o eixo X representa os cromossomos que as janelas estão localizadas. 103

Figura 13. Gráfico de manhattan plot mostrando os resultados para o estudo de associação genômica ampla realizado pelo método Bayes B ($\pi=0.995$) para peso a desmama ajustado para os 205 dias. O eixo Y representa a proporção da variância genética explicada por janelas de 1Mb e o eixo X representa os cromossomos que as janelas estão localizadas..... 103

Figura 14. Gráfico de manhattan plot mostrando os resultados para o estudo de associação genômica ampla realizado pelo método Bayes B ($\pi=0.995$) para peso ao sobreano ajustado para os 550 dias. O eixo Y representa a proporção da variância genética explicada por janelas de 1Mb e o eixo X representa os cromossomos que as janelas estão localizadas..... 104

Figura 15. Gráfico de manhattan plot mostrando os resultados para o estudo de associação genômica ampla realizado pelo método Bayes B ($\pi=0.995$) para ganho de peso pós-desmama ajustado para os 345 dias. O eixo Y representa a proporção da variância genética explicada por janelas de 1Mb e o eixo X representa os cromossomos que as janelas estão localizadas..... 104

Figura 16. Acurácia dos valores genômicos diretos (DGV) preditos a partir da validação cruzada usando agrupamento k-médias com o método BayesA_TagSNP. 111

Figura 17. Acurácia dos valores genômicos diretos (DGV) preditos a partir da validação cruzada usando agrupamento k-médias com o método BayesA. 111

Figura 18. Acurácia dos valores genômicos diretos (DGV) preditos a partir da validação cruzada usando agrupamento k-médias com o método BayesB. 112

LISTA DE TABELAS

Tabela 1. Estrutura dos dados para escores visuais e peso ajustado ao sobreano ajustado para os 550 dias de idade (PS550).	45
Tabela 2. Estimativas a posteriori dos componentes de (co)variâncias e herdabilidades ao sobreano utilizando modelo linear-linear.....	52
Tabela 3. Estimativas a posteriori dos componentes de (co)variâncias e herdabilidades ao sobreano utilizando modelo linear-limiar.....	53
Tabela 4. Estimativas a posteriori das correlações genéticas (r_g) para as características estudadas utilizando os modelos de limiar e linear.	55
Tabela 5. Correlações de Spearman entre os valores genéticos obtidos pelos modelos de limiar e linear avaliando diferentes níveis de proporção de touros selecionados, considerando todos os touros (100%) e os melhores 50%, 10%, 5% e 1%.	57
Tabela 6. Coeficientes de regressão (β^1) do desempenho fenotípico (escores) sobre o valor genético da característica, R^2 (pseudo coeficiente de determinação de McFadden) e critério de informação Aikaike (AIC) para cada modelo e característica.	58
Tabela 7. Estrutura do arquivo de dados para as características avaliadas nas raças Hereford e Braford.....	66
Tabela 8. Médias a posteriori e desvio-padrão (DP) das variâncias genéticas aditivas direta (σ_a^2), variâncias genéticas aditivas maternas (σ_m^2), (co)variâncias genéticas aditiva-materna (σ_{am}), variâncias de ambiente permanente materno (σ_{pe}^2) e residuais (σ_e^2), herdabilidades diretas (h_a^2) e maternas (h_m^2) para as características estudadas.....	73
Tabela 9. Número de animais (N), proporção de zebu, coeficiente de endogamia ¹ e parentesco genômico dentro e entre grupos (G_{ij}) para agrupamento k-médias em características medidas a desmama.	74
Tabela 10. Número de animais (N), proporção de zebu, coeficiente de endogamia ¹ e parentesco genômico dentro e entre grupos (G_{ij}) para agrupamento k-médias em características medidas ao sobreano.....	75
Tabela 11. Correlações genéticas do fenótipo ajustado com o valor genômico direto (DGV) predito da validação cruzada ($r_{g_{aa}}$) e coeficientes de regressão (β) usando diferentes métodos para características medidas a desmama.....	78

Tabela 12. Correlações genéticas do fenótipo ajustado com o valor genômico direto (DGV) predito da validação cruzada ($rg_{\alpha\alpha}$) e coeficientes de regressão usando diferentes métodos para características medidas ao sobreano.	79
Tabela 13. Ganhos em acurácia das metodologias ² de estimação de efeitos de marcadores em relação ao método tradicional BLUP nas características medidas a desmama.....	81
Tabela 14. Ganhos em acurácia das metodologias ² de estimação de efeitos de marcadores em relação ao método tradicional BLUP nas características medidas ao sobreano.	82
Tabela 15. Correlações genéticas do fenótipo ajustado com o valor genético aprimorado pela genômica (GEBV) predito da validação cruzada ($rg_{\alpha\alpha}$) e coeficientes de regressão (β) usando diferentes métodos para características medidas a desmama.	84
Tabela 16. Correlações genéticas do fenótipo ajustado com o valor genético aprimorado pela genômica (GEBV) predito da validação cruzada ($rg_{\alpha\alpha}$) e coeficientes de regressão (β) usando diferentes métodos para características medidas ao sobreano.....	85
Tabela 17. Ganhos em acurácia das metodologias ² de aprimoramento genômico em relação ao método tradicional BLUP nas características medidas a desmama.	89
Tabela 18. Ganhos em acurácia das metodologias ² de aprimoramento genômico em relação ao método tradicional BLUP nas características medidas ao sobreano.	90
Tabela 19. Grupos de animais de acordo com a estratégia de validação cruzada, K-médias (cl) ou aleatório (rand), média a posteriori da proporção de variância explicada pelos marcadores (h^2) considerando o método BayesB ($\pi=0,995$) e BayesA, o número de janelas top (TopW), número de SNPs em cada janela (TopSNPs) e o número de SNPs selecionados (TagSNPs).....	109
Tabela 20. Correlações genéticas do fenótipo ajustado com o valor genômico direto (DGV) predito da validação cruzada usando diferentes métodos para características de crescimento medidas a desmama e ao sobreano.	110
Tabela 21. Os termos MeSH significativamente sobre representados com genes estatisticamente associados com a característica de peso ao nascer (PN).	114

Tabela 22. Os termos MeSH significativamente sobre representados com genes estatisticamente associados com a característica de peso a desmama ajustado para os 205 dias (PD205)..... 115

Tabela 23. Os termos MeSH significativamente sobre representados com genes estatisticamente associados com a característica de peso ao sobreano ajustado para os 550 dias (PS550)..... 116

Tabela 24. Os termos MeSH significativamente sobre representados com genes estatisticamente associados com a característica de ganho pós-desmama ajustado para os 345 dias (GPD345)..... 117

SUMÁRIO

1. INTRODUÇÃO GERAL.....	18
2. REVISÃO DE LITERATURA	21
2.1 Seleção Genômica.....	21
2.2 Validação na seleção genômica.....	23
2.3 Métodos para estimar dos efeitos dos marcadores.....	25
2.4 Combinação das avaliações genômicas com a tradicional	28
2.5 Associação Genômica Ampla	29
2.6 Características de crescimento e escores visuais.....	32
2.6.1 Características de crescimento.....	32
2.6.2 Características de escores visuais.....	34
3. OBJETIVOS	38
3.1 Objetivos gerais	38
3.2 Objetivos específicos	38
4. HIPÓTESES.....	39
5. CAPÍTULO I	40
5.1 INTRODUÇÃO	41
5.2 MATERIAL E MÉTODOS.....	43
5.2.1 Dados e características avaliadas	43
5.2.2 Modelos de análise	45
5.2.3 Validação dos modelos.....	48
5.3 RESULTADOS E DISCUSSÃO	50
5.4 CONCLUSÕES.....	60
6. CAPÍTULO II	61
6.1 INTRODUÇÃO	62
6.2 MATERIAL E MÉTODOS.....	64
6.2.1 Arquivo de dados fenotípicos, genotípicos e pedigree.....	64
6.2.2 Análises e modelo estatístico	65
6.2.3 Predição genômica	67
6.2.4 Incorporação da informação tradicional com a genômica.....	68
6.2.5 Validação cruzada e acurácia de predição	70
6.3 RESULTADOS E DISCUSSÃO	73
6.3.1 Agrupamento da validação cruzada.....	73

6.3.2 Acurácia e viés das predições genômicas	75
6.3.3 Incorporação da informação tradicional com a genômica	82
7. CAPÍTULO III	92
7.1 INTRODUÇÃO	93
7.2 MATERIAL E MÉTODOS.....	95
7.2.1 Arquivo de dados fenotípicos, genotípicos e pedigree.....	95
7.2.2 Análises e modelo estatístico	96
7.2.3 Estudo de associação genômica ampla (GWAS) bayesiano	96
7.2.4 Seleção de Janelas e SNPs (Tag SNPs).....	98
7.2.5 Acurácia de predição dos painéis de SNPs selecionados	99
7.2.6 Análise de Enriquecimento Funcional.....	101
7.3 RESULTADOS E DISCUSSÃO	102
7.3.1 Estudo de associação genômica ampla (GWAS)	102
7.3.2 Seleção de marcadores moleculares (Tag SNPs)	106
7.3.3 Acurácia de predição dos painéis de SNPs selecionados	108
7.3.4 Enriquecimento Funcional	113
7.4 CONCLUSÕES	118
8. CONCLUSÃO GERAL	119
9. REFERÊNCIAS BIBLIOGRÁFICAS.....	121
ANEXOS	138

1. INTRODUÇÃO GERAL

A pecuária brasileira destaca-se pela importância socioeconômica no Brasil e no mundo por possuir o maior rebanho comercial e segundo maior produtor mundial de carne bovina. Além disso, a atividade contribui com grande parte do produto interno bruto (PIB) do país. O agronegócio corresponde a 21% do PIB total do Brasil (R\$ 1,26 trilhão), sendo que a pecuária é responsável por R\$ 400,7 bilhões, ou seja, 30% do PIB do agronegócio brasileiro (ABIEC, 2016). No entanto, a produção de carne bovina sofre grande concorrência de outros setores que utilizam em suas criações adoção de tecnologias avançadas, como suínos e aves.

Para não perder mercado para outros alimentos de origem animal, é imprescindível a melhoria da eficiência econômica e produtividade dos sistemas de produção em bovinos de corte. A aplicação de tecnologia, como o melhoramento genético animal, é uma das formas de aumentar o potencial produtivo dos animais, com a vantagem de ser de menor custo e gerar ganhos cumulativos ao longo dos anos, em relação às melhorias no ambiente de criação dos animais, como por exemplo, a nutrição, sanidade e o manejo (CARDELLINO & ROVIRA, 1987).

A utilização do melhoramento genético, através da seleção ou/e cruzamentos, irá permitir também a obtenção de animais eficientes nos mais diversos ambientes e em condições de atender as exigências do mercado consumidor (KOURY FILHO, 2009). As raças Hereford e Braford, possuem características desejadas atualmente à maioria dos sistemas de produção (ABHB, 2017), e de acordo com a ASBIA (2015), vem ocorrendo um aumento na utilização de raças britânicas, tanto do rebanho puro como para utilização em cruzamentos.

A tendência atual do mercado é buscar animais com maior velocidade de crescimento, os quais permanecem menos tempo em pastagens ou confinamentos, encurtando o ciclo de produção, possibilitando maior retorno econômico. Para isso, a seleção de animais com maiores pesos em idades jovens é torna-se uma prática importante (BOLIGON et al., 2009). As características de crescimento, amplamente utilizadas por programas de melhoramento, são um indicativo do ganho de peso, possuem a facilidade de mensuração e apresentarem herdabilidades de moderada a alta magnitude. Por outro lado, visando selecionar animais que também atendam às exigências de mercado, com boas características de carcaça como o acabamento precoce sem um tamanho maduro excessivo, os escores visuais de conformação,

precocidade e musculatura são uma boa alternativa (CARDOSO et al., 2001). Estas características vêm sendo utilizadas amplamente como critérios de seleção por diversos programas de melhoramento de bovinos de corte no Brasil.

O melhoramento genético dessas características, tradicionalmente, tem sido realizado com base na metodologia dos modelos mistos, incorporando dados de genealogia e do fenótipo do animal e de seus parentes para a obtenção dos valores genéticos dos indivíduos através da metodologia BLUP (Best Linear Unbiased Prediction) de HENDERSON (1975). Avanços consideráveis em produtividade provenientes do trabalho de avaliação genética pelos métodos tradicionais tem sido obtidos desde a sua implementação nos programas de melhoramento genéticos das diversas espécies.

No entanto, para escores visuais que são características categóricas, modelos lineares e de limiar estão sendo utilizados na estimação de parâmetros e avaliação genética (BOLIGON, 2016). Na avaliação genética de escores visuais em que os dados não apresentam uma distribuição normal, alguns cuidados devem ser considerados, uma vez que as metodologias de modelos lineares, na maioria das vezes, resultam em baixas estimativas de herdabilidade (MWANSA et al., 2002). Deste modo, recomenda-se o uso de modelos de limiar por detectar melhor a variabilidade genética quando comparados aos modelos lineares (MARCONDES et al., 2005).

Mais recentemente, a incorporação das informações moleculares como uma ferramenta adicional para melhorar a eficiência dos processos de seleção, foi denominada seleção assistida por marcadores (SAM) (LANDE & THOMPSON, 1990).

Com o avanço nas tecnologias de sequenciamento de DNA e genotipagem, e com a consequente detecção de milhares de polimorfismo de um único nucleotídeo (SNPs), tornou-se possível o desenvolvimento de novas metodologias para se estimar os efeitos destes marcadores. MEUWISSEN et al. (2001) apresentaram um novo método de seleção no qual se inclui a informação dos dados fornecidos pelos marcadores moleculares cobrindo todo genoma para predição do mérito genético, denominado como seleção genômica (SG), dessa forma permitiu-se um aumento na acurácia dos valores genéticos estimados e a redução do intervalo entre gerações (GODDARD & HAYES, 2009).

Um dos desafios para a efetiva implementação da seleção genômica em programas de avaliação e melhoramento genético, é a identificação de estratégias

apropriadas para combinar os resultados das avaliações genéticas tradicionais com os resultados das avaliações genômicas. Procedimentos multipassos (*multistep*) e passo único (*singlestep*) foram propostos e estão sendo testados e comparados por diferentes grupos de pesquisadores (VANRADEN et al., 2009; HAYES et al., 2009; MISZTAL et al., 2009; AGUILAR et al., 2010).

As informações dos marcadores moleculares além de contribuir para a obtenção de valores genéticos mais acurados, podem auxiliar a identificação de SNPs associados com a característica de interesse e explorar as regiões do genoma em que esses SNPs estão localizados, buscando genes e/ou loci de características quantitativas (QTL – Quantitative Trait Loci) relacionados à expressão fenotípica da característica (UTSUNOMIYA et al., 2013). Esse tipo de estudo é chamado de associação genômica ampla (GWAS).

Métodos GWAS podem ser empregados para detecção de QTLs, na construção de redes de genes e seleção genômica (EGGEN 2012; SNELLING et al., 2012). A identificação dos marcadores mais significativos associados às características de interesse, pode levar ao delineamento de painéis de marcadores de baixa densidade explicando assim boa parte da variância genética desses caracteres tornando-se soluções altamente desejáveis, com baixo custo e de extrema importância para a adoção da seleção genômica pela maioria dos setores de produção animal (VAN EENENNAAM et al., 2014).

Portanto, diante da importância da aplicação dessas tecnologias (seleção genômica e/ou GWAS), o presente trabalho busca contribuir para o conhecimento destas ferramentas nas raças Hereford e Braford e colaborar para a implementação destas técnicas em programas de melhoramento genético animal no Brasil.

2. REVISÃO DE LITERATURA

2.1 Seleção Genômica

O melhoramento genético de características de importância econômica, tradicionalmente tem sido realizado com base na metodologia dos modelos mistos, incorporando dados de genealogia e do fenótipo do animal e de seus parentes para obtenção dos valores genéticos dos indivíduos através do BLUP (Best Linear Unbiased Prediction; HENDERSON, 1975). A aplicação de métodos tradicionais para identificar e acasalar indivíduos com valor genético superior, através de dados fenotípicos e de parentesco, tem consistentemente gerado ganhos genéticos anuais médios da ordem de 1 a 2% na maioria das características produtivas avaliadas pelos programas de melhoramento do Brasil e do mundo (CARDOSO et al., 2013).

A incorporação das informações moleculares, como uma ferramenta adicional para melhorar a eficiência dos processos de seleção, tem sido implementada na indústria de carne bovina a algum tempo, por meio da seleção assistida por marcadores (SAM) (LANDE & THOMPSON, 1990; DEKKERS & HOSPITAL, 2002). Porém, a SAM tem se baseado na detecção de alguns poucos marcadores e, salvo algumas exceções, não tem proporcionado ganhos genéticos significativos em relação a seleção tradicional em programas de melhoramento (DEKKERS, 2004). Isso se deve ao fato de que, geralmente, as características de importância econômica são controladas por muitos pares de genes e, portanto, a informação destes poucos marcadores explica somente uma pequena parcela da variância genética total (HAYES, 2012).

Com as inovações nas tecnologias de sequenciamento de DNA e de genotipagem de marcadores moleculares, possibilitaram a detecção de milhares ou milhões de polimorfismo de um único nucleotídeo (SNP, "Single Nucleotide Polymorphism"). Estes SNPs, apresentam altos níveis de automação e acurácia e proporcionaram reduções drásticas nos custos de geração de dados (CAETANO, 2009). Os marcadores SNP tem como base as alterações mais elementares da molécula de DNA, ou seja, mutações em bases únicas da cadeia de bases nitrogenadas (Adenina, Citosina, Timina e Guanina). Além disso, os SNPs são bi-alelicos, podem ocorrer em regiões com função regulatória ou codificadoras e são extremamente abundantes no genoma (CAETANO, 2009). Estudos com humanos, e

com espécies de interesse zootécnico, mostram que pode haver milhões de polimorfismos SNP no genoma de um indivíduo (BOVINE GENOME SEQUENCING AND ANALYSIS CONSORTIUM, 2009). As principais tecnologias disponíveis atualmente para genotipagem de SNPs em bovinos de corte são o chip de genotipagem de SNPs da Illumina para bovinos (BovineSNP50), com capacidade de genotipar 54.609 SNPs e os painéis de alta densidade (HD - *High Density*) Illumina High Density Bovine Bead Chip Array (777,962 SNP) e Affymetrix Axiom Genome Wide BOS 1 Array (648,874 SNP), recentemente disponibilizados.

Esses desenvolvimentos tecnológicos na biologia molecular tornaram possível o desenvolvimento de métodos inovadores de avaliação genética que podem incorporar informações genômicas aos métodos tradicionais de avaliação genética. A inclusão de dados fornecidos pelos marcadores moleculares cobrindo todo genoma no processo de predição do mérito genético, além dos dados de desempenho e de pedigree, no processo denominado seleção genômica (MEUWISSEN et al., 2001), permite um aumento na acurácia dos valores genéticos estimados (GODDARD & HAYES, 2007).

A seleção genômica baseia-se no fato de que os marcadores e o QTL devem estar em suficiente desequilíbrio de ligação (LD), isto é, bastante associados, para garantir que os marcadores irão predizer consistentemente os efeitos do QTL entre gerações e/ou populações (HAYES et al., 2009; CALUS, et al., 2010). Como a posição do QTL geralmente não é conhecida, o LD médio entre marcadores adjacentes é referido como o parâmetro chave para avaliar a precisão com que o valor genético genômico (VGG) pode ser predito, de forma que quanto menor a extensão de LD em uma população, mais denso deve ser o painel marcador para trabalhar com seleção genômica (DE ROSS et al., 2008).

A implementação da seleção genômica compreende três etapas principais: (1) genotipagem de uma população referência (ou população de treinamento), caracterizada fenotipicamente, com conjuntos de SNPs em média e/ou alta densidade e posterior estimativa dos efeitos dos marcadores ou haplótipos (grupos de marcadores); (2) validação dos efeitos estimados em um grupo de animais que não pertence à população referência (população de validação) e, finalmente; (3) a predição dos valores genéticos de indivíduos candidatos à seleção (população de seleção), baseados nos genótipos dos marcadores e dos efeitos estimados desses marcadores da população de referência (HAYES et al., 2009).

2.2 Validação na seleção genômica

Para a predição dos valores genéticos genômicos serem utilizados para seleção, três populações são definidas na seleção genômica: treinamento, validação e seleção. Com base nestas, três estratégias populacionais podem ser empregadas: as populações podem ser fisicamente distintas (i), três populações diferentes; ii) exercer duas funções ao mesmo tempo, ou seja, mesma população é usada para treinamento e validação; iii) exercer três funções ao mesmo tempo, uma só população para treinamento, validação e seleção. De forma geral, a primeira e segunda estratégias são as mais utilizadas (GODDARD & HAYES, 2007).

A população de treinamento, pode ser também definida como de referência ou estimação. Nesta população, um grande número de marcadores é avaliado em um número moderado de indivíduos, os quais devem ter fenótipos avaliados para as características de interesse. Nesta etapa, são descobertos os marcadores que explicam os loci, bem como são estimados seus efeitos (RESENDE et al., 2008).

A população de validação possui conjunto de dados menor que a população de treinamento, e considera indivíduos avaliados para os marcadores e para as características de interesse. Para estimar a acurácia, os valores genéticos são preditos e submetidos a análise de correlação com os fenótipos observados. As equações de predição são testadas para verificar a acurácia nessa amostra independente (RESENDE et al., 2008).

Na população de seleção, há apenas a genotipagem dos animais e as equações de predição derivadas na população de treinamento são utilizadas na predição dos valores genéticos genômicos dos candidatos a seleção. No entanto, a acurácia seletiva, refere-se àquela estimada na população de validação (RESENDE et al., 2008).

A seleção genômica, geralmente tem sido empregada usando a estratégia em que a mesma população é utilizada tanto para estimação, como para validação. Para est tipo de situação, a técnica de validação cruzada tem sido usada para estimação da acurácia. Esta é uma técnica que serve para avaliar como os resultados de uma análise estatística vão generalizar para um conjunto de dados independente (KOHAVI, 1995). A validação cruzada consiste em separar o conjunto de dados em subconjuntos

complementares, realizando a análise em um subconjunto (conjunto de treinamento) e validando em um outro subconjunto (conjunto de validação). Existem vários métodos para particionar o conjunto de dados, dentre eles destacam-se: *holdout* (retenção), *k-fold* (k duplicado), *leave-one-out* (deixar um fora) e *sorteio aleatório* com ou sem repetição (KOHAVI, 1995).

A acurácia do modelo de seleção genômica é determinada pela correlação entre as estimativas dos valores genéticos preditos e os valores genéticos verdadeiros (LEGARRA et al., 2008). Como o valor genético verdadeiro não é conhecido, estimadores da acurácia com a seleção genômica foram propostos por LEGARRA et al., (2008) e HAYES et al., (2009). Os autores definem a correlação $r(\hat{y}, y)$ como a habilidade preditiva (para os dados não observados) e o termo acurácia para a correlação entre o valor genético total de um indivíduo (g) e sua estimativa (\hat{g}). Eles demonstraram que a acurácia pode ser estimada a partir da habilidade preditiva, através da fórmula: $r(g, \hat{g}) = r(y, \hat{y}) / h$. Em que, h é a raiz quadrada da herdabilidade.

Dentre os fatores que afetam a acurácia das predições genômicas, entre eles estão: i) o desequilíbrio de ligação (LD) entre marcadores e QTL nas populações de treinamento e validação; ii) o tamanho da população de treinamento (N); iii) a herdabilidade (h^2) da característica sobre investigação; iv) a arquitetura genética da característica e v) os métodos estatísticos utilizados nas predições dos valores genômicos diretos (VGD) (GODDARD, 2008; HAYES et al., 2009).

Aumentar a densidade dos marcadores irá melhorar a acurácia de predição, pois aumentará o número de QTL que estão em LD com marcadores e conseqüentemente irá capturar mais da variação genética (HEFFNER et al., 2011). A aplicação bem-sucedida da seleção genômica entre gerações depende da persistência da fase do LD entre marcadores e o QTL (DE ROOS et al., 2008), que é provável que seja uma função da relação genética entre populações (DE ROOS et al., 2008; TOOSI et al., 2010). Aumentar o N levará a uma melhor estimativa dos efeitos do SNP e, portanto, aumentará a precisão da predição (HAYES et al., 2009). DAETWYLER et al. (2010), com estudo de simulação, concluíram que as acurácias de predição aumentaram com o aumento da h^2 (herdabilidade) da característica, independentemente do número de QTL que controlam a característica ou do modelo de utilizado na predição dos efeitos dos marcadores. Os modelos de predição podem variar em relação à estimativa da acurácia entre características com diferentes

arquitecturas genéticas. Segundo HAYES et al., (2009), geralmente, a confiabilidade da predição diminui com o aumento da complexidade das características. DAETWYLER et al., (2010), comparando método bayesiano Bayes B e o BLUP, concluíram que a metodologia Bayes B foi mais precisa quando um menor número de loci controlam a característica.

2.3 Métodos para estimar dos efeitos dos marcadores

O modelo conceitual elementar para seleção genômica pode ser representado por:

$$y_i = \mu + \sum_{j=1}^{nHap} x_{ij}g_j + e_i$$

em que, y_i = fenótipo observado do animal i ; μ = média geral; x_{ij} = variável indicadora que relaciona o efeito do haplótipo g_j ao fenótipo observado do animal i ; e e_i é um erro aleatório. O valor genômico de um determinado animal i pode ser predito simplesmente somando-se as estimativas dos efeitos dos haplótipos (ou marcadores) disponíveis:

$$\hat{u}_i = \sum_{j=1}^{nHap} x_{ij}\hat{g}_j$$

Na maioria dos casos, o número de marcadores com efeito sobre a característica em questão excede largamente o número de observações fenotípicas disponíveis, gerando um modelo superparametrizado e, portanto, um grande desafio para a obtenção de estimativas com boas propriedades estatísticas (GIANOLA et al., 2009).

As diferenças nos métodos empregados na seleção genômica decorrem de diferentes pressuposições adotadas a respeito das distribuições dos efeitos dos marcadores para viabilizar a estimativa dos seus efeitos. De forma geral, esses métodos viabilizam a manutenção de todos os efeitos no modelo empregando penalizações via encolhimento das estimativas dos efeitos para zero (XU, 2007), destacando-se: BLUP (MEUWISSEN et al., 2001), regressão de cumeieira (WHITTAKER et al., 2000; GIANOLA et al., 2003), os métodos bayesianos (MEUWISSEN et al., 2001; HABIÉR et al., 2011) e o LASSO bayesiano (PARK & CASELLA, 2008; DE LOS CAMPOS et al., 2009).

Na metodologia do BLUP, os dados são analisados através de modelos mistos clássicos, nos quais os efeitos dos marcadores são considerados como pertencendo a uma mesma distribuição normal e assumindo-se que todos os marcadores possuem a mesma variância σ_g^2 , isto é:

$$g_j \stackrel{i.i.d}{\sim} N(0, \sigma_g^2)$$

Na abordagem Bayesiana, os dados experimentais observados são combinados com informações *a priori* sobre os parâmetros desconhecidos, gerando assim probabilidades *a posteriori* destes parâmetros. No caso da seleção genômica, os métodos bayesianos viabilizam pressuposições de variâncias diferentes para os efeitos de cada conjunto de haplótipos em um dado segmento cromossômico e se todos ou apenas uma fração dos haplótipos tem efeito diferente de zero, o que permite acomodar incertezas nas frações de haplótipos contendo variâncias diferentes de zero e heterogeneidade de variância (MEUWISSEN et al., 2001; GARRICK, 2010). No método Bayes A, a pressuposição é de que todos os marcadores têm efeito sobre a característica considerada, com distribuição *a priori* normal centrada em zero e com variância específica para cada marcador $\sigma_{g_j}^2$, derivada de uma distribuição qui-quadrado invertida escalonada $\chi_{(v, S^2)}^{-2}$, conforme apresentado a seguir:

$$g_j \stackrel{i.i.d}{\sim} N(0, S_{g_j}^2), \quad S_{g_j}^2 \sim C_{(v, S^2)}^{-2}$$

Como resultado da especificação acima, tem-se que a penalização é específica sobre cada efeito estimado. Por outro lado, os métodos Bayes B e Bayes C supõem que uma fração (π_0) dos marcadores não tem efeito e o processo para a determinação do grau de penalização é o que difere entre eles. No método Bayes B, o parâmetro π_0 é definido de forma arbitrária e para essa fração de marcadores a variância, bem como seus efeitos, é nula, enquanto que a fração $1 - \pi_0$ restante segue um modelo de variâncias heterogêneas análogo ao do Bayes A, conforme representado abaixo:

$$g_j \stackrel{i.i.d}{\sim} N(0, \sigma_{g_j}^2), \quad \sigma_{g_j}^2 \sim \begin{cases} \pi_0 : 0 \\ (1 - \pi_0) : \chi_{(v, S^2)}^{-2} \end{cases}$$

O Bayes C é similar ao Bayes B somente com variâncias homogêneas dos efeitos dos marcadores, com se descreve abaixo:

$$g_j \stackrel{i.i.d}{\sim} N(0, \sigma_{g_j}^2), \quad \sigma_{g_j}^2 \sim \begin{cases} \pi_0 : 0 \\ (1 - \pi_0) : \sigma_g^2 \end{cases}$$

Existem ainda outras variações do alfabeto bayesiano, tais como Bayes $\Sigma\pi$ e Bayes $D\pi$, que permitem estimar a fração de marcadores com efeito nulo (π_0) a partir do conjunto de dados (HABIER et al., 2011).

Existem diversos estudos comparando habilidade de predição de diversos métodos. Com dados simulados, HABIER et al. (2007); LUND et al. (2009); CLARCK et al. (2011) apresentaram resultados com melhor acurácia de predição utilizando as metodologias bayesianas. Outros autores (HAYES et al., 2009; HABIER et al. 2011) trabalharam com dados reais e não encontraram diferenças significativas entre os métodos frequentistas e bayesianos. No entanto, NEVES et al. (2014); CARDOSO et al., (2015), em bovinos de corte, compararam os resultados dessas metodologias para características de crescimento e resistência ao carrapato, respectivamente, e concluíram que as metodologias bayesianas foram mais acuradas que o GBLUP. Portanto, não há consenso sobre o melhor método a ser abordado e o importante é estudar a metodologia mais adequada para determinado conjunto de dados.

Nos EUA e na Austrália há algum tempo a seleção genômica vem sendo utilizada para melhorar a acurácia das avaliações genéticas nacionais na raça Angus. Acurácias estimadas pela *American Angus Association*, para as características de crescimento utilizando chip de 50k, foi de 0,64 para o peso ao nascer, de 0,54 para peso a desmama e de 0,66 para o peso ao sobreano (VAN EENENNAAM et al., 2012). Na Austrália (BREEDPLAN - *Australian beef cattle breeding*), as estimativas de acurácias foram de 0,40 para o peso ao nascer e de 0,37 para peso à desmama em bovinos da raça Angus (JOHNSTON et al., 2012).

Nos Estados Unidos, as avaliações do gado Holandês apresentam estimativas do valor genômico desde 2008 e aprimoramentos do processo estão em franco desenvolvimento. A base de dados de animais genotipados com o BovineSNP50 em março de 2017 já conta com mais de 180.000 animais genotipados, proporcionando ganhos médios de aproximadamente 40% na confiabilidade das avaliações de touros jovens para características produtivas (CDCB, 2017).

2.4 Combinação das avaliações genômicas com a tradicional

Um dos desafios para a efetiva implementação da seleção genômica em programas de avaliação e melhoramento genético é a identificação de estratégias apropriadas para combinar os resultados das avaliações genéticas tradicionais com os resultados das avaliações genômicas. Procedimentos multipassos (*multistep*) e passo único (*singlestep*) foram propostos e estão sendo testados e comparados por diferentes grupos de pesquisadores.

Nos procedimentos multipassos (mais de uma etapa), as diferenças esperadas na progênie (DEP) são obtidas através das análises convencionais (com base no fenótipo e no pedigree) e o valor genômico direto (VGD) é obtido com base nas equações de predição (através das informações de fenótipos e genótipos) separadamente. Posteriormente, ambas as fontes de informação (DEP e VGD) são combinadas por meio de diferentes índices ponderados para gerar as DEP Genômicas (DEPG), considerando a acurácia das estimativas obtidas e a herdabilidade da característica (VANRADEN, 2008; VANRADEN et al., 2009; HAYES et al., 2009).

No procedimento de passo único as informações dos SNP são combinadas com o fenótipo e o pedigree para a estimação simultânea dos valores genéticos. O método de uma etapa utiliza uma matriz combinando o parentesco tradicional com base no pedigree com o derivado das informações dos marcadores (MISZTAL et al., 2009; AGUILAR et al., 2010; CHRISTENSEN & MOGENS, 2010). Essa estratégia parece ser uma opção bastante interessante para características e populações que não seguem a estrutura clássica de bovinos de leite (onde muitos touros com avaliações de alta acurácia estão disponíveis), como no caso de frangos (CHEN et al., 2011), suínos (FORNI et al., 2011) e bovinos de corte (CARDOSO et al., 2011).

Uma terceira estratégia combina as informações dos marcadores, disponibilizadas na forma de valores genéticos moleculares, nas avaliações genéticas como uma nova característica correlacionada geneticamente com as demais, para produzir DEPs aprimoradas pela genômica (MACNEIL et al., 2010; GARRICK, 2010). Este método vem sendo utilizado na avaliação de bovinos de corte devido ao modelo de negócios adotado pelas empresas provedoras de serviços genômicos para o setor.

O modelo de trabalho para seleção genômica, deve ser realizado com base em processos que permitam a atualização constante das bases de dados utilizadas para avaliação dos animais, com informações moleculares, de parentesco e fenotípicas.

De acordo com CAETANO et al. (2012), num modelo de trabalho mais adequado, ao contratarem os serviços de genotipagem e de predição genômica, os criadores devem receber os resultados de ambos os procedimentos, ou seja, as informações dos genótipos de todos os marcadores para os indivíduos testados, assim como os valores genéticos calculados utilizando a informação desses marcadores. Com esse modelo, os produtores, além de identificar e selecionar animais superiores, podem também fornecer as informações geradas aos programas de melhoramento nos quais participam, para que estas sejam incorporadas e utilizadas para aprimorar os resultados obtidos, melhorando as acurácias das estimativas obtidas periodicamente (CARDOSO et al., 2013).

De maneira geral, as vantagens da seleção genômica em relação à seleção tradicional podem ser concentradas em três principais pontos: (1) uso da matriz de parentesco real e própria de cada característica, o que aumenta a acurácia de seleção; (2) possibilidade de seleção precoce dos animais, que se reflete em maior ganho genético por unidade de tempo; (3) utilização de mais informações (combinando dados fenotípicos, genotípicos e genealógicos) para correção de dados e realização das análise genômica, o que aumenta a acurácia das estimativas (RESENDE et al., 2010).

2.5 Associação Genômica Ampla

Os recentes avanços que ocorreram nas técnicas de genotipagem de marcadores moleculares, com a detecção de milhares de SNPs, além de contribuir para estimação dos valores genéticos na seleção genômica, também podem ser úteis para identificar loci responsáveis ou “ligados” a determinadas características, em estudos de associação genômica ampla (GWAS) (FERNANDO & GARRICK, 2013).

A GWAS explora a existência de desequilíbrio de ligação (LD – Linkage Disequilibrium) entre SNPs e variantes causais (BOLORMAA et al., 2013), sendo que, quanto maior a densidade dos painéis de marcadores, maiores tendem a ser os níveis de LD (RINCON et al., 2011). De maneira geral, com a GWAS, procura-se identificar SNPs associados com a característica de interesse e explorar as regiões do genoma em que esses SNPs estão inseridos, buscando genes e/ou loci de características

quantitativas (QTL – Quantitative Trait Loci) relacionados à expressão fenotípica da característica (UTSUNOMIYA et al., 2013).

Inicialmente, nas análises de GWAS eram usados métodos que se baseavam em um teste de um único marcador por vez (HIRSCHHORN & DALY, 2005). No entanto, em razão da dificuldade em realizar teste de significância em cada marcador, considerando que existem milhares de SNPs (WANG et al., 2012), outros métodos em que se ajustam os SNPs simultaneamente como efeitos aleatórios são capazes de explicar uma maior parte da variância genética (ONTERU et al., 2010; FAN et al., 2011) em relação aos que testam um marcador por vez. Portanto a escolha da metodologia é muito importante para o sucesso nas análises de GWAS.

Na maioria dos GWAS, o número k de marcadores disponíveis para análise excede largamente o número n de observações. Assim, a análise de regressão múltipla, com base nos quadrados mínimos ordinários não pode ser utilizado para estimar simultaneamente os efeitos de todos os k marcadores. No entanto, métodos de regressão bayesiana, que combinam informações a priori sobre os efeitos destes k marcadores com os fenótipos de características de interesse, podem ser usados para estimar todos os efeitos dos marcadores (MEUWISSEN et al., 2001). Embora esses métodos tenham sido originalmente propostos para a seleção genômica MEUWISSEN et al., (2001), eles também podem ser usados para GWAS (FAN et al., 2011; SUN et al., 2011).

Alguns métodos bayesianos, como por exemplo BayesB, apresentam como vantagem não assumirem a pressuposição de mesma variância para os efeitos de todos os SNP. No entanto, estes métodos não permitem a inclusão de informações de indivíduos com registros de fenótipos, mas não genotipados (WANG et al., 2012). Diante disso, WANG et al. (2012) propuseram um método de ssGBLUP que combina as informações de fenótipo, genótipo e pedigree. Comparando os resultados de diferentes metodologias para GWAS em frangos de corte (WANG et al. 2014), concluíram que, de maneira geral, todos os métodos apresentam alguma desvantagem.

O método BayesB é realizado com base em um modelo de mistura dos SNPs que explicam a variância genética e aqueles SNPs que não explicam a variância genética (π). Enquanto a proporção de SNPs que explica a variância genética pode tornar-se pequena, a variância genética total permanece constante e, assim, é

distribuída entre menos SNPs, resultando no que parece ser uma estimativa inflacionada de variância genética explicada por um SNP (WANG et al., 2014). O método ssGWAS também apresenta algumas limitações. Uma delas, é que o número de iterações é ditado por heurísticas no tempo. Estudos adicionais (inéditos) indicam que a precisão GWAS com ssGWAS é maximizada em 2-4 iterações, com uma única iteração criando parcelas ruidosas e com mais iterações suprimindo sinais de QTLs menores. Outra desvantagem da ssGWAS é a incapacidade de determinar o nível de significância (WANG et al., 2014). WANG et al., (2014) ainda destacaram que, a eficiência do método é influenciada por fatores como a proporção de marcadores que explicam a variância genética da característica, e do número de animais com dados de fenótipo e genótipo disponíveis.

Portanto, as análises de GWAS servem para identificar marcadores associados a uma característica e serem usados para a seleção genômica ou para mapeamento de QTL para uma região cromossômica. Deste modo, é necessário confirmar em populações independentes as associações que foram descobertas em uma população (BOLORMAA et al., 2011). Porém, pode ocorrer que a confirmação não seja bem-sucedida. Os autores destacam que a falta de confirmação de associação, pode ser devida a três razões: 1) a descoberta original foi um falso positivo; 2) a associação é específica para essa raça ou porque o QTL não segrega em outra raça ou porque os níveis de LD difere entre raças; ou 3) há uma falta de poder estatístico na população de descoberta ou validação, ou em ambas as populações.

De maneira ideal, um marcador SNP que está associado a um aumento em alguma característica de interesse na raça 1, estará também associado a um aumento em outras raças (BOLORMAA et al., 2011). No entanto, as raças bovinas diferem nos níveis de LD entre os marcadores no chip SNP 50K (DE ROOS et al., 2008), sendo assim, espera-se que eles diferem nos níveis de LD entre SNP e QTL. Isto pode significar que a associação entre SNP e QTL ainda é significativa em uma segunda raça, mas invertida no sinal. Provavelmente, essa associação seja simplesmente fraca e não significativa. Se o QTL estiver segregando na segunda raça, é provável que um SNP diferente, próximo ao QTL, irá agora mostrar uma associação significativa com o QTL. Portanto, para confirmar a detecção de um QTL, 3 tipos de evidência são investigados: O SNP descoberto esteja associado a característica na população descoberta, significativamente associado com a característica na segunda população

(validação); se a direção da associação é a mesma; e se há outro SNP na mesma região que mostra uma associação significativa com a característica de interesse (BOLORMAA et al., 2011).

Recentemente, várias regiões do genoma bovino associadas a características de crescimento foram identificadas para bovinos de corte. Uma região está no BTA “Bos taurus chromosome” 6, que contém genes candidatos incluindo NCAPG e LCORL (NISHIMURA et al., 2012; SAATCHI et al., 2014; SANTANA et al., 2014, SNELLING et al., 2010). Outra região, localizada no cromossomo BTA 14, contém PLAG1, CHCHD7 e outros genes relacionados, os quais afetam o crescimento em bovinos (NISHIMURA et al., 2012, PAUSCH et al., 2011; SAATCHI et al., 2014) e humanos (PRYCE et al., 2012; UTSUNOMIYA et al., 2013).

Embora muitos marcadores e genes utilizando GWAS já foram identificados, como sendo associados com características complexas, normalmente, a proporção da variância genética explicada por eles tem sido pequena (BOLORMAA et al., 2011). Esse fato pode ser justificado pela própria natureza das características quantitativas, uma vez que as mesmas são afetadas por muitos loci de pequeno efeito (DE LOS CAMPOS et al., 2013).

2.6 Características de crescimento e escores visuais

2.6.1 Características de crescimento

As características de desenvolvimento ponderal normalmente avaliadas pelos programas de melhoramento de bovinos de corte são: peso ao nascer (PN), peso à desmama (PesoD), peso ao sobreano (PesoS) e o ganho pós desmama (GPD).

O desempenho pré-desmama de bovinos é influenciado pelo genótipo do indivíduo (efeito genético direto), pelo genótipo da mãe para caracteres maternos e pelos correspondentes efeitos ambientais materno permanente e temporário (WILLHAM, 1972). O genótipo da mãe afeta o fenótipo dos bezerros através de uma amostra de seus genes e também através de seu genótipo para habilidade materna, principalmente para produção de leite (CARDOSO et al., 2001).

O peso ao nascer é um importante preditor de facilidade de parto. Touros com DEPs muito altas para PN, não são recomendados para novilhas com pequena abertura pélvica. A seleção para PN não deve ser conduzida para os extremos, pois maiores taxas de sobrevivência perinatal são observadas quando PN's estão próximos da média.

Para o PN, estimativas de herdabilidade direta encontradas na literatura são de média a alta magnitude (KOOTTS et al., 1994). CARDOSO et al., (2001) analisando dados de animais Angus criados de forma extensiva, encontraram valores de h^2 de 0,29. CUNDIFF et al. (1975) relataram valores de 0,22 para PN, avaliando animais cruzados Angus x Hereford. SNELLING et al. (1996), trabalhando com animais da raça Hereford, estimaram valores de h^2 de 0,55. Esses valores indicam que, apesar da dificuldade de medir essa característica a campo, os efeitos genéticos diretos são de moderados a altamente herdáveis.

O peso a desmama serve para identificação do potencial de crescimento do animal, juntamente com a habilidade materna da vaca e sua produção leiteira (BIF, 2010). A estimativa de herdabilidade (h^2) para peso a desmama ajustado para os 205 dias (PD205), segundo vários autores indicam ser de baixa a média magnitude para efeito direto. BOLDMAN et al. (1991), para a raça Hereford, estimaram o valor de 0,22 de h^2 para PD205. MEYER (1993) também analisando dados de animais Hereford, estimaram valores entre 0,19 e 0,32. DE MATTOS et al. (2000), estimaram componentes de (co)variância de animais da raça Hereford provenientes de três países, Uruguai, Canadá e USA, e foram encontrados valores de 0,18; 0,16 e 0,16 para h^2 do PD205, respectivamente. Por outro lado, as herdabilidades maternas para PD205 são baixas de acordo com os valores encontrados na literatura para essa característica, variando de 0,10 a 0,16 (CARDOSO et al., 2001; MEYER et al., 2004; BOLIGON et al., 2013). Outro fator interessante em relação a essa característica, é que as covariâncias genéticas entre os efeitos genéticos direto e materno geralmente são negativas (CARDOSO et al., 2001), o que pode ser devido a estimativa do efeito materno como função linear do efeito direto (quando existem touros que não possuem filhas desmamando produtos).

O período pós-desmama é importante na avaliação genética de bovinos de corte por corresponder a uma fase próxima do produto final e por melhor representar

o ambiente de criação, não sendo diretamente influenciado por efeitos maternos (CARDOSO et al., 2004).

CARDOSO et al., (2004) analisando dados de bovinos da raça Angus, encontraram h^2 de 0,20 para o ganho pós-desmama (GPD). CARDOSO & TEMPELMAN (2004) estimaram herdabilidade para o GPD345 de 0,15, usando um modelo multirracial para as raças Hereford e Braford. Trabalhando com animais da raça Brangus – Ibagé, OLIVEIRA et al. (2010), encontraram valores de h^2 de 0,29 para GPD. Para peso ao sobreano (PS550) o valor estimado de herdabilidade encontrado na literatura é de média magnitude (KOURY FILHO et al., 2010) variando entre 0,29 e 0,36. SILVA et al. (2012), estudando animais da raça Brangus, encontraram valores de h^2 de 0,34 para PS550.

2.6.2 Características de escores visuais

Estas características são importantes para identificarmos animais mais equilibrados para não avaliar somente o ganho de peso dos animais. Estes são observados quanto à conformação (C), precocidade (P), musculatura (M), tamanho (T) e umbigo (U).

Para os escores de C, P, M e T, São atribuídas notas, com variação de 1 a 5, na desmama e no sobreano, sempre de forma relativa à média do grupo de contemporâneos (GC). Os escores mais altos indicam presença mais marcante da característica. Em cada GC, deverão existir animais de todas as notas (de 1 a 5), independente de qualquer comparação absoluta com outro grupo ou rebanho da mesma raça. A distribuição ideal das notas é: 40% deverão receber nota 3, 20% nota 2, 20% nota 4, 10% nota 1 e 10% nota 5 (CAMPOS, 2011).

A conformação (C) estima a quantidade de carne na carcaça. Esta característica é influenciada pelo tamanho (principalmente pelo comprimento) e pelo grau de musculosidade. O modo de avaliação é olhar pata o animal vivo, mas imaginar a carcaça do animal, em um frigorífico. É uma estimativa potencial para produção de carne que o animal possui (CARDOSO et al., 2001; CARDOSO et al., 2004; CAMPOS, 2011).

A precocidade de terminação (P) é a capacidade de armazenar gordura, indicando rapidez para atingir acabamento. Ao contrário do proposto por LONG (1973), no sistema Ankony, através da característica ausência de gordura excessiva, avalia-se a capacidade de o animal chegar a um grau de acabamento mínimo de carcaça, com peso vivo não elevado. Animais com maior profundidade de costelas, maior caixa torácica, silhueta cheia, virilhas pesadas e em início de deposição de gordura subcutânea, principalmente na base da cauda, indicam maior precocidade de terminação. Animais altos, esguios, sem caixa, com silhueta de gazela e extremamente enxutos são mais tardios. É importante combinar as precocidades de crescimento (ganho de peso) e de terminação para produzir animais equilibrados e obter a mesma classe de novilho em um menor período de tempo (CARDOSO et al., 2001; CARDOSO et al., 2004; CAMPOS, 2011).

O escore de musculatura (M) é uma avaliação do desenvolvimento de massa muscular como um todo, observada em pontos como antebraço, paleta, lombo, garupa e, principalmente, a largura e profundidade dos quartos traseiros. Analisando-se os animais parados, nota-se que os de musculatura mais desenvolvida apresentam os membros afastados, tanto de frente como de trás. O animal de musculatura forte tem o peito amplo e é mais largo na parte inferior do corpo, recebendo notas mais altas para esta característica. Já o animal de musculatura débil ou fraca em o peito fechado e é mais largo na parte superior do corpo, recebendo notas mais baixas (CARDOSO et al., 2001; CARDOSO et al., 2004; CAMPOS, 2011).

O escore de tamanho (T) compreende o comprimento e, principalmente, a altura do animal. DEPs de T moderado são desejáveis para uma maior e mais eficiente produção de carne em sistemas de ciclo curto. Para a avaliação de tamanho, é utilizado escore 5 para o animal com maior tamanho do grupo avaliado, mas este não é necessariamente o desejado para a raça trabalhada. O escore 1 é destinado ao menor animal do grupo avaliado (CARDOSO et al., 2001; CARDOSO et al., 2004; CAMPOS, 2011).

O escore para tamanho do prepúcio/umbigo (U) é uma característica funcional, por isso, avalia-se o tamanho e o formato do prepúcio e/ou umbigo. Machos com prepúcio muito longo, frequentemente, sofre lesões causadas pela vegetação e têm seu desempenho reprodutivo comprometido. No processo de avaliação do “U”, os animais sempre são analisados em relação a um tamanho e formato de prepúcio/umbigo ideal, e não ao GM. São dadas notas maiores para prepúcio e/ou

umbigo maiores, e notas menores para prepúcio e/ou umbigo menores (CAMPOS et al., 2011).

A estimação de parâmetros e a avaliação genética de escores visuais geralmente são realizadas utilizando-se modelos lineares. Porém, dados categóricos na maioria das vezes não apresentam distribuição normal dos dados e modelos de limiar são mais adequados teoricamente (GIANOLA & FOULLEY, 1983). Os modelos de limiar admitem que existe uma variável subjacente de distribuição contínua, em relação à variável discreta (SORENSEN et al., 1995), e as estimativas referentes a uma determinada categoria são observadas se os valores da escala subjacente estiverem localizados entre os limiares que definem essa categoria. Assim, a distribuição de probabilidade das estimativas, para dados categóricos, depende da posição da média da distribuição subjacente contínua, em relação aos limiares fixos (FARIA et al., 2008).

As estimativas de herdabilidades diretas para os escores visuais encontradas na literatura geralmente são moderadas. CARDOSO et al., (2001) analisando dados de animais da raça Angus na fase de desmama (D) e utilizando modelo linear, encontraram valores de 0,18, 0,19, 0,19 e 0,21 para C, P, M e T, respectivamente, e CARDOSO et al., (2004) avaliando os animais na fase de sobreano (S) estimaram valores de h^2 para C, P, M e T de 0,19, 0,25, 0,26 e 0,24, respectivamente. KOURY FILHO et al. (2010) relataram valores de herdabilidades diretas variando entre 0,13 e 0,25 para os escores de C, P e M na fase de desmama e de 0,24 a 0,32 ao sobreano para as mesmas características também com modelo linear. BOLIGON et al. (2013) estimaram valores de h^2 para os escores na fase de desmama de 0,21 (C), 0,22 (P) e 0,20 (M) e para fase de sobreano de 0,43 (C), 0,40 (P) e 0,40 (M), trabalhando com animais da raça Nelore e usando modelo de limiar. Para o escore de umbigo, BIGNARDI et al., (2010) utilizando modelo linear encontraram valores de 0,15 de h^2 para U a desmama e 0,27 na fase de sobreano em animais da raça Nelore. GORDO et al., (2012) estimaram h^2 maiores, de 0,38 também analisando dados de bovinos Nelore e com modelo linear. Também analisando dados em bovinos Nelore, BOLIGON et al., (2016), compararam as estimativas de modelos e encontraram valores de h^2 com modelo de limiar para U à desmama de 0,22 e ao sobreano de 0,42, e com modelo linear, 0,16 e 0,29, à desmama e ao sobreano, respectivamente.

Os resultados destes estudos indicam existir variabilidade genética para essas características e, apesar de serem avaliados de maneira subjetiva, quando aplicados

critérios e por avaliadores treinados, podem servir para identificar indivíduos com maior precocidade de terminação, melhor musculatura e conformação frigorífica, e sem tamanho excessivo na maturidade (FARIA et al., 2009).

Portanto, espera-se que a utilização de metodologias adequadas e o uso da genômica possa incrementar os ganhos genéticos para características de crescimento e escores visuais, muito utilizadas em programas de melhoramento genético no Brasil.

3. OBJETIVOS

3.1 Objetivos gerais

Comparar metodologias que combinem informações de genética quantitativa e molecular para estudos seleção e associação genômica ampla de características economicamente importantes em bovinos Hereford e Braford.

3.2 Objetivos específicos

- Estimação de parâmetros e valores genéticos para os escores visuais utilizando modelo linear e de limiar e validação desses modelos utilizando regressão multinomial;
- Comparar a acurácia de diferentes metodologias de estimação de efeitos de marcadores SNP nas características de crescimento e escores visuais em bovinos de corte;
- Utilizar e comparar diferentes estratégias de incorporação de dados genômicos nas avaliações genéticas, através das metodologias de multipassos (multistep) e de passo único (singlestep);
- Detectar SNP's em associação com as características de crescimento e propor um painel de baixa densidade com base nos marcadores selecionados para as raças Hereford e Braford.

4. HIPÓTESES

- Os parâmetros genéticos para os escores visuais quando estes possuem distribuição normal são semelhantes utilizando modelos lineares ou de limiar;
- A utilização da informação de marcadores moleculares aumenta a acurácia dos valores genéticos estimados para características de importância econômica;
- O uso de marcadores moleculares combinada com a avaliação genética tradicional nas metodologias de múltiplos passos e de passo único possibilita a estimação de valores genéticos mais acurados para as características estudadas em relação aos métodos tradicionais;
- Existem marcadores moleculares com maior efeito sobre as características em estudo, sendo possível a seleção de SNPs para desenvolvimento de painéis de baixa densidade.

5. CAPÍTULO I

MODELOS DE LIMIAR E LINEAR PARA AVALIAÇÃO GENÉTICA POR ESCORES
VISUAIS EM BOVINOS HEREFORD E BRAFORD

5.1 INTRODUÇÃO

A avaliação por escores visuais vem sendo amplamente utilizada em programas de melhoramento de gado de corte no Brasil, por serem características de fácil medição, baixo custo e alcançarem resposta significativa a seleção. O uso destes escores, permite identificar animais com maior desenvolvimento ponderal e, ao mesmo tempo, melhor conformação morfológica, precocidade de acabamento e musculosidade (KOURY FILHO et al., 2010).

A avaliação do escore de umbigo através do tamanho e posição, é importante para seleção de touros em sistemas de produção extensivos, com base em composição heterogênea de pastagens, pois reprodutores com umbigo longo e pendurado seriam expostos a traumas e lesões, prejudicando seu desempenho (BOLIGON et al., 2016).

Considerando que muitas características morfológicas não apresentam distribuição normal dos dados, deve-se ter cuidado em relação à metodologia aplicada na análise genética desses caracteres. Tanto em pesquisa como em programas de melhoramento, tem-se optado pelos modelos lineares em relação aos de limiar na avaliação genética de características categóricas, pela facilidade de aplicação e pelo menor tempo gasto de processamento (FARIA et al., 2008). Contudo, teoricamente, os métodos lineares são descritos como não apropriados para se quantificar a natureza discreta dos dados categóricos (MARCONDES et al., 2005), pois algumas dessas características assumem distribuição não-normal.

Alguns estudos relataram herdabilidades variando de 0,23 a 0,34 utilizando modelo linear (CARDOSO et al., 2004; KOURY FILHO et al., 2010; BIGNARDI et al., 2011; GORDO et al., 2016) e de 0,16 a 0,43 quando aplicado modelo de limiar (BOLIGON et al., 2011; BOLIGON et al. 2012; BOLIGON et al., 2013) para os escores visuais de umbigo, conformação, precocidade, musculatura e tamanho ao sobreano em diferentes raças de bovinos. Além disso, escores visuais de conformação, precocidade e musculatura, estão correlacionados positivamente com algumas características de crescimento e de carcaça (YOKOO et al. 2009; KOURY FILHO et al. 2010; GROSSO et al., 2014; GORDO et al., 2016).

O presente estudo foi desenvolvido com o objetivo de comparar as estimativas de parâmetros e valores genéticos a partir de modelos animal linear e de limiar e

validar esses modelos através da relação entre o desempenho fenotípico e o mérito genético para as características de escores visuais de umbigo, conformação, precocidade, musculatura e tamanho ao sobreano, obtidas em análises bayesianas bicaracterísticas em bovinos das raças Hereford e Braford.

5.2 MATERIAL E MÉTODOS

5.2.1 Dados e características avaliadas

Foram utilizados registros de 127.539 animais das raças Hereford e Braford nascidos entre os anos de 1991 a 2012, pertencentes ao programa de melhoramento Conexão Delta G e mantidos em 23 fazendas localizadas no estado do Rio Grande do Sul, Brasil. Foram estudadas as características de escores visuais de umbigo (U), conformação de carcaça (C), precocidade de acabamento (P), musculatura (M) e tamanho (T) ao sobreano e a característica contínua, peso ao sobreano ajustado para 550 dias de idade (PS550).

Os animais foram pesados ao nascimento, à desmama (aproximadamente aos 205 dias de idade) e ao sobreano (aproximadamente aos 550 dias de idade), além de serem avaliados visualmente ao sobreano, recebendo notas individuais de um a cinco quanto à U, C, P, M e T, em que a nota cinco refere-se à expressão máxima da característica.

A característica PS550 foi calculada como segue:

$$PS550 = \frac{PSR - PDR}{IDS - IDD} \times 345 + PD205$$

em que: PDR é o peso à desmama real do animal, em kg; PSR é o peso ao sobreano real do animal, em kg; IDD é a idade na desmama, em dias; IDS é a idade ao sobreano, em dias e PD205 é o peso a desmama ajustado para os 205 dias de idade. Foram utilizadas apenas medidas de PSR quando a IDS do animal se encontrava dentro do intervalo de 330 a 670 dias.

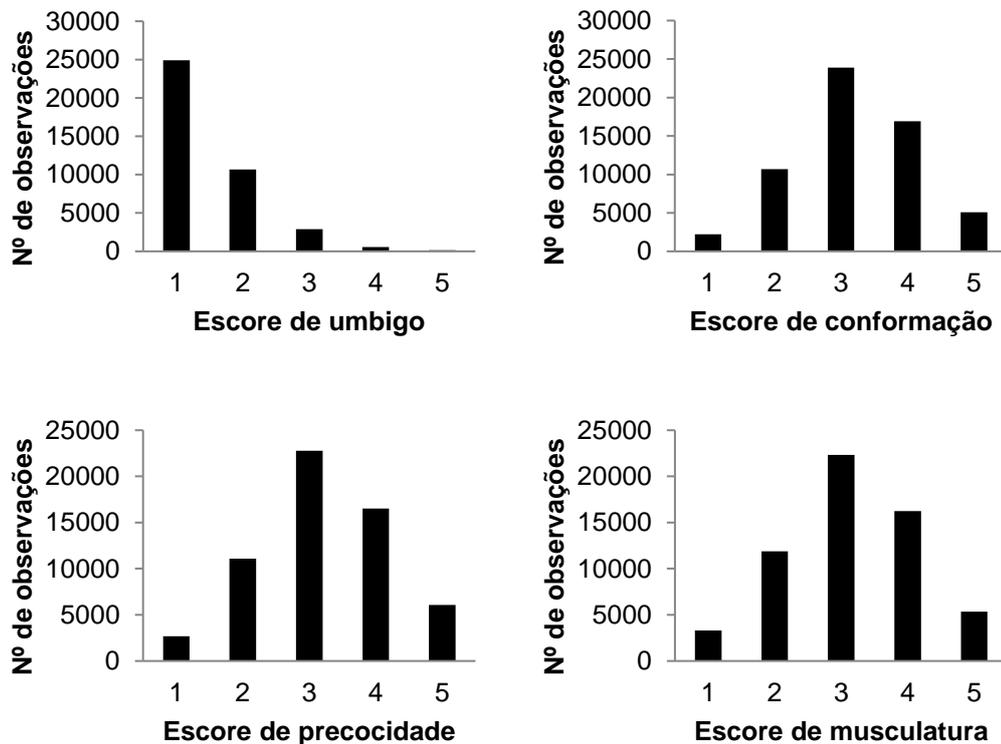
A avaliação visual do escore de umbigo foi avaliada com base na bainha e tamanho e posição do umbigo, correspondendo à dobra umbilical em fêmeas e à formação combinada da bainha e da prega umbilical em machos. As notas foram atribuídas em uma escala de 1 a 5, com escores mais altos referenciados a animais com bainhas e umbigos maiores e mais pesados e escores menores para aqueles com uma bainha e um umbigo mais retraídos.

O escore visual de conformação foi atribuído considerando o comprimento e a profundidade do corpo, associado ao desenvolvimento muscular e à harmonia geral do indivíduo, sendo um indicativo da quantidade de carne na carcaça do animal. Com a avaliação do escore visual de precocidade, procurou-se estimar a capacidade do

indivíduo em armazenar gordura, o que é indicação de rapidez para atingir o acabamento. O escore visual de musculatura é um indicativo do grau de desenvolvimento muscular apresentado pelo animal. O escore para tamanho corporal é determinado principalmente pelo comprimento corporal, profundidade de costelas e altura do animal.

Neste sistema de avaliação para os escores de C, P, M e T, recomenda-se que os avaliadores distribuam as notas o mais próximo possível de uma distribuição normal, com frequência aproximada de 10% para as notas extremas 1 e 5, 20% para as notas 2 e 4 e 40% para nota 3. Os animais devem ser avaliados em relação a média do grupo de contemporâneos para essas características. No entanto, para o escore de umbigo, os animais são avaliados em relação a um tamanho e formato de prepúcio ou umbigo ideal, independente da média e distribuição dos escores do grupo contemporâneo ao qual pertencem.

Na Figura 1 é apresentada a distribuição dos escores de umbigo, conformação, precocidade, musculatura e tamanho nos rebanhos estudados.



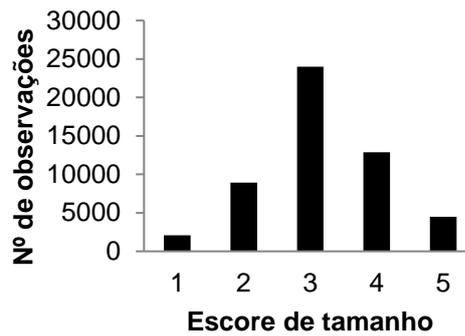


Figura 1. Distribuição das observações dos escores visuais de umbigo, conformação, precocidade, musculatura e tamanho.

5.2.2 Modelos de análise

A definição do grupo de contemporâneos (GC) incluiu as variáveis sexo, fazenda ao sobreano, ano e estação de nascimento e grupo de manejo ao sobreano. Para os escores visuais foram eliminados os GC com menos de cinco animais e também sem variabilidade. Para a característica PS550 foram excluídos GC com menos de oito animais e medidas de 3,5 desvios padrão acima ou abaixo da média do seu GC. A estrutura geral do arquivo de dados analisado após as consistências está descrita na Tabela 1.

Tabela 1. Estrutura dos dados para escores visuais e peso ajustado ao sobreano ajustado para os 550 dias de idade (PS550).

Características	Número de observações	Médias \pm DP ¹	Número de GC ²	Número de Touros	Número de Vacas
Umbigo	39.196	1,47 \pm 0,72	2.373	835	25.913
Conformação	58.825	3,20 \pm 0,96	3.518	921	32.642
Precocidade	59.117	3,21 \pm 1,01	3.552	923	32.727
Musculatura	59.077	3,14 \pm 1,02	3.547	923	32.721
Tamanho	52.395	3,17 \pm 0,94	3.115	909	30.394
PS550	59.638	315,91 \pm 75,17	3.575	923	32.593

¹DP: desvios padrão; ²GC: grupos de contemporâneos

Os componentes de (co)variâncias e os parâmetros genéticos foram estimados por meio de análises bayesianas bicaracterísticas, considerando um modelo animal linear-linear e um modelo linear-limiar, utilizando a característica PS550 como linear em todas as análises simultaneamente com cada um dos escores visuais, que foram alternativamente considerados através de modelos lineares ou de limiar. A combinação das características no modelo linear-limiar foi feita com o programa

computacional THRGIBBSF90 (MISZTAL et al., 2002), que permite o estudo de características contínuas e categóricas na mesma análise. No modelo linear-linear foi utilizado o software GIBBS1F90 (MISZTAL et al., 2002). As estimativas a posteriori foram obtidas com a utilização do aplicativo POSTGIBBSF90 (MISZTAL et al., 2002).

O modelo completo pode ser representado em notação matricial como:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{a} + \mathbf{e} \quad [1]$$

em que: \mathbf{y} é o vetor dos caracteres observados; $\boldsymbol{\beta}$ é o vetor dos efeitos fixos; \mathbf{a} é o vetor dos efeitos genéticos aditivos diretos; \mathbf{e} é o vetor dos efeitos residuais; \mathbf{X} e \mathbf{Z} são as matrizes de incidência que relacionam as observações aos efeitos fixos e aleatórios, respectivamente. Considerou-se que $E[\mathbf{y}] = \mathbf{X}\boldsymbol{\beta}$, $\text{Var}(\mathbf{a}) = \mathbf{A} \otimes \boldsymbol{\Sigma}\mathbf{a}$ e $\text{Var}(\mathbf{e}) = \mathbf{I}_N \otimes \boldsymbol{\Sigma}\mathbf{e}$, em que: $\boldsymbol{\Sigma}\mathbf{a}$ é a matriz de covariâncias genéticas aditivas; $\boldsymbol{\Sigma}\mathbf{e}$ é a matriz de covariâncias residuais; \mathbf{A} é a matriz de parentesco; \mathbf{I} é uma matriz identidade; N é o número de animais com registros e \otimes denota o produto direto entre as matrizes. Considerou-se que os vetores \mathbf{a} e \mathbf{e} são independentes.

De acordo com o enfoque bayesiano, os vetores $\boldsymbol{\beta}$ e \mathbf{a} são parâmetros de locação de uma distribuição condicional $\mathbf{y} \mid \boldsymbol{\beta}, \mathbf{a}$. A distribuição a priori de $\boldsymbol{\beta}$ foi considerada ser uniforme. Para as matrizes de (co)variâncias dos efeitos aleatórios residuais, foi considerada uma distribuição Wishart invertida. Deste modo, a distribuição de \mathbf{y} para a característica que tem distribuição contínua (scores visuais e PS550 para o modelo linear), dado os parâmetros de locação e escala, podem ser descritos de acordo com (VAN TASSEL & VAN VLECK, 1996), como segue:

$$(\mathbf{y} \mid \boldsymbol{\beta}, \mathbf{a}, \mathbf{R}) \sim M[\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{a}, \mathbf{I}_N\mathbf{R}]$$

Como efeitos fixos no modelo, foram considerados os efeitos de GC e classe de idade da vaca. Foram utilizadas classes para idade da mãe, em um total de 15 classes, em que a classe 1 corresponde a vacas com 2 anos e a classe 15 corresponde a vacas com mais de 16 anos de idade. Também foram usadas, como covariáveis, os efeitos linear e quadrático da idade do animal à mensuração e os efeitos lineares aditivos de proporção de Nelore e heterozigose individual.

No modelo de limiar, considera-se que a escala subjacente apresenta distribuição normal contínua, representada como:

$$\mathbf{U} \mid \boldsymbol{\theta} \sim N(\mathbf{W}\boldsymbol{\theta}, \mathbf{I}\sigma_e^2),$$

em que: \mathbf{U} é o vetor da escala base de ordem r ; $\boldsymbol{\theta} = (\boldsymbol{\beta}', \mathbf{a}')$ é o vetor dos parâmetros de locação de ordem s , com $\boldsymbol{\beta}$ definido sob o ponto de vista frequentista como efeitos fixos, com \mathbf{a} , como efeitos aleatórios genéticos aditivos diretos; \mathbf{W} é a matriz de

incidência conhecida, de ordem r por s ; I é a matriz de identidade de ordem r por r e σ_e^2 é a variância residual. Quando se considera que a variável na distribuição subjacente não é observável, a parametrização $\sigma_e^2 = 1$ é geralmente adotada para que se possa identificá-la na função de verossimilhança (GIANOLA & SORENSEN, 2002). Tal pressuposição é padrão em análises para dados categóricos em modelo de limiar. Entretanto, quando há pelo menos três categorias distintas, uma pressuposição alternativa consiste em fixar os valores de dois limiares, tornando possível a estimação de σ_e^2 (SORENSEN et al., 1995).

As características categóricas foram determinadas por variáveis contínuas não observáveis, em escala subjacente, em que são fixados valores iniciais de limiares, tais que: $t_1 < t_2 \dots < t_{j-1}$, com $t_0 = -\infty$ e $t_j = \infty$, em que: j é o número de categorias. Os dados observáveis são dependentes da variável subjacente, que é limitada entre dois limiares não observáveis (GIANOLA & FOULLEY, 1983). Assim, as categorias ou escores de y_i (escores visuais), para cada animal i , são definidos por U_i , na escala subjacente:

$y_i = 1$, para $t_0 < U_i \leq t_1$; 2, para $t_1 < U_i \leq t_2$; 3, se $t_2 < U_i \leq t_3$; 4, se $t_3 < U_i \leq t_4$, e 5 se $t_4 < U_i \leq t_5$, para $i = 1, \dots, n$,

em que: n é o número de observações. Após as especificações dos limiares t_0 a t_5 , é necessário que um dos limiares (de t_1 a t_4) seja ajustado a uma constante arbitrária. Neste estudo, entretanto, considerou-se $t_1 = 0$ e $t_2 = 1$, tal que o vetor de limiares estimáveis foi definido como $t = t_3$ e t_4 , tornando possível a estimativa de σ_e^2 .

Foram implementados diferentes tamanhos de cadeia devido a diferença de convergência para cada característica e modelo. Para o modelo linear, foram usados 400 mil ciclos, em que os primeiros 100 mil foram descartados. As amostras foram armazenadas a cada 10 ciclos. Para o modelo de limiar, cadeias com comprimento de 800 mil para U e C, e de 1,2 milhão de ciclos para P, M e T, em que os primeiros 200 mil foram descartados. As amostras foram armazenadas a cada 20 ciclos. O período de descarte e o intervalo de amostragem foram estabelecidos empiricamente com base em gráficos de trajetória e autocorrelações. A convergência foi verificada com a inspeção gráfica, valores amostrados x iterações, e com o critério proposto por Geweke (1992) como teste de hipótese, por meio do pacote BOA, do programa R versão 2.10.1 (SMITH, 1997). Para todas as características, as médias das variâncias *a posteriori*, estimadas em cada uma das análises bicaracterísticas, foram usadas para a estimação da herdabilidade média *a posteriori*.

As predições dos valores genéticos foram utilizadas para calcular as correlações de posto ou “Spearman” entre os modelos comparados para todos os touros e para os 1%, 10% e 50% melhores reprodutores.

5.2.3 Validação dos modelos

Nesta etapa, foi realizada validação sobre os animais jovens. Para isso, utilizou-se a técnica de validação cruzada, separando os animais jovens, candidatos a seleção, nascidos entre 2011 e 2012, em um grupo à parte. Para as análises de validação cruzada foram excluídos os dados desse grupo de animais jovens (conjunto de validação) e usados somente os dados dos animais nascidos até 2010 (conjunto de treinamento) para obter as predições genéticas de todos os animais. Com isso, a informação fenotípica usada para validar a informação genética não contribuiu para a avaliação genética, e, portanto, não houve covariância ambiental entre as duas fontes de dados.

Para comparação dos modelos, foi utilizada uma metodologia semelhante à usada por McHUGH et al., (2014), empregando uma regressão múltipla. O modelo utilizado nessa regressão foi semelhante ao empregado na equação [1], porém a única exceção foi a substituição do efeito genético aleatório do animal por uma variável independente de mérito genético para cada característica e modelo estudado pré-estimado nas análises de validação cruzada. Essas estimativas de valor genético foram incluídas nos modelos como variáveis contínuas.

O \log das probabilidades para as características de escores visuais foram estimados usando regressão logística no software R, usando a biblioteca “nnet”, assumindo uma distribuição multinomial, como descrito abaixo:

$$\log\{P(Y = j | X)\} = GC + IDV + PropNel + Heter + IDS + IDS^2 + VG + e$$

em que, $\log\{P(Y = j | X)\}$, é o log das probabilidades de J categorias de Y ($j=1, \dots, J-1$) dado as variáveis independentes, GC e IDV são os efeitos fixos de grupo contemporâneos e classes de idade da vaca, *PropNel* é a covariável de proporção de nelore, *Heter* é a covariável de heterozigose, IDS e IDS^2 são as covariáveis de idade do animal linear e quadrática e VG é a covariável do valor genético para cada característica e modelo utilizado. Foi usado para todas as características o escore 1

como nível de referência. A comparação dos modelos logísticos multinomiais foi realizada com base no “pseudo” R^2 de McFadden (McFadden, 1970).

5.3 RESULTADOS E DISCUSSÃO

Neste estudo, para o escore de umbigo, os animais são avaliados em relação a um tamanho e formato de prepúcio ou umbigo ideal. Notas maiores para prepúcio e/ou umbigo de maiores tamanhos e notas menores para prepúcio e/ou umbigo menores. Muitos animais obtiveram notas 1 e 2 e poucos receberam notas 4 e 5 (Figura 1), devido ao fato dessa característica ser avaliada independente da média e distribuição dos escores do grupo de contemporâneo ao qual pertencem. No entanto, os escores visuais de conformação, precocidade, musculatura e tamanho foram obtidos em relação à média do grupo de contemporâneos em que o animal foi avaliado. Devido a esse fato, um grande número de animais recebeu nota 3 e poucos receberam notas 1 e 5 (Figura 1).

Resultados obtidos com o programa POSTGIBBSF90 mostraram que o período de burn-in e tamanho das amostras foram suficientes para atingir a convergência dos parâmetros genéticos estudados pelo teste de Geweke (1992), tanto no modelo linear como no de limiar. Poucas diferenças entre a média, moda e mediana para as variâncias foram observadas, portanto, somente a média foi usada como medida de tendência central da distribuição *a posteriori* para o cálculo da herdabilidade e correlações genéticas. Estimativas das densidades posteriores e traços das herdabilidades para U, C, P, M e T são mostrados nas Figuras 2 e 3 para os modelos analisados. As distribuições das cadeias não mostraram grande dispersão, ou seja, as oscilações permaneceram estáveis, indicando assim que o período de burn-in considerado nas análises foi confiável e permitiu a convergência das cadeias e que as distribuições *a posteriori* dos parâmetros estimados foram relativamente simétricas (GELFAND & SMITH, 1990).

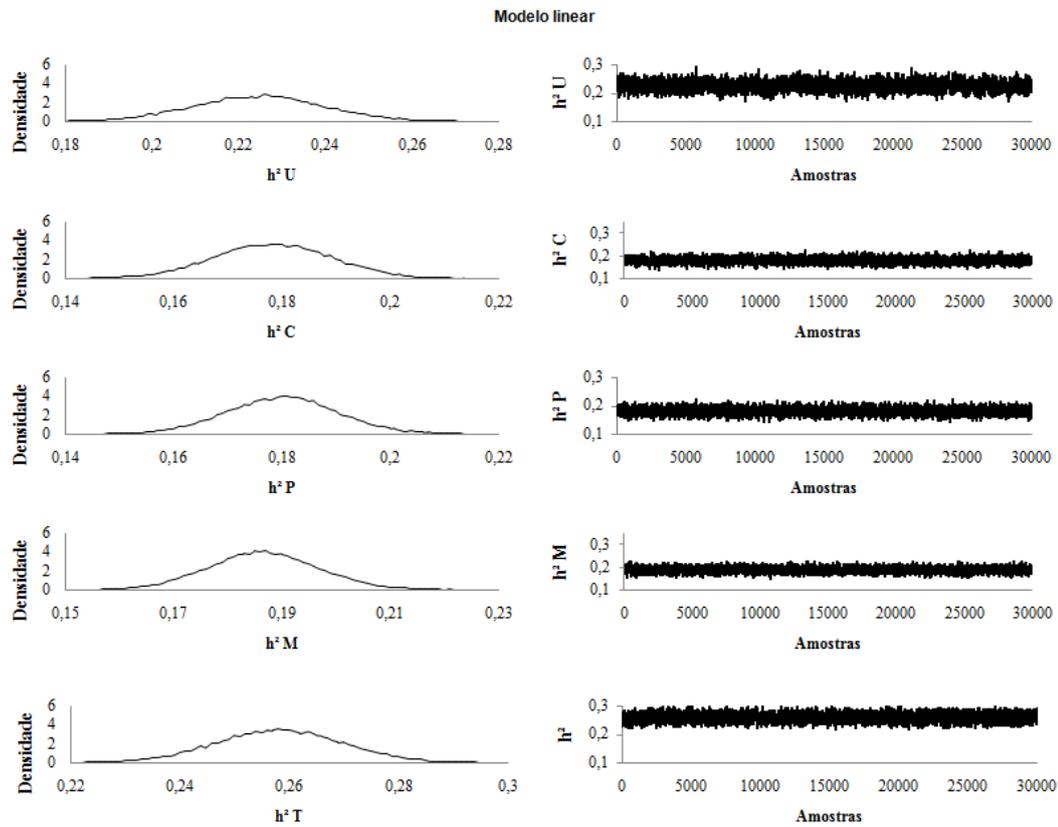


Figura 2. Estimativa da densidade a posteriori do lado esquerdo e gráfico de traço da herdabilidade estimada do lado direito para umbigo (U), conformação (C), precocidade (P), musculatura (M) e tamanho (T) para o modelo linear.

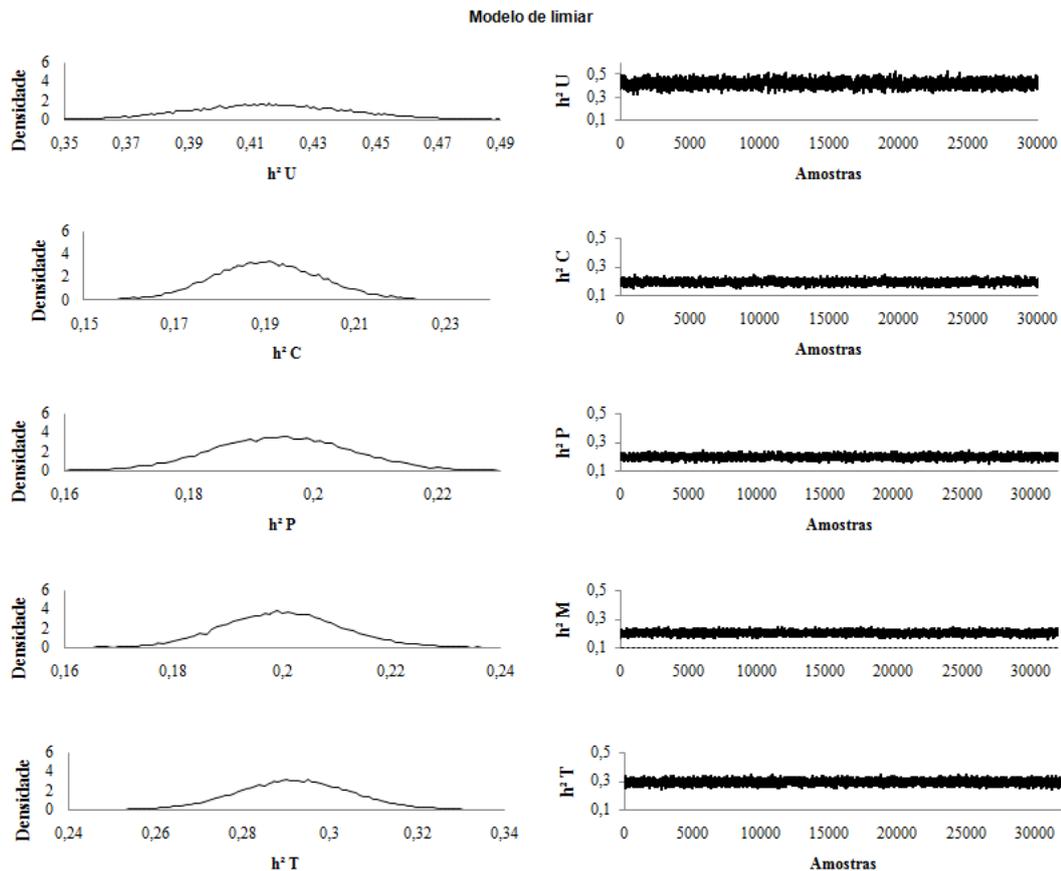


Figura 3. Estimativa da densidade a posteriori do lado esquerdo e gráfico de traço da herdabilidade estimada do lado direito para umbigo (U), conformação (C), precocidade (P), musculatura (M) e tamanho (T) para o modelo de limiar.

Para o escore de umbigo foi obtida herdabilidade média superior quando se utilizou o modelo de limiar ($0,42 \pm 0,02$) em relação ao modelo linear ($0,22 \pm 0,02$), provavelmente este ocorrido aconteceu pelo fato da distribuição para essa característica ser discrepante em relação a distribuição normal (Figura 1).

As estimativas de herdabilidade para os escores visuais de C, P, M e T ao sobreano (Tabelas 2 e 3) foram de baixa a moderada magnitude, variando entre 0,18 a 0,26 com o modelo linear e 0,17 a 0,29 com o modelo de limiar. Os valores médios de herdabilidades e os intervalos de maior densidade *a posteriori* contendo 95% das observações para C, P, M e T foram muito próximos com os modelos linear e de limiar, pois a distribuição se aproxima da normal.

Tabela 2. Estimativas a posteriori dos componentes de (co)variâncias e herdabilidades ao sobreano utilizando modelo linear-linear.

Características	Parâmetros ³	Linear-Linear			
		Médias \pm DP ¹	Moda	Mediana	HPD (95%) ²
Umbigo	σ_a^2	0,09	0,10	0,09	0,08 a 0,10
	σ_e^2	0,31	0,31	0,31	0,30 a 0,32
	h^2	$0,22 \pm 0,02$	-	-	0,21 a 0,24
Conformação	σ_a^2	0,13	0,13	0,13	0,12 a 0,15
	σ_e^2	0,62	0,62	0,62	0,60 a 0,63
	h^2	$0,18 \pm 0,01$	-	-	0,16 a 0,19
Precocidade	σ_a^2	0,16	0,16	0,16	0,14 a 0,17
	σ_e^2	0,70	0,70	0,70	0,69 a 0,72
	h^2	$0,19 \pm 0,01$	-	-	0,17 a 0,19
Musculatura	σ_a^2	0,16	0,16	0,16	0,14 a 0,18
	σ_e^2	0,70	0,70	0,70	0,68 a 0,72
	h^2	$0,19 \pm 0,01$	-	-	0,17 a 0,20
Tamanho	σ_a^2	0,18	0,18	0,18	0,17 a 0,20
	σ_e^2	0,53	0,53	0,53	0,51 a 0,54
	h^2	$0,26 \pm 0,01$	-	-	0,25 a 0,27
PS550	σ_a^2	251,41	250,50	251,40	229,23 a 273,59
	σ_e^2	754,13	754,20	754,20	733,77 a 774,49
	h^2	$0,25 \pm 0,01$	-	-	0,24 a 0,26

¹DP: desvios padrão; ²HPD: Intervalos de maior densidade *a posteriori*;

³Variâncias genéticas aditivas direta (σ_a^2); variâncias residuais (σ_e^2); herdabilidades diretas (h_a^2).

Tabela 3. Estimativas a posteriori dos componentes de (co)variâncias e herdabilidades ao sobreano utilizando modelo linear-limiar.

Características	Parâmetros ³	Linear-Limiar			
		Médias±DP ¹	Moda	Mediana	HPD (95%) ²
Umbigo	σ_a^2	0,407	0,404	0,407	0,36 a 0,46
	σ_e^2	0,572	0,574	0,572	0,52 a 0,62
	h^2	0,42±0,02	-	-	0,37 a 0,47
Conformação	σ_a^2	0,039	0,039	0,039	0,03 a 0,04
	σ_e^2	0,168	0,168	0,168	0,16 a 0,17
	h^2	0,19±0,01	-	-	0,17 a 0,21
Precocidade	σ_a^2	0,287	0,287	0,282	0,25 a 0,33
	σ_e^2	1,187	1,210	1,20	1,09 a 1,25
	h^2	0,19±0,01	-	-	0,17 a 0,22
Musculatura	σ_a^2	0,293	0,293	0,294	0,24 a 0,32
	σ_e^2	1,174	1,177	1,185	1,11 a 1,22
	h^2	0,20±0,01	-	-	0,18 a 0,22
Tamanho	σ_a^2	0,473	0,492	0,476	0,40 a 0,54
	σ_e^2	1,147	1,171	1,164	0,78 a 1,17
	h^2	0,29±0,01	-	-	0,27 a 0,32
PS550	σ_a^2	251,12	250,80	251,10	229,00 a 273,70
	σ_e^2	754,37	755,40	754,40	734,00 a 273,70
	h^2	0,25±0,01	-	-	0,23 a 0,27

¹DP: desvios padrão; ²HPD: Intervalos de maior densidade a *posteriori*.

³Variâncias genéticas aditivas direta (σ_a^2); variâncias residuais (σ_e^2); herdabilidades diretas (h_a^2).

Para o escore de U, com o modelo linear, BIGNARDI et al., (2012) encontraram valor de herdabilidade de 0,19 ao sobreano, inferior ao do presente estudo, de 0,22. Por outro lado, GORDO et al., (2012) encontraram valor superior, de 0,32 também com modelo linear. BOLIGON et al., (2016) usando modelo de limiar, estimaram valor de herdabilidade de 0,41.

As herdabilidades médias estimadas por BOLIGON et al. (2011; 2012; 2013) para os escores visuais de C, P, M e T, foram superiores em relação as obtidas no presente estudo, com médias variando de 0,24 a 0,43 para C, de 0,31 a 0,40 para P e de 0,30 a 0,40 para M, usando modelo de limiar e analisando dados de animais da raça Nelore. Por outro lado, EVERLING et al. (2012; 2014) reportaram herdabilidades médias similares as obtidas no presente trabalho, com valores de 0,17 e 0,19 para C e P, respectivamente, analisando dados de bovinos Angus, com um modelo de limiar.

Para a raça Nelore, PEDROSA et al. (2010) estimaram herdabilidades médias semelhantes ao presente estudo, utilizando um modelo animal linear para os escores de conformação, precocidade e musculatura, com valores entre 0,19 a 0,22, no entanto, as notas atribuídas para os escores variavam de 1 a 9, neste sistema de avaliação. KOURY FILHO et al. (2010) e GORDO et al. (2016) também usando um

modelo animal linear e bovinos Nelore, relataram valores levemente superiores aos estimados no presente estudo, de 0,24 a 0,34, de 0,25 a 0,32 e 0,23 a 0,27, respectivamente, para os escores de conformação, precocidade e musculatura. Para o escore de tamanho, CARDOSO et al. (2004) encontraram estimativa semelhante de herdabilidade ao do presente estudo, de 0,24, analisando dados de animais da raça Angus com um modelo animal linear.

VEZELA et al. (2011) analisando escores de carcaça de bovinos de corte na República Tcheca, pelo sistema SEUROP, encontraram valores maiores para o modelo de limiar para escores de conformação e de cobertura de gordura, em comparação ao modelo linear. PÉREZ-CABAL & CHARFEDDINE (2015) estimaram valores maiores de herdabilidade usando modelo de limiar quando comparado com o linear, analisando características de saúde do casco em gado de leite na Espanha. Entretanto, resultados semelhantes ao do presente estudo foram encontrados por FARIA et al. (2008), analisando dados de animais da raça Nelore e concluíram que não há diferença entre o modelo linear e o de limiar na obtenção de estimativas de herdabilidades e correlações genéticas de características categóricas morfológicas. PIRES et al. (2010) trabalhando com bovinos Nelore em análises bicaracterísticas observaram que ao considerar o modelo linear-linear, as estimativas de herdabilidade foram de maior magnitude que aquelas obtidas no modelo linear-limiar para escores visuais.

As diferenças nas estimativas de herdabilidades dos escores visuais em diferentes pesquisas podem ser atribuídas às variações entre observadores ou pelo sistema de avaliação visual, que é variável entre os programas de melhoramento no Brasil. A discrepância entre os resultados também pode ser devido ao tipo de modelo utilizado para estimação dos componentes de (co)variâncias e diferenças genéticas entre as populações estudadas. De modo geral, os resultados no presente trabalho para os escores visuais de U, C, P, M e T indicam que pode se obter progresso genético para essas características, pois parte da variação é atribuída aos genes com efeitos aditivos.

As estimativas *a posteriori* das correlações genéticas (Tabela 4) foram todas positivas e variaram de baixa a alta magnitudes entre a característica PS550 e os escores visuais de U, C, P, M e T. Os resultados foram praticamente os mesmos para todas as características usando tanto o modelo linear-linear como o linear-limiar. As

estimativas de correlações genéticas indicam que a seleção para PS550 irá também promover aumento nos escores visuais de C, P, M e T.

Tabela 4. Estimativas a posteriori das correlações genéticas (rg) para as características estudadas utilizando os modelos de limiar e linear.

Parâmetros	Médias \pm DP	Moda	Mediana	HPD(95%)
Modelo linear-linear				
rg PS550-U	0,17 \pm 0,04	0,15	0,17	0,10 a 0,25
rg PS550-C	0,76 \pm 0,02	0,74	0,76	0,72 a 0,80
rg PS550-P	0,33 \pm 0,03	0,31	0,33	0,27 a 0,39
rg PS550-M	0,45 \pm 0,03	0,41	0,44	0,39 a 0,50
rg PS550-T	0,85 \pm 0,01	0,83	0,85	0,82 a 0,88
Modelo linear-limiar				
rg PS550-U	0,16 \pm 0,04	0,14	0,16	0,08 a 0,24
rg PS550-C	0,76 \pm 0,02	0,75	0,75	0,72 a 0,80
rg PS550-P	0,33 \pm 0,02	0,32	0,32	0,26 a 0,39
rg PS550-M	0,44 \pm 0,02	0,41	0,43	0,38 a 0,49
rg PS550-T	0,84 \pm 0,01	0,83	0,84	0,81 a 0,87

PS550: Peso ao sobreano ajustado para os 550 dias de idade; C: Conformação; P: Precocidade; M: Musculatura; T: Tamanho; U: Umbigo;

rg: Correlação genética; DP: desvios padrão; HPD: Intervalos de maior densidade a *posteriori*.

As correlações genéticas estimadas estão abaixo dos valores estimados por KOURY FILHO et al. (2010), trabalhando com animais da raça Nelore e usando um modelo animal linear, com correlações estimadas entre peso ao sobreano e os escores de conformação, precocidade e musculatura de 0,83, 0,58 e 0,59, respectivamente. GORDO et al. (2012) também estimaram correlações genéticas superiores em relação aos resultados encontrados neste estudo usando um modelo animal linear e trabalhando com bovinos Nelore, entre peso ao sobreano e os escores de estrutura, precocidade, musculatura e umbigo de 0,91; 0,41; 0,55 e 0,38, respectivamente. EVERLING et al. (2012; 2014) analisando dados de animais da raça Angus estimaram correlação praticamente nula entre ganho de peso da desmama ao sobreano e o escore de conformação e de 0,48 entre o ganho de peso e o escore de precocidade, com a utilização de um modelo animal de limiar. BIGNARDI et al. (2011) estimaram correlação praticamente nula entre peso ao sobreano e o escore de umbigo analisando dados de bovinos Nelore.

As altas correlações genéticas obtidas entre os escores de C e T e PS550 indicam que a maioria dos genes que influenciam uma dessas características são os

mesmos que estão atuando nas outras. Esses resultados eram esperados, pois a definição da característica de conformação sugere que os escores devem ser atribuídos de acordo com a estimativa visual da quantidade de carne na carcaça do animal, e o escore de tamanho deve ser atribuído em relação ao comprimento corporal e altura do animal. No entanto, a baixa correlação genética obtida entre o escore U e o PS550 indica que a maior parte dos genes que estão atuando em uma característica não influenciam o desempenho da outra.

De modo geral, os resultados dos parâmetros genéticos foram semelhantes nos dois modelos para os escores visuais para C, P e M, exceto para a herdabilidade do escore de T, que foi levemente superior com o modelo de limiar. Para o escore U, maior valor de herdabilidade foi observado com o modelo de limiar. De acordo com SILVA et al., (2005), seria mais apropriado para estimação dos parâmetros genéticos de dados categóricos a utilização de um modelo de limiar, por considerar o caráter discreto dos dados ao relacionar a resposta observada na escala categórica a uma escala subjacente normal contínua. Além disso, utilizando a metodologia linear em dados categóricos que não são normalmente distribuídos, alguns pressupostos são violados (GIANOLA & FOULLEY, 1983). Um dos principais, de acordo com GIANOLA (1982) é que a variância de uma característica categórica é uma função de sua esperança, e resulta em heterogeneidade da variância. GIANOLA (1982) ainda destaca que as soluções do BLUP não consideram a restrição de que a soma das probabilidades de resposta tenha que ser igual a um entre todas as categorias de resposta e que a variância genética aditiva na escala observada depende da incidência da característica na população dada.

Outro fato importante em relação ao modelo de limiar, é que na prática os valores genéticos devem ser transformados para a escala de probabilidade para obtenção das DEPs (Diferença Esperada na Progênie). Desta forma, a unidade da DEP para as características de escores visuais é dada em porcentagem, com o objetivo de facilitar a interpretação desses valores. Para as características com cinco escores, a DEP predita irá expressar a probabilidade de um animal deixar filhos que atinjam o escore ideal para a característica (VARONA, 2006).

Em relação as correlações de posto ou “Spearman” (Tabela 5) dos valores genéticos dos touros, entre os modelos de linear-linear com os de linear-limiar, para o escore de U, variaram de 60% a 96%, indicando que ocorreram na classificação dos touros utilizando um modelo ou outro, principalmente quando a intensidade de seleção

for mais alta (menor proporção de selecionados), pois a correlação observada foi de média magnitude.

Quanto aos escores de C, P, M e T, os valores das correlações variaram entre 96% e 99% quando se utilizou todos os reprodutores. Quando foram selecionados os melhores 50%, 10%, 5% e 1% dos melhores touros, as correlações foram de 95% a 99%, para C, P e M. Para o escore de tamanho, ao selecionar os animais top 1%, a correlação de classificação baixou para 81%. De modo geral, para C, P, M e T não são esperadas grandes mudanças na classificação dos animais em ambos os modelos, pois as correlações foram de alta magnitude. Resultados semelhantes foram encontrados por MARCONDES et al. (2005) e FARIA et al. (2008) que não detectaram diferenças na classificação dos animais por meio de modelos lineares ou de limiar.

Tabela 5. Correlações de Spearman entre os valores genéticos obtidos pelos modelos de limiar e linear avaliando diferentes níveis de proporção de touros selecionados, considerando todos os touros (100%) e os melhores 50%, 10%, 5% e 1%.

Características	Proporção dos touros selecionados				
	100%	50%	10%	5%	1%
Umbigo	0,96	0,91	0,83	0,71	0,60
Conformação	0,99	0,98	0,96	0,96	0,95
Precocidade	0,99	0,99	0,97	0,97	0,96
Musculatura	0,99	0,99	0,97	0,96	0,96
Tamanho	0,99	0,99	0,98	0,97	0,81

É sabido, que a variação genética contribui significativamente para diferenças fenotípicas entre os indivíduos através das características de importância econômica (McHUGH et al., 2014). A contribuição relativa da variação genética aditiva às diferenças no desempenho fenotípico, associada ao conhecimento dos efeitos ambientais sistemáticos e do parentesco entre os animais, são utilizadas para estimar os valores genéticos dos indivíduos. As medidas de mérito genético utilizadas neste foram equivalentes aos valores genéticos (VG).

Para todas as características, exceto para U com o modelo linear, a direção das associações entre o fenótipo e o VG estão de acordo com o esperado (Tabela 6). Houve incremento no fenótipo por unidade de mudança no VG para, C, P, M e T.

Para o escore de U, utilizando o VG estimado pelo modelo linear, houve inconsistência, pois, os coeficientes de regressão relacionando os valores genéticos como os escores fenotípicos dos indivíduos diminuí para o escore 5 em relação aos escores 3 e 4. Esta associação não está correta, pois espera-se que animais com um escore maior para umbigo também tenham valores genéticos com valores mais altos.

No entanto, utilizando os VG estimados pelo modelo de limiar para a mesma característica, estão de acordo com o esperado, pois o aumento de 1 unidade no VG irá resultar na chance de estar em uma classe maior de escore.

Tabela 6. Coeficientes de regressão (β^1) do desempenho fenotípico (escores) sobre o valor genético da característica, R^2 (pseudo coeficiente de determinação de McFadden) e critério de informação Aikaike (AIC) para cada modelo e característica.

Características	Escores	Modelo linear		Modelo de limiar	
		β	R^2	β	R^2
Umbigo	2	3,23	0,33	1,52	0,37
	3	5,34		2,64	
	4	6,74		2,83	
	5	4,12		3,69	
Conformação	2	0,90	0,45	1,49	0,45
	3	2,16		3,87	
	4	3,06		4,92	
	5	4,51		5,96	
Precocidade	2	1,74	0,37	1,56	0,38
	3	2,81		2,49	
	4	4,02		3,53	
	5	4,88		4,27	
Musculatura	2	1,33	0,41	0,74	0,41
	3	2,00		1,15	
	4	2,91		1,70	
	5	4,34		2,57	
Tamanho	2	0,90	0,45	0,62	0,45
	3	2,02		1,51	
	4	3,11		2,37	
	5	4,15		3,04	

¹ Coeficientes de regressão apresentados em escala logarítmica.

Além disso, os valores de pseudo R^2 estimado na regressão multinomial foram diferentes utilizando o valor genético dos modelos linear (0,33) e de limiar (0,37). O valor mais alto foi quando se usou o VG estimado com o modelo limiar, indicando um melhor ajuste do modelo. Segundo LOUVIERE et al., (2000), pseudo R^2 de McFadden é usado para indicar o ajuste do modelo geral, e sugeriram que valores de R^2 entre 0,2 e 0,4 deveriam ser considerados como representando um bom ajuste.

Para as características de C, P, M e T o log das probabilidades (β) aumentou linearmente por unidade de aumento no VG (Tabela 6). O maior aumento foi para a característica de C com o VG estimado pelo modelo de limiar, aumentando linearmente de 1,49 para o escore 2 vs escore 1, até 5,96 para o escore 5 vs 1, por unidade de aumento no VG. Os valores de pseudo R^2 para essas características foram muito semelhantes e não houve diferença em relação ao ajuste do modelo usando o VG do modelo linear ou de limiar.

Os resultados do presente estudo, para a característica de U, indicam que o melhor modelo a ser utilizado é o de limiar, mesmo com a maior dificuldade de interpretação das DEP's com esse modelo. Por outro lado, para os escores de C, P e M, não há diferença na utilização de um modelo ou outro sendo o mais indicado o uso do modelo linear. Para o escore de T, apesar da herdabilidade ter sido levemente superior com o modelo de limiar, seria melhor o uso do modelo linear por causa da facilidade de interpretação dos valores genéticos e aplicação.

5.4 CONCLUSÕES

Para o escore de umbigo em que a distribuição não se assemelha a normal, ocorreu mudança na estimativa da herdabilidade, com vantagem para o modelo de limiar em relação ao modelo linear. A classificação dos animais também foi influenciada para essa característica ao comparar os modelos analisados.

Não há diferença entre o modelo linear e o de limiar na obtenção de estimativas de parâmetros genéticos, quando os escores visuais possuem uma distribuição semelhante à normal, como para as características de conformação, precocidade, musculatura e tamanho, sendo o mais indicado o modelo linear. O uso de qualquer um dos modelos não influenciou a classificação dos animais quanto aos valores genéticos preditos para essas características.

Os resultados da validação cruzada por regressão demonstraram que a seleção genética para os escores visuais também irá promover melhorias no desempenho fenotípico, para conformação, precocidade, musculatura e tamanho. Para o escore de umbigo, o modelo de limiar apresentou melhor associação entre o mérito genético e o desempenho fenotípico.

6. CAPÍTULO II

ACURÁCIA DE PREDIÇÃO GENÔMICA PARA CARACTERÍSTICAS DE
CRESCIMENTO E ESCORES VISUAIS EM BOVINOS HEREFORD E BRAFORD

6.1 INTRODUÇÃO

O melhoramento genético de características de importância econômica tradicionalmente tem sido realizado com base na metodologia dos modelos mistos, incorporando dados de genealogia e fenótipo do animal para a obtenção dos valores genéticos dos indivíduos através do BLUP (Best Linear Unbiased Prediction; HENDERSON, 1975). Na prática, utilizando a metodologia tradicional para uma dada intensidade de seleção, pode-se aumentar a acurácia dos valores genéticos preditos, mas o intervalo de gerações é estendido até que os registros fenotípicos dos indivíduos ou de seus descendentes estejam disponíveis para a predição dos valores genéticos (SAATCHI et al., 2011), desse modo reduzindo os ganhos genéticos possíveis.

Com o avanço nas tecnologias de sequenciamento de DNA e genotipagem, e a consequente detecção de milhares de polimorfismo de um único nucleotídeo (SNPs), tornou-se possível o desenvolvimento de novas metodologias para estimar os efeitos destes polimorfismos e também combinar as informações genômicas com as tradicionalmente utilizadas. A inclusão de dados fornecidos pelos marcadores moleculares cobrindo todo o genoma no processo de predição do mérito genético além dos dados de desempenho e de pedigree, é denominado seleção genômica (MEUWISSEN et al., 2001). Com a abordagem da seleção genômica, os ganhos genéticos são proporcionados principalmente pela maior acurácia de seleção e pela redução do intervalo de gerações (GODDARD & HAYES, 2007).

Existem vários métodos utilizados em seleção genômica, destacando-se: BLUP (MEUWISSEN et al., 2001), regressão de cumeieira (WHITTAKER et al., 2000; GIANOLA et al., 2003), os métodos bayesianos (MEUWISSEN et al., 2001; HABIER et al., 2011) e o LASSO bayesiano (PARK & CASELLA, 2008; DE LOS CAMPOS et al., 2009). As diferenças entre esses métodos decorrem de diferentes pressuposições adotadas a respeito das distribuições dos efeitos dos marcadores para viabilizar a estimativa dos seus efeitos. De forma geral, esses métodos viabilizam a manutenção de todos os efeitos no modelo empregando penalizações via encolhimento das estimativas dos efeitos para zero (XU, 2007).

Os efeitos estimados dos marcadores (SNPs) podem ser utilizados para prever o valor genético de qualquer animal genotipado sendo chamado de valor

genômico direto (DGV – Direct Genomic Value). A magnitude da acurácia do DGV irá depender de vários fatores, tais como o tamanho da população de referência (DAETWYLER et al., 2010), a herdabilidade da característica (GODDARD & HAYES, 2009), a extensão do desequilíbrio de ligação entre os marcadores e os QTL's (quantitative trait loci; WOLC et al., 2011), além dos métodos estatísticos empregados para a estimação dos efeitos dos marcadores (SU et al., 2010).

O DGV pode ser ainda combinado com o valor genético tradicional para o cálculo do valor genético aprimorado pela genômica (GEBV – Genomic Breeding Value; VAN RADEN et al., 2009). Procedimentos multipassos (multistep; VAN RADEN et al., 2009; McNEIL et al., 2010) e passo único (singlestep; MISZTAL et al., 2009; AGUILAR et al., 2010) foram propostos e estão sendo testados para identificar a estratégia mais apropriada para implementação da seleção genômica em programas de avaliação e melhoramento genético.

As características de crescimento e escores visuais são amplamente utilizadas como critérios de seleção em bovinos de corte no Brasil e, portanto, espera-se que a utilização de predições genômicas para essas características possam melhorar a eficiência dos programas de melhoramento e, conseqüentemente, aumentar os ganhos genéticos. Entretanto, para que esses resultados desejáveis sejam alcançados torna-se necessário determinar, entre os métodos disponíveis atualmente, o mais indicado a ser utilizado em avaliações genômicas. O presente estudo foi desenvolvido com o objetivo de se avaliar as acurácias e o viés das predições genômicas, para características de crescimento (pesos e ganhos em peso) e escores visuais (conformação, precocidade, musculatura e tamanho) obtidas na desmama e ao sobreano em bovinos das raças Hereford e Braford, utilizando diferentes métodos.

6.2 MATERIAL E MÉTODOS

6.2.1 Arquivo de dados fenotípicos, genotípicos e pedigree

Foram utilizados registros fenotípicos de 126.290 animais das raças Hereford e Braford nascidos entre os anos de 1991 a 2012, pertencentes ao programa de melhoramento Conexão Delta G, localizados no estado do Rio Grande do Sul, Brasil. As características estudadas foram: peso ao nascer (PN), peso a desmama ajustado aos 205 dias de idade (PD205), escores visuais atribuídos na desmama de conformação (CD), precocidade (PD), musculatura (MD), tamanho (TD), peso ao sobreano ajustado aos 550 dias de idade (PS550), ganho em peso pós-desmama ajustado para aos 345 dias de idade (GPD345) e escores visuais atribuídos ao sobreano de conformação (CS), precocidade (PS), musculatura (MS), tamanho (TS).

Os animais foram pesados ao nascimento, à desmama (aproximadamente aos 205 dias de idade) e ao sobreano (aproximadamente aos 550 dias de idade). Além de pesados, foram também avaliados visualmente na desmama e ao sobreano sendo atribuídas notas individuais de um a cinco para à conformação, precocidade, musculatura e tamanho, em que a nota 5 refere-se à expressão máxima de cada característica.

A característica PD205 foi calculada da seguinte forma:

$$PD205 = \frac{PDR - PN}{IDD} \times 205 + PN$$

em que: PDR é o peso real à desmama do animal, em kg, PN é o peso ao nascer, em kg, e IDD é a idade na desmama, em dias. Foram consideradas as medidas de PDR apenas quando a IDD do animal se encontrava no intervalo de 100 a 300 dias.

A característica GPD345 foi calculada como segue:

$$GPD345 = \frac{PS550 - PD205}{IDS - IDD} \times 345$$

em que: IDS é a idade ao sobreano, em dias.

A característica PS550 foi calculada da seguinte forma:

$$PS550 = \frac{PSR - PDR}{IDS - IDD} \times 345 + PD205$$

em que: PSR é o peso real ao sobreano do animal, em kg. Foram utilizadas apenas medidas de PSR quando a IDS do animal se encontrava no intervalo de 330 a 670 dias.

Foi utilizado um conjunto de 3.545 animais genotipados com o Illumina BovineSNP50 Bead Chip (50K; ILLUMINA, San Diego, CA) e 131 touros pais com o Illumina High-Density Bovine Bead Chip Array (HD; ILLUMINA). O controle de qualidade foi implementado usando R/SNPStats package (CLAYTON, 2014). Foram removidas amostras com taxa de genotipagem (Call Rate – CR) < 0,90, heterozigocidade com 3 desvios acima ou abaixo da média observada desse critério na população genotipada, e com erros de identificação de amostras. Esses erros foram constatados quando o sexo determinado através dos genótipos não correspondia ao sexo do indivíduo associado com tal amostra e quando duas amostras estavam associadas a indivíduos diferentes, mas tinham genótipos idênticos. Apenas SNPs mapeados para autossomas, com CR > 0,98, frequências do alelo menor (MAF) > 0,03, e desvio de Equilíbrio de Hardy-Weinberg não altamente significativo ($P > 10^{-7}$), foram consideradas para análises subsequentes. Além disso, apenas o SNP com maior MAF foi mantido quando havia SNPs na mesma posição ou os genótipos estavam altamente correlacionados ($r > 0,98$). Os dados do chip de alta densidade (HD) foram filtrados para selecionar somente SNPs que também estavam presentes no painel de 50K. Após a edição, um total de 41.045 marcadores SNPs e 3.592 amostras foram utilizadas nas análises, incluindo, 2934 animais da raça Braford e 658 da raça Hereford. Dados perdidos em genótipos (0,86% de todos os genótipos) foram imputados usando o software FImpute (SARGOLZAEI et al., 2011).

6.2.2 Análises e modelo estatístico

Primeiramente, foram estimados os componentes de (co)variâncias e parâmetros genéticos usando inferência bayesiana por amostragem de Gibbs, com o programa Gibbs1f90 (MISZTAL et al., 2002) em análises uni-características. Foram considerados 800.000 ciclos, com descarte inicial de 200.000 ciclos e intervalo de amostragem de 20. Posteriormente, os componentes de (co)variâncias estimados foram utilizados na predição dos valores genéticos, usando o software BLUPF90 (MISZTAL et al., 2002). O seguinte modelo estatístico, em notação matricial, foi utilizado nas análises:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{Wm} + \mathbf{Spe} + \mathbf{e} \quad [1]$$

em que: \mathbf{y} é o vetor de observações, \mathbf{X} , \mathbf{Z} , \mathbf{W} e \mathbf{S} são as matrizes de incidência que relacionam as observações aos vetores \mathbf{b} de efeitos fixos, \mathbf{u} de efeito aleatório genético aditivo direto, \mathbf{m} de efeito aleatório genético materno e \mathbf{pe} de efeito de ambiente permanente materno, e \mathbf{e} é o vetor dos efeitos aleatórios residuais.

Como efeitos fixos, foram considerados os efeitos de grupos contemporâneos (GC – mesma fazenda, sexo, ano e estação de nascimento, grupo de manejo e data da desmama ou sobreano) e classes de idade da vaca (total de 15 classes, em que a classe 1 correspondia a vacas com 2 anos de idade e a 15 a vacas com mais de 16 anos de idade). Além disso, foram incluídas, como covariáveis, os efeitos linear e quadrático da idade do animal na mensuração e os efeitos lineares aditivos e maternos de proporção de Nelore na composição racial e heterozigose individual e materna. Para as características medidas ao sobreano (PS550, GPD345, CS, PS, MS e TS) não foram considerados os efeitos genéticos maternos e de ambiente permanente materno, efeitos maternos de proporção de Nelore e heterozigose materna.

Para as características contínuas (PN, PD205, GPD345 e PS550), foram excluídos GC com menos de cinco animais e medidas com 3,5 desvios padrão acima ou abaixo da média do GC ao qual pertenciam. Para os escores visuais, foram eliminados os GC com menos de cinco animais e sem variabilidade. A estrutura geral do arquivo de dados analisado está descrita na Tabela 7.

Tabela 7. Estrutura do arquivo de dados para as características avaliadas nas raças Hereford e Braford.

Características	Nº de observações	Média±DP ¹	Nº de GC ¹	Nº de Touros	Nº de vacas
PN (kg)	105.596	33,00±5,32	1.045	2.162	41.279
PD205 (kg)	113.022	179,53±32,87	2.884	2.226	43.886
CD (1 a 5)	112.696	3,11±1,04	2.864	2.219	43.767
PD (1 a 5)	112.723	3,18±1,07	2.868	2.219	43.783
MD (1 a 5)	112.735	3,08±1,08	2.870	2.219	43.798
TD (1 a 5)	100.975	3,04±0,98	2.535	2.006	39.956
PS550 (kg)	60.422	315,87±77,37	3.285	1.971	30.546
GPD345 (kg)	59.617	136,34±106,45	3.261	1.965	30.295
CS (1 a 5)	60.641	3,20±0,96	3.278	1.964	30.701
PS (1 a 5)	60.879	3,20±1,01	3.304	1.964	30.784
MS (1 a 5)	60.848	3,14±1,02	3.300	1.964	30.773
TS (1 a 5)	53.849	3,17±0,95	2.902	1.789	27.559

¹DP: desvios-padrão; GC: grupos de contemporâneos.

6.2.3 Predição genômica

Foi utilizada a abordagem proposta por GARRICK et al. (2009) para calcular os valores genéticos desregredidos (DEBV – deregressed breeding values) e correspondentes ponderadores (\mathbf{w}) = $\{w_j\}$ para todos os animais, livre do efeito médio dos pais a partir dos valores genéticos e as respectivas confiabilidades. Os DEBVs ponderados foram usados como variáveis respostas para estimar os efeitos dos marcadores através de diferentes métodos. Em todos os casos, o seguinte modelo linear foi ajustado:

$$DEBV_j = \mu + \sum_{K=1}^K M_{jk} g_k + \varepsilon_j \quad , \quad [2]$$

em que: $DEBV_j$ é o DEBV do animal j , μ é a média geral, $K=41.045$ é o total de marcadores SNPs no painel, M_{jk} é a covariável que representa o genótipo do animal j no marcador K pelo número de alelos B na codificação Illumina A / B (ILLUMINA, 2006), g_k é o efeito aleatório do marcador K e ε_j é o componente residual com distribuição normal e variância heterogênea, $\varepsilon_j \sim N(0, w_j^{-1} \sigma_\varepsilon^2)$, em que σ_ε^2 é a variância residual e w_j é o ponderador como uma função da confiabilidade do $DEBV_j$ (GARRICK et al., 2009).

Devido ao fato de o número de marcadores com efeito sobre as características exceder o número de observações fenotípicas disponíveis, restrições devem ser aplicadas no modelo para tornar os efeitos dos marcadores estimáveis (GIANOLA et al., 2009). As diferenças nos métodos empregados na seleção genômica decorrem de diferentes pressuposições adotadas a respeito das distribuições dos efeitos dos marcadores: 1) SNP_BLUP (MEUWISSEN et al., 2001): os efeitos dos marcadores são considerados como pertencendo a uma mesma distribuição normal e assumindo-se que todos os marcadores possuem a mesma variância σ_g^2 , isto é, $g_k \sim N(0, \sigma_g^2)$ para todo k ; 2) BayesB (MEUWISSEN et al., 2001): assume um modelo de mistura, em que uma fração dos marcadores tem efeito nulo ($\pi=0.95$), enquanto que a outra fração de marcadores restante ($1 - \pi$), tem efeitos diferentes de zero, os quais são gerados a partir de uma distribuição normal, $g_k \sim N(0, \sigma_{g^k}^2)$ com variância específica $\sigma_{g^k}^2$,

derivada de uma distribuição qui-quadrado invertida escalonada, isto é, $\sigma_{g^k}^2 \sim \chi_{(v, S^2)}^{-2}$; 3) BayesC (HABIER et al., 2011): também é um modelo de mistura similar ao BayesB, porém assume variância comum homogênea para distribuição dos efeitos dos marcadores não nulos ($1 - \pi$), isto é, $g_k \sim N(0, \sigma_g^2)$.

Os valores genômicos diretos (DGV) foram obtidos como $DGV_j = \sum_{k=1}^k M_{jk} \hat{g}_k$, em

que as estimações dos efeitos dos SNPs foram representadas pela sua média a posteriori nos métodos bayesianos.

O método SNP_BLUP foi implementado empregando seu equivalente BLUP genômico (GBLUP; STRANDEN & GARRICK, 2009) usando os softwares da família BLUPf90 (MISZTAL et al., 2002). Os métodos bayesianos BayesB e BayesC foram implementados usando o software Gensel (FERNANDO & GARRICK, 2009). Para os métodos bayesianos, foi utilizado um tamanho de cadeia de 42.000 amostras depois de 2.000 ciclos de burn-in e hiperparâmetros para distribuição dos parâmetros foram especificados com base nos componentes de variância estimados no modelo [1]. A convergência foi verificada a partir do critério proposto por GEWEKE (1992), por meio do pacote “BOA”, do programa R versão 2.10.1 (SMITH, 1997).

6.2.4 Incorporação da informação tradicional com a genômica

Foram usadas três abordagens alternativas para combinar o DGV com os dados de pedigree e desempenho para obtenção dos valores genéticos aprimorados pela genômica (GEBV).

A primeira abordagem foi realizada com base no índice de seleção de VANRADEN et al. (2009). O primeiro conjunto do índice são os DGVs (DGV), o segundo conjunto foi calculado para um subconjunto dos animais genotipados usando o parentesco tradicional e seu DEBV, excluindo os dados dos animais não genotipados (SPA) e o terceiro conjunto, é o EBV tradicional obtido a partir do modelo (PA) [1]. Esta abordagem permite calcular o GEBV de cada animal como a diferença entre as predições genômicas e as tradicionais adicionados à média dos pais com o seguinte índice proposto por VANRADEN et al., (2009):

$$GEBV = b_1 DGV + b_2 SPA + b_3 PA \quad [3]$$

em que, $\mathbf{b} = [b_1 \ b_2 \ b_3] = \mathbf{c}'\mathbf{V}^{-1}$, e \mathbf{c}' é um vetor com as confiabilidades correspondentes aos 3 termos do índice e \mathbf{V} é uma matriz simétrica 3x3 com os elementos de \mathbf{c} na diagonal e a seguinte função das 3 confiabilidades fora da diagonal:

$$V_{12} = V_{23} = r_{SPA}^2, \text{ e}$$

$$V_{13} = r_{SPA}^2 + (r_{DGV}^2 - r_{SPA}^2)(r_{PA}^2 - r_{SPA}^2)/(1 - r_{SPA}^2).$$

Para todos os animais, V_{11} (r_{DGV}^2) e V_{33} (r_{PA}^2) foram construídas para serem maiores do que V_{22} (r_{SPA}^2) para garantir matrizes positivas definidas (VANRADEN et al., 2009).

Em uma segunda abordagem, o GEBV foi realizado com base em uma análise bi-característica entre o fenótipo e o DGV usando o pedigree tradicional (MACNEIL et al., 2010). Diferentes modelos lineares foram assumidos para cada característica. O mesmo modelo [1] foi usado para os dados fenotípicos, enquanto que para o DGV, o seguinte modelo foi assumido $DGV = \mu + \alpha_j + \varepsilon_j$, em que, μ é a média geral, α_j é o efeito genético aditivo do j-ésimo animal e ε_j é o efeito residual. Nesse modelo bivariado assumiu-se para os efeitos genéticos aditivos dos fenótipos (\mathbf{a}) e para $\alpha = \{\alpha_j\}$ a seguinte distribuição:

$$\begin{bmatrix} \mathbf{a} \\ \alpha \end{bmatrix} \sim N \left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{A}\sigma_a^2 & \mathbf{A}\sigma_{a\alpha} \\ \mathbf{A}\sigma_{a\alpha} & \mathbf{A}\sigma_\alpha^2 \end{bmatrix} \right),$$

em que: σ_a^2 é a variância genética aditiva para o DGV e $\sigma_{a\alpha}$ é a covariância genética entre o DGV e o registro fenotípico. No entanto, assumiu-se para os efeitos residuais aleatórios para o fenótipo (\mathbf{e}) e para o DGV $\varepsilon = \{\varepsilon_j\}$ a seguinte distribuição:

$$\begin{bmatrix} \mathbf{e} \\ \varepsilon \end{bmatrix} \sim N \left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{I}\sigma_e^2 & 0 \\ 0 & \mathbf{I}\sigma_\varepsilon^2 \end{bmatrix} \right),$$

em que: neste modelo bi-característica, há um empréstimo de informações entre o valor genético tradicional e o DGV através da covariância ($\sigma_{a\alpha}$) e ambas as estimativas de a_j e α_j são valores candidatos para representar o GEBV do animal j .

A terceira e última metodologia utilizada foi combinando diretamente as informações fenotípicas, de pedigree e genômicas usando o single step GBLUP (MISZTAL et al., 2009; AGUILAR et al., 2010; CHRISTENSEN & LUND, 2010). Em

ssGBLUP, foi utilizado o mesmo modelo linear [1], porém a matriz de parentesco tradicional (**A**) foi substituída pela matriz **H**, que inclui a informação genômica. Embora **H** seja complexa (LEGARRA et al., 2009), sua inversa tem forma simples (AGUILAR et al., 2010):

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

em que: **G** é a matriz de parentesco genômica construída como em VANRADEN (2008) usando frequências alélicas e **A₂₂** é a matriz do numerador do parentesco somente para os animais genotipados. As análises foram realizadas usando o software da família BLUPf90 e o ssGBLUP foi implementado usando o programa preGSf90, com as opções padrão para ajuste das matrizes **G** e **A₂₂** e tornando **G** positiva definida (AGUILAR & MIZSTAL, 2014).

6.2.5 Validação cruzada e acurácia de predição

Os animais genotipados para as características estudadas foram divididos em quatro ou cinco grupos, por duas estratégias de validação cruzada, usando o programa R (R CORE TEAM, 2013): agrupamento em função do parentesco genômico, pelo método K-médias, e de forma aleatória, de acordo com SAATCHI et al. (2011) com a finalidade de formar grupos de treinamento e validação independentes. Essas estratégias foram utilizadas para representar dois cenários de seleção genômica para que os animais tenham parentesco mais distante ou mais próximo ao conjunto de treinamento, respectivamente pelos métodos K-médias e aleatório.

O método K-médias agrupou os indivíduos em função do parentesco genômico, de forma que o relacionamento genético dentro de cada grupo fosse maior que entre grupos. A outra estratégia particionou os animais selecionando-se os indivíduos ao acaso (SAATCHI et al., 2011).

A validação cruzada foi realizada no grupo que não foi usado no treinamento, ou seja, em cada análise foram excluídos os dados de um grupo (conjunto de validação) e os dados dos grupos restantes (conjunto de treinamento) foram usados para estimar os efeitos dos marcadores e obter as predições genômicas para os animais do conjunto de validação. Esse procedimento foi repetido para todos os

grupos, de forma que foram geradas predições por validação cruzada para todos os animais com genótipo e fenótipo em cada característica. A mesma estratégia de validação cruzada foi aplicada usando o pedigree tradicional (BLUP) como forma de comparação com as predições genômicas e também para gerar os GEBVs nos métodos de aprimoramento genômico.

As acurácias de predição dos animais genotipados combinada para todos conjuntos de validação foram derivadas como a correlação genética entre o fenótipo observado (y) e o valor genômico direto (DGV) ou o valor genético genômico (GEBV), estimadas em um modelo animal bicaracterística, como demonstrado abaixo:

$$\begin{bmatrix} y \\ \text{DGV/GEBV} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{Z}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \end{bmatrix}$$

em que: $\boldsymbol{\beta}_1$ e $\boldsymbol{\beta}_2$ são os vetores de efeitos fixos (mesmos efeitos usados em [1] para $\boldsymbol{\beta}_1$ e efeito de classe dos grupos usados na validação para $\boldsymbol{\beta}_2$); $\boldsymbol{\alpha}_1$ e $\boldsymbol{\alpha}_2$ são os vetores de efeitos genéticos aditivos para as duas características, $Var(\alpha_1) = A^* \sigma_{\alpha_1}^2$, $Var(\alpha_2) = A^* \sigma_{\alpha_2}^2$ e $Cov(\sigma_1, \sigma_2) = A^* \sigma_{\sigma_1, \sigma_2}$, onde A^* é uma matriz do numerador de parentesco, em que as covariâncias entre indivíduos de diferentes grupos foram zeradas, conforme SAATCHI et al. (2013):

$$A^* = \begin{bmatrix} A_{11} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & A_{22} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & A_{33} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & A_{44} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & A_{55} \end{bmatrix}$$

Os componentes de (co)variâncias e correlações genéticas foram estimados por meio de inferência bayesiana com base no método de Monte Carlo via Cadeias de Markov, utilizando 100.000 ciclos e período de descarte inicial de 50.000 amostras usando o software Gibbs2f90 (MISZTAL et al., 2002).

Dentro de um dado grupo (g), as acurácias preditas foram alternativamente estimadas como a correlação ($\hat{r}_{a\hat{a}g}$) entre o predito ($\hat{\mathbf{a}}$) e o verdadeiro (\mathbf{a}) valor genético, como proposto por LEGARRA et al. (2008):

$$\hat{r}_{a\hat{a}g} = r(\mathbf{y}_g^*, \hat{\mathbf{y}}_g^*) / h \quad [4]$$

em que: \mathbf{y}_g^* é o vetor dos fenótipos dos animais pertencentes ao grupo g ajustados para os efeitos fixos como na equação [1] e $\hat{\mathbf{y}}_g^*$ são os valores preditos a partir da validação cruzada, o DGV para predições diretas e GEBV para os métodos de aprimoramento genômico e h é a raiz quadrada da herdabilidade da característica. Para cada método, o coeficiente de regressão do \mathbf{y}_g^* sobre o $\hat{\mathbf{y}}_g^*$ foi calculado para avaliar o grau de inflação/deflação das predições genômicas (DGV/GEBV). Modelos não viesados são esperados ter coeficiente de regressão de 1.

6.3 RESULTADOS E DISCUSSÃO

Os componentes de (co)variância e herdabilidades de cada característica encontram-se na Tabela 8.

Tabela 8. Médias a posteriori e desvio-padrão (DP) das variâncias genéticas aditivas direta (σ_a^2), variâncias genéticas aditivas maternas (σ_m^2), (co)variâncias genéticas aditiva-materna (σ_{am}), variâncias de ambiente permanente materno (σ_{pe}^2) e residuais (σ_e^2), herdabilidades diretas (h_a^2) e maternas (h_m^2) para as características estudadas.

Características ¹	σ_a^2	σ_m^2	σ_{am}	σ_{pe}^2	σ_e^2	$h_a^2 \pm DP$	$h_m^2 \pm DP$
PN	4,98 (±0,25)	1,23 (±0,13)	-0,93 (±0,14)	0,61 (±0,08)	12,21 (±0,18)	0,28 (±0,01)	0,07 (±0,01)
PD205	123,39 (±9,06)	50,29 (±5,17)	-27,48 (±5,22)	133,34 (±3,85)	284,03 (±6,20)	0,22 (±0,01)	0,09 (±0,01)
CD	0,129 (±0,01)	0,077 (±0,01)	-0,047 (±0,01)	0,136 (±0,01)	0,569 (±0,01)	0,15 (±0,01)	0,09 (±0,01)
PD	0,134 (±0,01)	0,061 (±0,01)	-0,046 (±0,01)	0,135 (±0,01)	0,659 (±0,01)	0,15 (±0,01)	0,06 (±0,01)
MD	0,159 (±0,01)	0,066 (±0,01)	-0,045 (±0,01)	0,142 (±0,01)	0,655 (±0,01)	0,16 (±0,01)	0,07 (±0,01)
TD	0,109 (±0,01)	0,054 (±0,01)	-0,016 (±0,01)	0,105 (±0,01)	0,497 (±0,01)	0,15 (±0,01)	0,07 (±0,01)
PS550	273,14 (±11,69)	-	-	-	761,30 (±10,55)	0,26 (±0,01)	-
GPD345	143,24 (±11,25)	-	-	-	1267,94 (±12,37)	0,10 (±0,01)	-
CS	0,148 (±0,01)	-	-	-	0,622 (±0,01)	0,19 (±0,01)	-
PS	0,164 (±0,01)	-	-	-	0,701 (±0,01)	0,19 (±0,01)	-
MS	0,168 (±0,01)	-	-	-	0,704 (±0,01)	0,19 (±0,01)	-
TS	0,195 (±0,01)	-	-	-	0,531 (±0,01)	0,27 (±0,01)	-

¹PN: peso ao nascer; PD205: peso a desmama ajustado para os 205 dias de idade; CD: conformação a desmama; PD: precocidade a desmama; MD: musculatura a desmama; TD: tamanho a desmama; PS550: peso ao sobreano ajustado para os 550 dias de idade; GPD345: ganho pós-desmama ajustado para os 345 dias de idade; CS: conformação ao sobreano; PS: precocidade ao sobreano; MS: musculatura ao sobreano; TS: tamanho ao sobreano; GC: grupos de contemporâneos.

6.3.1 Agrupamento da validação cruzada

Agrupamento pelo método k-médias produziu grupos com número de animais desbalanceados para as características medidas a desmama e ao sobreano (Tabelas 9 e 10), demonstrando que a média de parentesco genômico (G_{ij}) dentro de um grupo foram maiores do que a média entre grupos. Diferentemente de SAATCHI et al. (2011; 2012; 2013), que usaram a matriz de parentesco tradicional para formação dos grupos, neste trabalho utilizamos o parentesco genômico entre os animais para o

agrupamento conforme BODDHIREDDY et al., (2014), devido a presença de informações incompletas de pedigree para parte dos animais utilizados no presente estudo. Os resultados mostram que o agrupamento k-médias particionou os grupos primeiramente pela composição racial (um grupo contendo praticamente animais da raça Hereford) e após, para os animais Braford, os grupos foram formados de acordo com o grau de endogamia e parentesco genômico.

Os resultados para o agrupamento de forma aleatória foram muito diferentes em relação ao método k-médias. Nesta metodologia de agrupamento, os grupos tiveram número de animais balanceado, com média similar de endogamia de 0,01 e composição racial de 30%, e não houve diferença nos valores de G_{ij} dentro e entre grupos, sendo estes próximos de zero (dados não mostrados), devido a centralização da matriz de parentesco genômico (VANRADEN, 2008). Portanto, o objetivo de formar grupos distintos para representar cenários com animais de parentesco mais distante e outro, com parentesco mais próximo, para as populações de treinamento e validação foi atingido.

Tabela 9. Número de animais (N), proporção de zebu, coeficiente de endogamia¹ e parentesco genômico dentro e entre grupos (G_{ij}) para agrupamento k-médias em características medidas a desmama.

Características ²	Grupo	N	Proporção de zebu	Coeficiente de endogamia	G_{ij} dentro de grupo	G_{ij} entre grupos
PN	1	597	0,01±0,06	0,08±0,03	0,13±0,03	-0,02±0,03
	2	270	0,37±0,04	-0,01±0,03	0,10±0,03	-0,001±0,03
	3	656	0,37±0,06	0,004±0,02	0,05±0,04	-0,01±0,05
	4	524	0,33±0,09	-0,003±0,03	0,009±0,04	-0,002±0,05
	5	1145	0,36±0,06	-0,007±0,03	0,007±0,04	-0,004±0,03
PD205	1	603	0,01±0,07	0,08±0,02	0,13±0,04	-0,02±0,06
	2	669	0,37±0,06	0,004±0,03	0,05±0,05	-0,01±0,05
	3	1167	0,36±0,06	-0,007±0,03	0,007±0,03	-0,004±0,03
	4	545	0,33±0,09	-0,002±0,03	0,008±0,03	-0,002±0,03
	5	271	0,37±0,05	-0,01±0,02	0,10±0,07	-0,001±0,04
CD, PD e MD	1	603	0,01±0,07	0,08±0,02	0,13±0,04	-0,02±0,05
	2	668	0,36±0,06	0,004±0,03	0,05±0,03	-0,01±0,03
	3	1168	0,34±0,06	-0,007±0,03	0,007±0,03	-0,003±0,03
	4	545	0,37±0,09	0,002±0,03	0,008±0,04	-0,001±0,05
	5	271	0,36±0,05	-0,01±0,02	0,10±0,06	-0,009±0,03
TD	1	568	0,01±0,07	0,08±0,02	0,13±0,04	-0,02±0,05
	2	266	0,37±0,05	-0,01±0,02	0,11±0,07	-0,009±0,03
	3	1126	0,36±0,06	-0,007±0,03	0,007±0,03	-0,003±0,03
	4	657	0,38±0,06	0,003±0,03	0,04±0,03	-0,01±0,03
	5	520	0,33±0,09	-0,002±0,03	0,009±0,04	-0,001±0,04

¹Diagonal dos elementos da matriz de parentesco genômica menos 1 ($G_{ii} - 1$)

²PN: peso ao nascer; PD205: peso a desmama ajustado para os 205 dias de idade; CD: conformação a desmama; PD: precocidade a desmama; MD: musculatura a desmama; TD: tamanho a desmama.

Tabela 10. Número de animais (N), proporção de zebu, coeficiente de endogamia¹ e parentesco genômico dentro e entre grupos (G_{ij}) para agrupamento k-médias em características medidas ao sobreano.

Características ²	Grupo	N	Proporção de zebu	Coeficiente de endogamia	G_{ij} dentro de grupo	G_{ij} entre grupos
PS550	1	229	0,37±0,05	-0,01±0,03	0,11±0,03	-0,008±0,03
	2	510	0,02±0,08	0,08±0,03	0,12±0,03	-0,02±0,03
	3	853	0,35±0,07	-0,007±0,02	0,03±0,04	-0,001±0,05
	4	871	0,37±0,06	0,002±0,03	0,03±0,04	-0,01±0,05
GPD345	1	462	0,02±0,08	0,08±0,02	0,13±0,04	-0,02±0,06
	2	538	0,38±0,06	0,002±0,03	0,05±0,05	-0,01±0,05
	3	873	0,36±0,06	-0,007±0,03	0,01±0,03	-0,004±0,03
	4	522	0,35±0,07	-0,003±0,03	0,004±0,03	-0,001±0,03
CS, MS e PS	1	507	0,34±0,08	-0,003±0,02	0,005±0,07	-0,001±0,04
	2	938	0,35±0,06	-0,007±0,03	0,01±0,04	-0,004±0,05
	3	493	0,02±0,07	0,08±0,03	0,13±0,03	-0,02±0,03
	4	566	0,38±0,06	0,001±0,03	0,05±0,03	-0,01±0,04
TS	1	469	0,38±0,06	0,002±0,003	0,05±0,06	-0,009±0,05
	2	477	0,34±0,08	-0,005±0,003	0,004±0,03	-0,001±0,03
	3	492	0,02±0,07	0,08±0,02	0,13±0,05	-0,002±0,05
	4	826	0,36±0,06	-0,006±0,002	0,01±0,03	-0,004±0,03

¹Diagonal dos elementos da matriz de parentesco genômica menos 1 ($G_{ij} - 1$).

²PS550: peso ao sobreano ajustado para os 550 dias de idade; GPD345: ganho pós-desmama ajustado para os 345 dias de idade; CS: conformação ao sobreano; PS: precocidade ao sobreano; MS: musculatura ao sobreano; TS: tamanho ao sobreano.

6.3.2 Acurácia e viés das predições genômicas

De modo geral, as correlações genéticas entre o fenótipo e o DGV obtido a partir da validação cruzada, utilizando diferentes metodologias para estimação dos efeitos dos marcadores, foram de moderada a alta magnitude para as características medidas na desmama e ao sobreano, variando entre 0,19 e 0,78 para os métodos de agrupamento k-médias e aleatório (Tabelas 11 e 12). Os métodos genômicos foram similares ao BLUP. Os maiores valores de correlação genética estimada foram com o método BayesB, de 0,74 e 0,78 com o agrupamento aleatório para as características CD e PN, respectivamente. Para todas as características, foram observadas maiores correlações na validação cruzada com os grupos formados aleatoriamente em relação ao K-médias.

Entre as características de crescimento, usando o agrupamento K-médias, as metodologias BayesB e GBLUP tiveram as maiores correlações genéticas, com valores idênticos de 0,42 para PN. Para o agrupamento aleatório, as maiores correlações genéticas foram obtidas utilizando o método BayesB, de 0,72 e 0,78, para PS550 e PN, respectivamente. As menores acurácias estimadas foram para GPD345

e PD205 o que pode ser devido a menor herdabilidade dessas características em relação a PN e PS550.

Para PN e PD205, as estimativas de correlações genéticas do presente estudo foram similares as obtidas por SAATCHI et al. (2013), que estimaram médias variando de 0,25 a 0,42, usando o BayesB e BayesC, analisando dados de bovinos Hereford, com agrupamento K-médias. No entanto, para PS550, o valor de acurácia deste estudo foi maior do que o encontrado pelos autores citados acima, de 0,27. SAATCHI et al. (2011; 2012), estimaram valores maiores aos encontrados nesse estudo, trabalhando com bovinos da raça Angus, Simental e Limousin, respectivamente, para as características de crescimento PN, PD205 e PS550, usando critérios de formação de grupos semelhantes e estimação dos efeitos dos marcadores com metodologia bayesiana. Diferenças nos valores estimados de correlação genética entre os estudos podem ocorrer devido a diferentes herdabilidades, quantidade de informação e arquitetura genética para cada característica.

Para as características de escores visuais medidos a desmama e sobreano, os maiores valores de correlação genética foram estimados utilizando as metodologias bayesianas (BayesB e BayesC) de 0,38 para TD e 0,45 para TS, usando o agrupamento k-médias. Quando os grupos foram formados de forma aleatória, os maiores valores também foram para esses métodos, de 0,74 e 0,73 para CD e TS, respectivamente. NEVES et al. (2014), analisando dados de bovinos da raça Nelore, estimaram valores de acurácia para CD, PD e MD de 0,21, 0,43 e 0,43, respectivamente, e de 0,18 para CS, 0,68 para PS e 0,70 para MS, utilizando os métodos GBLUP, BayesC e BayesLasso. A diferença nos valores encontrados no estudo de NEVES et al. (2014) em comparação ao presente estudo, pode ser devida a utilização de touros com alto valor de acurácia no conjunto de treinamento e a utilização um conjunto menor de animais genotipados do que no presente estudo. Outra razão para diferentes valores de acurácia é o padrão de desequilíbrio de ligação diferente entre populações de bovinos de corte de origem indica e taurina (O'BRIEN et al., 2014).

De forma geral, para a maioria das características estudadas, os métodos bayesianos foram superiores ao método GBLUP, fato que pode ser devido a pressuposição para os efeitos dos marcadores em cada metodologia. O GBLUP assume um modelo infinitesimal com muitos locus de pequenos efeitos e pressupõe que os efeitos de QTL possuem distribuição normal e variância constante para todos

segmentos cromossômicos. Já nos métodos bayesianos, os dados observados são combinados com informações a priori sobre os parâmetros desconhecidos, gerando assim probabilidades a posteriori destes parâmetros. Esses métodos viabilizam pressuposições de variâncias diferentes para os efeitos de cada conjunto de marcadores em um dado segmento cromossômico e considera que apenas uma fração desses marcadores tem efeito diferente de zero, como em BayesB e BayesC (MEUWISSEN et al., 2001; GARRICK, 2010). Portanto, métodos bayesianos que consideram que em muitos marcadores não existe variação genética, podem representar melhor a realidade em relação aos efeitos dos SNPs, considerando que as análises genômicas geralmente envolvem diversos marcadores, espera-se não possuam efeito.

A média dos coeficientes de regressão de y^* sobre as predições genômicas (DGV), calculados como indicadores do viés de predição, variaram entre 0,44 e 1,43 para o método de agrupamento k-médias e de 0,77 a 1,48 para a metodologia aleatória (Tabelas 11 e 12). De forma geral, os coeficientes para as características medidas na desmama foram maiores que 1 para os métodos bayesianos, portanto, as predições genômicas foram subestimadas. Para a metodologia frequentista as predições foram menores que 1 ou superestimadas tanto para o método de agrupamento K-médias, como para o aleatório. Para as características avaliadas na fase de sobreano, as predições foram superestimadas, tanto para os métodos frequentistas como para os bayesianos, com o agrupamento k-médias, e subestimadas para os métodos BayesC, BayesB e GBLUP, com os grupos formados de forma aleatória. O viés de predição pode ser um problema, particularmente se as predições de animais genotipados (DGV) e não genotipados (EBV) precisam ser comparados para a classificação dos animais e decisões de seleção. Neste caso, as previsões do DGV conteriam valores extremos, favorecendo os indivíduos que possuem informações genômicas (CARDOSO et al., 2015).

As acurácias dentro dos grupos, calculadas com base na equação [4], para todas as características, foram mais baixas para o grupo composto na sua maior parte por animais da raça Hereford pelo método de agrupamento K-médias (Figuras 4, 5 e 6), indicando que o conjunto de treinamento composto pela sua maioria de animais da raça Braford não irá estimar predições acurada para os animais Hereford no conjunto de validação. Estes resultados já eram esperados, pois CARDOSO et al. (2015), trabalhando com o mesmo conjunto de dados, mas para a característica de resistência

ao carrapato, utilizando a mesma metodologia para formação dos grupos, também encontraram menores acurácias para o grupo onde continham somente animais da raça Hereford. SAATCHI et al. (2013), utilizando uma população de treinamento com dados de bovinos Hereford da América do Norte, também não produziram previsões acuradas para os rebanhos da mesma raça no Uruguai e na Argentina. HABIÉR et al. (2007) mostraram que a acurácia das previsões genômicas nos candidatos a seleção diminui à medida que a média de parentesco dos indivíduos do conjunto de treinamento diminui.

Tabela 11. Correlações genéticas do fenótipo ajustado com o valor genômico direto (DGV) predito da validação cruzada (rg_{aa}) e coeficientes de regressão (β) usando diferentes métodos para características medidas a desmama.

Características ¹	Método ²	K-médias		Aleatório	
		rg_{aa}	β	rg_{aa}	β
PN	BLUP	0,40±0,15	0,76	0,63±0,20	0,91
	GBLUP	0,42±0,11	0,72	0,76±0,17	0,86
	Bayes C	0,41±0,09	0,79	0,77±0,16	0,92
	Bayes B	0,42±0,08	0,71	0,78±0,17	0,85
PD205	BLUP	0,29±0,05	1,03	0,45±0,12	1,18
	GBLUP	0,23±0,03	0,64	0,45±0,09	0,75
	Bayes C	0,27±0,05	1,03	0,45±0,12	1,03
	Bayes B	0,28±0,05	0,84	0,46±0,13	0,83
CD	BLUP	0,33±0,12	1,15	0,70±0,18	1,23
	GBLUP	0,35±0,08	1,13	0,73±0,19	1,07
	Bayes C	0,35±0,09	1,19	0,74±0,17	1,01
	Bayes B	0,34±0,09	1,31	0,73±0,18	1,10
PD	BLUP	0,20±0,06	1,08	0,24±0,08	1,12
	GBLUP	0,19±0,06	0,77	0,23±0,07	0,80
	Bayes C	0,20±0,06	1,26	0,24±0,08	0,96
	Bayes B	0,23±0,06	0,95	0,28±0,07	0,84
MD	BLUP	0,27±0,20	1,15	0,70±0,19	1,24
	GBLUP	0,22±0,06	0,91	0,73±0,21	1,02
	Bayes C	0,29±0,07	0,85	0,71±0,20	1,04
	Bayes B	0,31±0,19	1,07	0,72±0,22	1,09
TD	BLUP	0,31±0,07	1,07	0,39±0,09	1,14
	GBLUP	0,34±0,09	1,07	0,38±0,10	1,01
	Bayes C	0,38±0,11	1,23	0,42±0,12	1,09
	Bayes B	0,38±0,11	1,22	0,42±0,11	1,10

¹PN: peso ao nascer; PD205: peso a desmama ajustado para os 205 dias de idade; CD: conformação a desmama; PD: precocidade a desmama; MD: musculatura a desmama; TD: tamanho a desmama.

²BLUP: BLUP tradicional; GBLUP: BLUP genômico; BayesC: modelo misturado bayesiano com distribuição normal e ponto de massa em zero e probabilidade ($\pi=0,95$); BayesB: modelo misturado bayesiano com distribuição-t e ponto de massa em zero e probabilidade ($\pi=0,95$).

Tabela 12. Correlações genéticas do fenótipo ajustado com o valor genômico direto (DGV) predito da validação cruzada ($r_{g_{aa}}$) e coeficientes de regressão usando diferentes métodos para características medidas ao sobreano.

Características ¹	Método ²	K-médias		Aleatório	
		$r_{g_{aa}}$	β	$r_{g_{aa}}$	β
PS550	BLUP	0,30±0,14	0,85	0,68±0,18	0,86
	GBLUP	0,31±0,09	0,76	0,70±0,18	0,89
	Bayes C	0,33±0,10	1,04	0,71±0,17	1,14
	Bayes B	0,33±0,10	0,78	0,72±0,17	0,89
GPD345	BLUP	0,21±0,08	1,06	0,44±0,22	0,98
	GBLUP	0,21±0,09	0,86	0,45±0,22	1,18
	Bayes C	0,23±0,10	0,52	0,41±0,25	1,40
	Bayes B	0,24±0,11	0,71	0,45±0,22	0,88
CS	BLUP	0,28±0,13	0,87	0,67±0,19	0,95
	GBLUP	0,26±0,09	0,78	0,68±0,19	0,91
	Bayes C	0,26±0,08	1,08	0,68±0,19	1,28
	Bayes B	0,27±0,08	0,82	0,70±0,12	1,03
PS	BLUP	0,33±0,24	0,78	0,47±0,22	0,88
	GBLUP	0,34±0,22	0,73	0,60±0,20	1,03
	Bayes C	0,35±0,21	0,74	0,61±0,20	1,11
	Bayes B	0,36±0,21	0,70	0,61±0,21	1,06
MS	BLUP	0,22±0,05	0,88	0,52±0,18	1,03
	GBLUP	0,24±0,04	0,90	0,55±0,16	1,06
	Bayes C	0,24±0,04	0,97	0,56±0,17	1,18
	Bayes B	0,26±0,05	0,91	0,56±0,16	1,09
TS	BLUP	0,35±0,11	0,76	0,63±0,20	0,86
	GBLUP	0,42±0,08	0,96	0,72±0,15	1,03
	Bayes C	0,42±0,07	1,02	0,72±0,14	1,09
	Bayes B	0,45±0,07	1,04	0,73±0,15	1,09

¹PS550: peso ao sobreano ajustado para os 550 dias de idade; GPD345: ganho pós-desmama ajustado para os 345 dias de idade; CS: conformação ao sobreano; PS: precocidade ao sobreano; MS: musculatura ao sobreano; TS: tamanho ao sobreano.

²BLUP: BLUP tradicional; GBLUP: BLUP genômico; BayesC: modelo misturado bayesiano com distribuição normal e ponto de massa em zero e probabilidade ($\pi=0,95$); BayesB: modelo misturado bayesiano com distribuição-t e ponto de massa em zero e probabilidade ($\pi=0,95$).

GBLUP

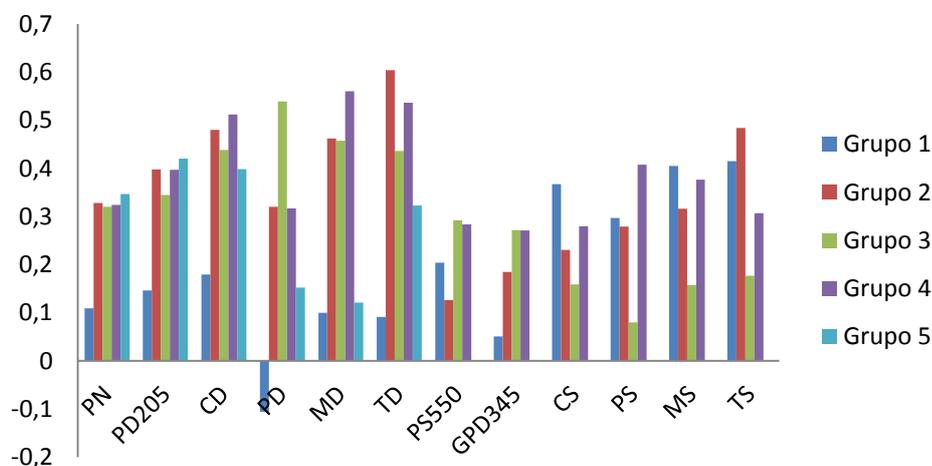


Figura 4. Acurácia dos valores genômicos diretos (DGV) preditos a partir da validação cruzada usando agrupamento k-médias com o método GBLUP.

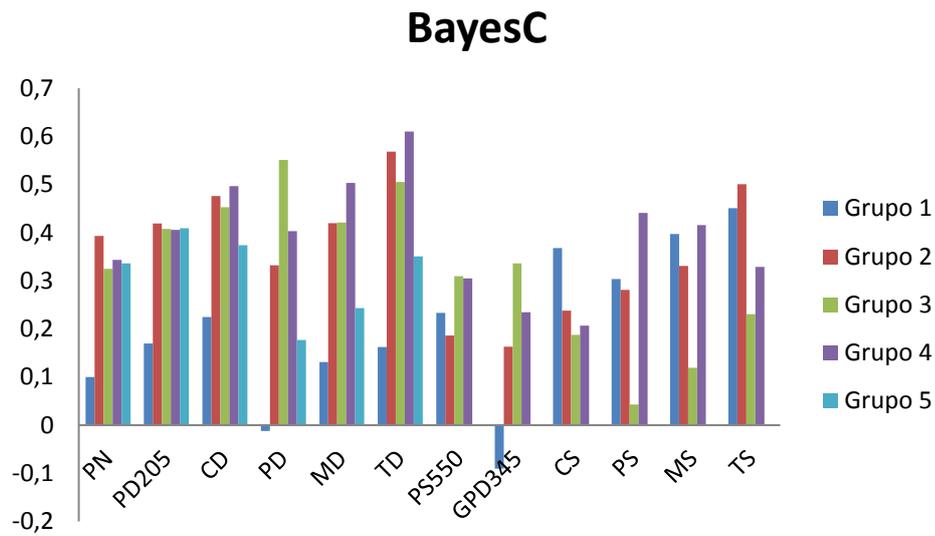


Figura 5. Acurácia dos valores genômicos diretos (DGV) preditos a partir da validação cruzada usando agrupamento k-médias com o método BayesC.

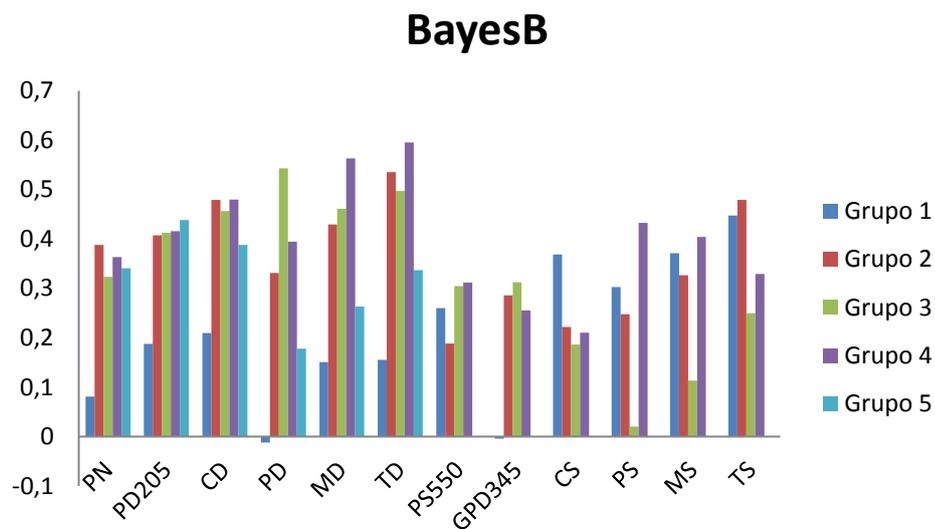


Figura 6. Acurácia dos valores genômicos diretos (DGV) preditos a partir da validação cruzada usando agrupamento k-médias com o método BayesB.

De forma geral, maiores ganhos em acurácia utilizando as metodologias de estimação dos efeitos dos marcadores em relação ao método BLUP (Tabelas 13 e 14), foram obtidos com os métodos bayesianos, tanto para as características medidas ao sobreano como a desmama. Para as características medidas a desmama, maior ganho foi alcançado com o método BayesB para PN, de 23,8% com agrupamento aleatório. Para fase de sobreano, maior ganho foi obtido com as metodologias BayesB e BayesC, de 29,8%, para PS, utilizando grupos formados aleatoriamente. Os

métodos bayesianos (BayesB e BayesC) tiveram maiores médias de ganho em relação ao método frequentista (GBLUP).

No entanto, para algumas características, o método tradicional teve desempenho melhor que o genômico, principalmente usando o K-médias. Isso pode ser devido ao fato de que a metodologia para o cálculo do DEBV proposta por GARRICK et al., (2009) remove a média dos pais, e o foco da estimativa é realizado com base na segregação mendeliana. De acordo SAATCHI et al., (2011) apesar deste método não excluir a informação de outros parentes, que são informativos para as meioses que produziram o touro que está sendo desregredido, se a maioria dos animais genotipados pertencerem a grandes linhagens paternas, a estimativa com o método tradicional pode ser superior. Outro fato que pode ter ocorrido em relação aos ganhos com o método tradicional, é pela maior distância genética do agrupamento K-médias, influenciando na menor estimativa da acurácia para o DGV.

Tabela 13. Ganhos em acurácia das metodologias² de estimação de efeitos de marcadores em relação ao método tradicional BLUP nas características medidas a desmama.

Características ¹	Ganho em acurácia (%)		
	GBLUP	BayesC K-médias	BayesB
PN	5,00	2,5	5,0
PD205	-20,7	-6,9	-3,4
CD	6,1	6,1	3,0
PD	-5,0	0,0	15,0
MD	-18,5	7,4	14,8
TD	9,7	22,6	22,6
Média	-3,9	5,3	9,5
		Aleatório	
PN	20,6	22,2	23,8
PD205	0,0	0,0	2,2
CD	4,3	5,7	4,3
PD	-4,2	0,0	16,2
MD	4,3	1,4	2,9
TD	-2,6	7,7	7,7
Média	3,7	6,2	9,5
Média Geral	-0,8	5,7	9,5

¹PN: peso ao nascer; PD205: peso a desmama ajustado para os 205 dias de idade; CD: conformação a desmama; PD: precocidade a desmama; MD: musculatura a desmama; TD: tamanho a desmama.

²GBLUP: BLUP genômico; BayesC: modelo misturado bayesiano com distribuição normal e ponto de massa em zero e probabilidade ($\pi=0,95$); BayesB: modelo misturado bayesiano com distribuição-t e ponto de massa em zero e probabilidade ($\pi=0,95$).

Tabela 14. Ganhos em acurácia das metodologias² de estimação de efeitos de marcadores em relação ao método tradicional BLUP nas características medidas ao sobreano.

Características ¹	Ganho em acurácia (%)		
	GBLUP	BayesC K-médias	BayesB
PS550	3,3	10,0	10,0
GPD345	0,0	9,5	14,3
CS	-7,1	-7,1	-3,6
PS	3,0	6,1	9,1
MS	9,1	13,6	18,2
TS	20,0	20,0	28,6
Média	4,7	8,7	12,8
	Aleatório		
PS550	2,9	1,4	1,4
GPD345	2,3	-6,8	2,3
CS	1,5	1,5	4,5
PS	27,7	29,8	29,8
MS	5,8	7,7	7,7
TS	14,3	14,3	15,9
Média	9,1	8,0	10,3
Média Geral	6,9	8,3	11,5

¹PS550: peso ao sobreano ajustado para os 550 dias de idade; GPD345: ganho pós-desmama ajustado para os 345 dias de idade; CS: conformação ao sobreano; PS: precocidade ao sobreano; MS: musculatura ao sobreano; TS: tamanho ao sobreano.

²GBLUP: BLUP genômico; BayesC: modelo misturado bayesiano com distribuição normal e ponto de massa em zero e probabilidade ($\pi=0,95$); BayesB: modelo misturado bayesiano com distribuição-t e ponto de massa em zero e probabilidade ($\pi=0,95$).

6.3.3 Incorporação da informação tradicional com a genômica

Incorporação da informação histórica de pedigree e fenótipos com o DGV foram desempenhadas com os métodos GBLUP e BayesB para estimação dos valores genéticos aprimorados pela genômica (GEBV). Foram utilizados 3 métodos alternativos e para a maioria das características as correlações genéticas entre o fenótipo e a predição genômica foram melhoradas (Tabelas 15 e 16). O método BayesB foi o escolhido para combinar as informações tradicionais entre os métodos bayesianos, pois foi o método com maior ganho entre os bayesianos e porque assume pressuposições mais distintas para os efeitos dos marcadores em relação ao GBLUP (CARDOSO et al., 2015). As análises utilizando o pedigree tradicional contribuíram bastante como informação adicional para cálculo do GEBV, pelo fato de a maioria das características possuírem um número de observações consideravelmente grande.

A primeira abordagem utilizada foi o índice de seleção proposto por VANRADEN et al. (2009), que é uma função das confiabilidades médias dos animais usados na predição. Neste índice, subtraindo-se o subconjunto dos animais genotipados (SPA), apenas a informação adicional contida nas predições genômicas

é adicionado ao conjunto completo (PA), com isso evitando a dupla contagem das predições do pedigree tradicional no subconjunto de dados utilizados para o treinamento dos efeitos dos marcadores (VANRADEN et al., 2009). Para todas as características foi obtido aumento na correlação genética quando comparado ao método tradicional usando a média dos pais (BLUP), com ambos os métodos GBLUP e BayesB, com aumento na acurácia variando entre 0,18 e 0,19 para PS550 e GPD345, com o método de agrupamento k-médias. Com os grupos formados aleatoriamente foram alcançados também aumentos de 0,18 e 0,19, porém para as características para PS e PD205, respectivamente.

Nas análises bivariadas (segunda abordagem), o DGV foi considerado como uma característica correlacionada como proposto por MACNEIL et al. (2010) e produziu dois GEBV para cada animal, um para os dados de fenótipo (a) e outro para o DGV (α). Nesta metodologia, há uma troca de informações através da $\sigma_{a\alpha}$, que é diretamente dependente da acurácia de predição (Tabelas 11 e 12). No presente estudo, o GEBV(a) foi um melhor preditor, com aumento de correlação genética de até 0,14 e 0,18 em relação ao BLUP tradicional, pelo método k-médias e aleatório, respectivamente, para características medidas a desmama. Para a fase de sobreano, houve aumento de 0,21 para grupos k-médias e de 0,18 para aleatório. Nesta metodologia bivariada, para algumas características a combinação do DGV com a informação tradicional não foi superior ao BLUP, principalmente para o GEBV(α). Este fato pode ocorrer devido a quantidade de informações históricas de pedigree utilizado no método BLUP.

Os maiores valores de correlações genéticas para a maioria das características, entre as metodologias testadas, foram para o método ssGBLUP (Tabelas 15 e 16). O maior aumento foi para PN, de 0,22 com o agrupamento k-médias e de 0,15 para TD e TS, quando os grupos foram formados de forma aleatória. Esta abordagem tem sido utilizada em avaliações genéticas de gado de leite e corte (SAATCHI et al., 2013). A metodologia combina diretamente todas as fontes de informação (fenótipo, pedigree e genômica) (MISZTAL et al, 2009; AGUILAR et al., 2010; CHRISTENSEN & LUND, 2010), e a matriz inversa de \mathbf{H}^{-1} , necessária para implementação do método, envolve simplesmente a habitual \mathbf{A}^{-1} tradicional, e uma submatriz correspondente aos animais genotipados, que é aumentada pela adição da diferença entre a inversa da matriz de parentesco genômica e a inversa da matriz de

parentesco para os animais genotipados. No entanto, este método considera que todos os locus tem o mesmo peso (SAATCHI et al., 2013).

Tabela 15. Correlações genéticas do fenótipo ajustado com o valor genético aprimorado pela genômica (GEBV) predito da validação cruzada (rg_{aa}) e coeficientes de regressão (β) usando diferentes métodos para características medidas a desmama.

Características ¹	Método ²	K-médias		Aleatório	
		rg_{aa}	β	rg_{aa}	β
PN	GBLUP_Biv(a)	0,43±0,09	0,62	0,63±0,18	0,51
	GBLUP_Biv(α)	0,43±0,11	0,73	0,75±0,17	0,87
	IS_GBLUP	0,53±0,17	0,77	0,77±0,15	0,88
	BayesB_Biv(a)	0,43±0,09	0,62	0,63±0,18	0,53
	BayesB_Biv(α)	0,40±0,08	0,72	0,76±0,16	0,87
	IS_BayesB	0,52±0,13	0,78	0,77±0,13	0,92
	ssGBLUP	0,53±0,12	0,76	0,78±0,16	1,12
	PD205	GBLUP_Biv(a)	0,40±0,03	1,35	0,63±0,03
GBLUP_Biv(α)		0,24±0,04	0,76	0,65±0,16	0,86
IS_GBLUP		0,34±0,05	0,73	0,65±0,11	0,87
BayesB_Biv(a)		0,40±0,03	1,34	0,65±0,15	1,23
BayesB_Biv(α)		0,25±0,04	0,75	0,63±0,04	0,86
IS_BayesB		0,36±0,06	0,79	0,64±0,12	0,82
ssGBLUP		0,40±0,06	0,86	0,65±0,12	0,94
CD		GBLUP_Biv(a)	0,38±0,09	1,14	0,82±0,17
	GBLUP_Biv(α)	0,33±0,07	1,19	0,82±0,18	1,15
	IS_GBLUP	0,42±0,11	1,02	0,79±0,13	1,06
	BayesB_Biv(a)	0,38±0,09	0,09	0,82±0,17	0,72
	BayesB_Biv(α)	0,33±0,07	0,07	0,82±0,17	1,19
	IS_BayesB	0,36±0,14	1,08	0,79±0,12	1,06
	ssGBLUP	0,43±0,11	1,08	0,80±0,13	1,14
	PD	GBLUP_Biv(a)	0,31±0,07	1,31	0,36±0,07
GBLUP_Biv(α)		0,20±0,06	1,02	0,22±0,06	0,91
IS_GBLUP		0,30±0,07	0,83	0,37±0,10	0,90
BayesB(a)		0,35±0,06	1,21	0,38±0,08	1,19
BayesB(α)		0,19±0,05	0,84	0,21±0,05	0,86
IS_BayesB		0,32±0,08	0,89	0,38±0,10	0,94
ssGBLUP		0,33±0,08	0,89	0,39±0,09	0,96
MD		GBLUP_Biv(a)	0,39±0,18	1,17	0,73±0,20
	GBLUP_Biv(α)	0,22±0,07	0,94	0,72±0,21	1,06
	IS_GBLUP	0,28±0,09	0,86	0,80±0,17	1,03
	BayesB_Biv(a)	0,41±0,20	0,96	0,67±0,23	0,75
	BayesB_Biv(α)	0,32±0,19	1,08	0,71±0,22	1,09
	IS_BayesB	0,35±0,18	0,91	0,81±0,16	1,05
	ssGBLUP	0,46±0,17	1,01	0,82±0,16	1,10
	TD	GBLUP_Biv(a)	0,43±0,05	1,29	0,48±0,04
GBLUP_Biv(α)		0,37±0,09	1,15	0,39±0,11	1,11
IS_GBLUP		0,37±0,06	0,98	0,43±0,08	0,99
BayesB_Biv(a)		0,37±0,06	0,91	0,39±0,06	0,92
BayesB_Biv(α)		0,38±0,11	1,25	0,42±0,12	1,12
IS_BayesB		0,35±0,05	0,92	0,44±0,08	0,94
ssGBLUP		0,47±0,10	1,08	0,54±0,12	1,09

¹PN: peso ao nascer; PD205: peso a desmama ajustado para os 205 dias de idade; CD: conformação a desmama; PD: precocidade a desmama; MD: musculatura a desmama; TD: tamanho a desmama.

²GBLUP: BLUP genômico; BayesB: modelo misturado bayesiano com distribuição-t e ponto de massa em zero e probabilidade ($\pi=0,95$); Biv: análise bivariada combinando o fenótipo e o DGV produzindo 2 GEBV, um para dados fenotípicos (a) e um para os dados dos valores genômicos diretos; IS: Índice de seleção ponderando o DGV, média dos pais usando todos os animais (PA) e média dos pais usando apenas os dados dos animais genotipados (SPA); ssGBLUP: GBLUP de passo único.

Tabela 16. Correlações genéticas do fenótipo ajustado com o valor genético aprimorado pela genômica (GEBV) predito da validação cruzada (rg_{aa}) e coeficientes de regressão (β) usando diferentes métodos para características medidas ao sobreano.

Características ¹	Método ²	K-médias		Aleatório	
		rg_{aa}	β	rg_{aa}	β
PS550	GBLUP_Biv(a)	0,39±0,12	0,73	0,68±0,19	0,49
	GBLUP_Biv(α)	0,32±0,10	0,80	0,70±0,18	0,93
	IS_GBLUP	0,48±0,14	0,85	0,75±0,15	0,88
	BayesB_Biv(a)	0,40±0,12	0,74	0,68±0,18	0,50
	BayesB_Biv(α)	0,33±0,10	0,81	0,72±0,18	0,94
	IS_BayesB	0,49±0,15	0,89	0,76±0,14	0,91
	ssGBLUP	0,52±0,17	0,79	0,78±0,14	0,80
GPD345	GBLUP_Biv(a)	0,24±0,09	0,86	0,44±0,22	0,65
	GBLUP_Biv(α)	0,21±0,09	0,90	0,48±0,22	1,20
	IS_GBLUP	0,35±0,12	1,06	0,52±0,22	0,95
	BayesB_Biv(a)	0,26±0,10	0,85	0,43±0,22	0,70
	BayesB_Biv(α)	0,24±0,11	0,75	0,45±0,22	0,92
	IS_BayesB	0,39±0,14	0,98	0,53±0,22	0,84
	ssGBLUP	0,25±0,09	0,73	0,57±0,22	0,66
CS	GBLUP_Biv(a)	0,37±0,12	0,78	0,66±0,19	0,58
	GBLUP_Biv(α)	0,29±0,09	0,81	0,68±0,19	0,94
	IS_GBLUP	0,39±0,13	0,79	0,74±0,16	0,90
	BayesB_Biv(a)	0,37±0,12	0,79	0,66±0,19	0,58
	BayesB_Biv(α)	0,27±0,08	0,84	0,70±0,18	1,05
	IS_BayesB	0,40±0,12	0,82	0,75±0,16	0,97
	ssGBLUP	0,42±0,11	0,78	0,78±0,15	0,92
PS	GBLUP_Biv(a)	0,35±0,23	0,69	0,48±0,19	0,66
	GBLUP_Biv(α)	0,36±0,22	0,75	0,60±0,20	1,05
	IS_GBLUP	0,46±0,22	0,87	0,65±0,19	1,02
	BayesB(a)	0,35±0,22	0,70	0,48±0,20	0,65
	BayesB(α)	0,34±0,21	0,72	0,61±0,20	1,08
	IS_BayesB	0,45±0,23	0,86	0,65±0,19	1,04
	ssGBLUP	0,47±0,22	0,79	0,65±0,19	0,94
MS	GBLUP_Biv(a)	0,28±0,05	0,83	0,51±0,15	0,76
	GBLUP_Biv(α)	0,24±0,05	0,93	0,56±0,16	1,09
	IS_GBLUP	0,32±0,06	0,92	0,63±0,16	1,02
	BayesB_Biv(a)	0,28±0,05	0,83	0,51±0,14	0,75
	BayesB_Biv(α)	0,24±0,04	0,95	0,57±0,17	1,12
	IS_BayesB	0,32±0,06	0,93	0,63±0,17	1,04
	ssGBLUP	0,33±0,05	0,81	0,66±0,16	0,99
TS	GBLUP_Biv(a)	0,43±0,08	0,76	0,68±0,15	0,66
	GBLUP_Biv(α)	0,42±0,08	1,03	0,73±0,15	1,07
	IS_GBLUP	0,55±0,17	0,85	0,75±0,14	1,14
	BayesB_Biv(a)	0,45±0,09	0,78	0,75±0,14	1,55
	BayesB_Biv(α)	0,43±0,08	1,11	0,77±0,15	1,14
	IS_BayesB	0,55±0,17	0,89	0,75±0,15	1,01
	ssGBLUP	0,56±0,10	0,91	0,78±0,13	1,02

¹PS550: peso ao sobreano ajustado para os 550 dias de idade; GPD345: ganho pós-desmama ajustado para os 345 dias de idade; CS: conformação ao sobreano; PS: precocidade ao sobreano; MS: musculatura ao sobreano; TS: tamanho ao sobreano.

²GBLUP: BLUP genômico; BayesB: modelo misturado bayesiano com distribuição-t e ponto de massa em zero e probabilidade ($\pi=0,95$); Biv: análise bivariada combinando o fenótipo e o DGV produzindo 2 GEBV, um para dados fenotípicos (a) e um para os dados dos valores genômicos diretos; IS: Índice de seleção ponderando o DGV, média dos pais usando todos os animais (PA) e média dos pais usando apenas os dados dos animais genotipados (SPA); ssGBLUP: GBLUP de passo único.

O padrão dos resultados das acurácias dentro de grupos para os GEBVs (equação 4) foram similares aos obtidos com os DGV, pois os grupos que possuem o maior número de animais da raça Hereford, tiveram os resultados de menores acurácias em relação com os outros grupos (Figuras 7, 8, 9, 10 e 11), porém a magnitude dos valores de acurácia foi maior para GEBV comparada à acurácia do DGV. Este fato ocorreu porque as acurácias entre grupos utilizando o método tradicional (BLUP) apresentaram-se semelhantes entre todos os grupos (dados não mostrados), devido ao cálculo do EBV utilizando a média dos pais possuir informação adicional de seus ancestrais. Portanto, quando ao DGV foi adicionada a informação histórica de pedigree e de desempenho para o cálculo do GEBV, os valores de acurácia entre grupos aumentaram mesmo para os grupos composto pela sua maioria de animais Hereford.

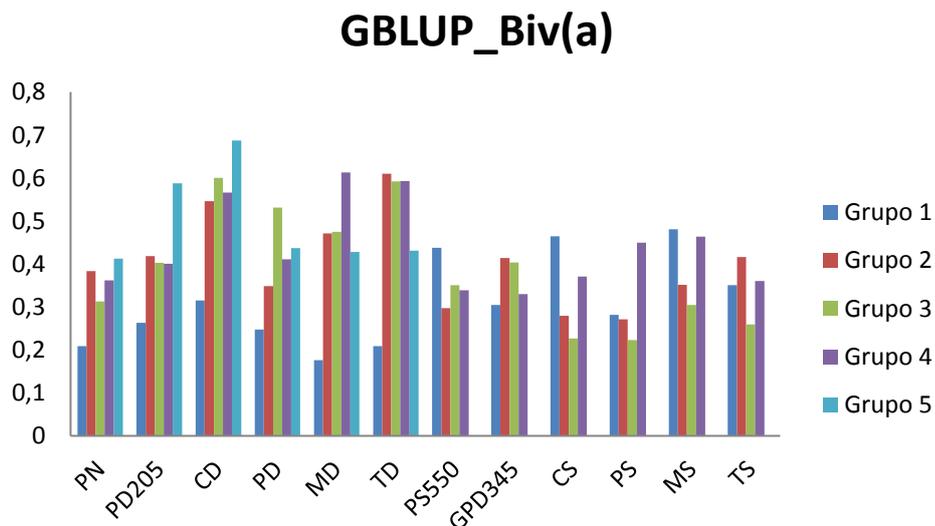


Figura 7. Acurácia dos valores genéticos aprimorados pela genômica (GEBV) preditos a partir da validação cruzada usando agrupamento k-médias com o método Biv_GBLUP(a).

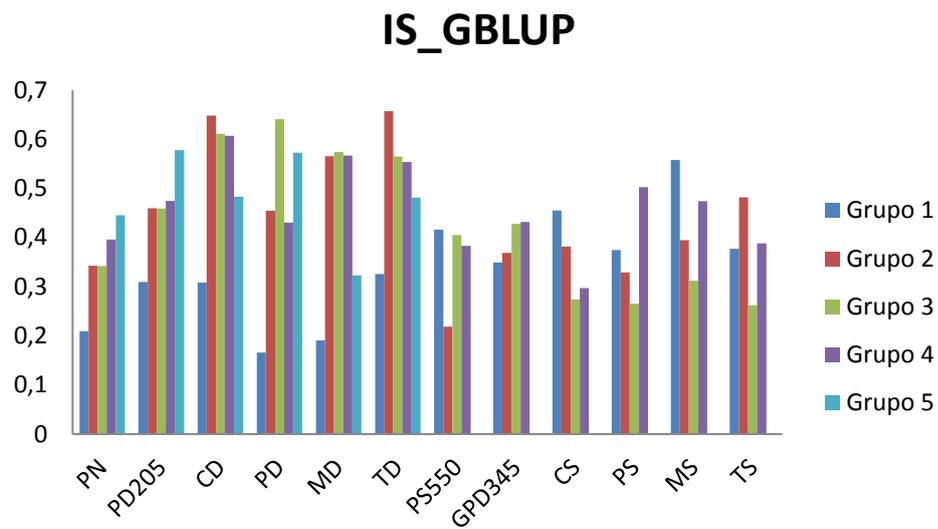


Figura 8. Acurácia dos valores genéticos aprimorados pela genômica (GEBV) preditos a partir da validação cruzada usando agrupamento k-médias com o método IS_GBLUP.

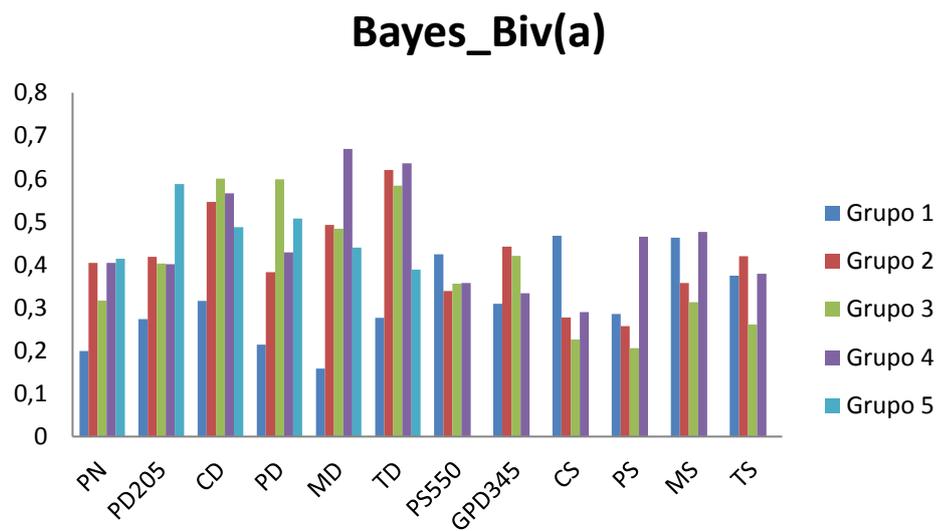


Figura 9. Acurácia dos valores genéticos aprimorados pela genômica (GEBV) preditos a partir da validação cruzada usando agrupamento k-médias com o método Biv_BayesB(a).

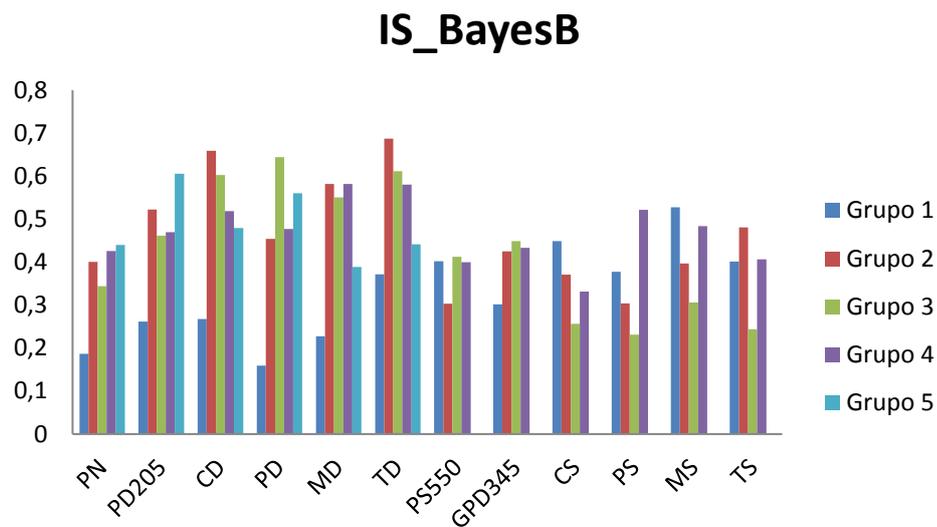


Figura 10. Acurácia dos valores genéticos aprimorados pela genômica (GEBV) preditos a partir da validação cruzada usando agrupamento k-médias com o método IS_BayesB.

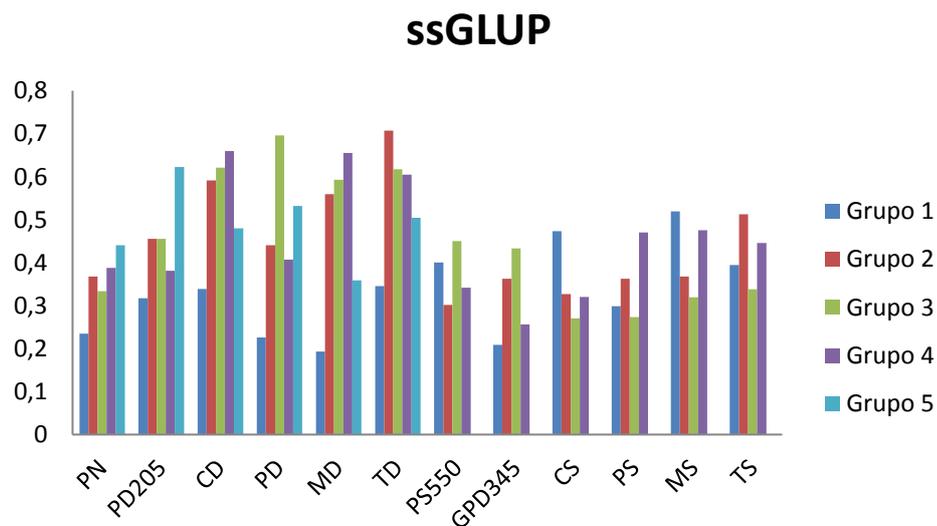


Figura 11. Acurácia dos valores genéticos aprimorados pela genômica (GEBV) preditos a partir da validação cruzada usando agrupamento k-médias com o método ssGLUP.

Os coeficientes de regressão para o fenótipo ajustado sobre o GEBV, tanto para as características de desmama como de sobreano, em qualquer um dos métodos utilizados, em média foram menores que 1, portanto, produziram previsões superestimadas. Para reduzir o viés das previsões genômicas, futuras pesquisas devem ser realizadas, usando por exemplo, o método single step bayesiano (GARRICK et al., 2014), o qual considera diferentes pressuposições para os efeitos dos marcadores, diferentemente do ssGBLUP.

Para todas as características foram obtidos ganhos em acurácia em relação ao método tradicional (Tabelas 17 e 18). O método com a melhor média de ganho em relação ao BLUP foi o ssGBLUP, com média de 40,70% entre todas as características medidas a desmama e de 36,70 % para as avaliadas na fase de sobreano. De acordo com AGUILAR et al., (2010) o procedimento de uma única etapa fornece uma estrutura unificada, elimina vários pressupostos e parâmetros e reduz o viés das predições por usa todas as fontes informações permitindo calcular avaliações genômicas mais acuradas do que os procedimentos de múltiplas etapas.

Tabela 17. Ganhos em acurácia das metodologias² de aprimoramento genômico em relação ao método tradicional BLUP nas características medidas a desmama.

Características ¹	Ganho em acurácia (%)				
	GBLUP_Biv	IS_GBLUP	BayesB_Biv	IS_BayesB	ssGBLUP
	K-médias				
PN	7,5	32,5	7,5	30,0	32,5
PD205	37,9	17,2	37,9	24,1	37,9
CD	15,2	27,3	15,2	9,1	30,3
PD	55,0	50,0	75,0	60,0	65,0
MD	44,4	3,7	51,9	29,6	70,4
TD	38,7	19,4	22,6	12,9	51,6
Média	33,2	25,0	35,01	27,6	47,9
	Aleatório				
PN	19,0	22,2	20,6	22,2	23,8
PD205	40,0	44,4	44,4	42,2	44,4
CD	17,1	12,9	17,1	12,9	14,3
PD	50,0	54,2	58,3	58,3	62,5
MD	4,3	14,3	1,4	15,7	17,1
TD	23,1	10,3	7,7	12,8	38,5
Média	25,6	26,4	24,9	27,3	33,4
Média Geral	29,3	25,7	29,2	27,5	40,7

¹PN: peso ao nascer; PD205: peso a desmama ajustado para os 205 dias de idade; CD: conformação a desmama; PD: precocidade a desmama; MD: musculatura a desmama; TD: tamanho a desmama.

²GBLUP: BLUP genômico; BayesB: modelo misturado bayesiano com distribuição-t e ponto de massa em zero e probabilidade ($\pi=0,95$); Biv: análise bivariada combinando o fenótipo e o DGV produzindo 2 GEBV, um para dados fenotípicos (a) e um para os dados dos valores genômicos diretos; IS: Índice de seleção ponderando o DGV, média dos pais usando todos os animais (PA) e média dos pais usando apenas os dados dos animais genotipados (SPA); ssGBLUP: GBLUP de passo único.

Tabela 18. Ganhos em acurácia das metodologias² de aprimoramento genômico em relação ao método tradicional BLUP nas características medidas ao sobreano.

Características ¹	Ganho em acurácia (%)				
	GBLUP_Biv	IS_GBLUP	BayesB_Biv	IS_BayesB	ssGBLUP
	K-médias				
PS550	30,0	60,0	33,3	63,3	73,3
GPD345	14,3	66,7	23,8	85,7	19,0
CS	23,3	30,0	23,3	30,0	43,3
OS	9,1	39,4	6,1	36,4	42,4
MS	27,3	45,5	27,3	45,5	50,0
TS	22,9	54,3	28,6	54,3	62,9
Média	21,1	49,3	27,7	52,5	48,5
	Aleatório				
PS550	2,9	10,3	5,9	11,8	14,7
GPD345	9,1	18,2	2,3	20,5	29,5
CS	1,5	10,4	4,5	11,9	16,4
OS	27,7	38,3	29,8	38,3	38,3
MS	7,7	21,2	9,6	21,2	26,9
TS	15,9	19,0	22,2	19,0	23,8
Média	10,8	19,6	12,3	20,4	25,0
Média Geral	15,9	34,4	18,1	36,4	37,7

¹PS550: peso ao sobreano ajustado para os 550 dias de idade; GPD345: ganho pós-desmama ajustado para os 345 dias de idade; CS: conformação ao sobreano; PS: precocidade ao sobreano; MS: musculatura ao sobreano; TS: tamanho ao sobreano.

²GBLUP: BLUP genômico; BayesB: modelo misturado bayesiano com distribuição-t e ponto de massa em zero e probabilidade ($\pi=0,95$); Biv: análise bivariada combinando o fenótipo e o DGV produzindo 2 GEBV, um para dados fenotípicos (a) e um para os dados dos valores genômicos diretos; IS: Índice de seleção ponderando o DGV, média dos pais usando todos os animais (PA) e média dos pais usando apenas os dados dos animais genotipados (SPA); ssGBLUP: GBLUP de passo único.

Dentre os métodos testados, apesar do ssGBLUP ter tido os melhores resultados de acurácia para a maioria das características, de acordo com CARDOSO et al., (2015), a escolha da metodologia testada irá depender do modelo de negócios adotado para os testes genômicos e a estrutura dos programas de avaliação genética. Por exemplo, se companhias privadas entregam o DGV sem os genótipos individuais, como aconteceu inicialmente na American Angus Association (AAA) desde 2009, análises multi passos como o modelo bivariado proposto por MACNEIL et al., (2010) é indicado, bem como o índice de seleção, usado em avaliações de gado de leite (VANRADEN et al., 2009). Para melhorar a predição e simplificar os procedimentos, AAA está em transição para implementação do ssGBLUP (LOURENÇO et al., 2016). De acordo com os autores, testes iniciais com ssGBLUP mostraram um aumento da acurácia de 25% para as características de crescimento em comparação com as avaliações tradicionais.

6.4 CONCLUSÕES

Entre as metodologias para estimação dos efeitos dos marcadores, maiores valores de acurácia foram estimados usando o método bayesiano BayesB. As acurácias para os grupos que continham um maior número de animais Hereford, de forma geral, foram mais baixas do que os valores para os grupos que apresentavam animais Braford, sugerindo a inclusão de mais animais com fenótipos e genótipos desta raça na população de treinamento, para melhorar a estimação dos efeitos dos marcadores.

Menores estimativas de acurácia foram estimadas utilizando o método de agrupamento K-médias, indicando que a acurácia as predições genômicas diminui à medida que o parentesco entre os animais no conjunto de treinamento é mais distante dos animais a serem avaliadas e selecionados com uso de informação genômica.

Os ganhos obtidos com os métodos combinando as informações tradicionais com a genômica em relação ao método BLUP tradicional indicam que as predições genômicas devem ser usadas como ferramenta para melhorar os ganhos genéticos, mas sendo usadas no programa de avaliação genética de forma integrada com as informações históricas de pedigree e fenótipos. Neste sentido, os melhores resultados foram obtidos com o método ssGBLUP.

7. CAPÍTULO III

SELEÇÃO DE MARCADORES MOLECULARES UTILIZANDO ESTUDO DE ASSOCIAÇÃO GENÔMICA AMPLA BAYESIANO PARA CARACTERÍSTICAS DE CRESCIMENTO EM BOVINOS DA RAÇA HEREFORD E BRAFORD

7.1 INTRODUÇÃO

A seleção para características de crescimento utilizando dados de pedigree e fenótipos através do uso de metodologias tradicionais, está sendo efetiva ao longo dos anos. No entanto, esta pode ser aprimorada se polimorfismos de DNA responsáveis por diferenciais no crescimento puderem ser determinados (SNELLING et al., 2010).

Com estudos de associação genômica ampla (GWAS), pode-se identificar SNPs associados com a característica de interesse e explorar as regiões do genoma em que esses SNPs estão inseridos, buscando genes e/ou loci de características quantitativas (QTL – Quantitative Trait Loci) relacionados à expressão fenotípica da característica (UTSUNOMIYA et al., 2013).

Inicialmente, nas análises de GWAS eram usados métodos que se baseavam em um teste de um único marcador por vez (HIRSCHHORN & DALY, 2005). No entanto, em razão da dificuldade em realizar teste de significância em cada marcador, considerando que existem milhares de SNPs (WANG et al., 2012), outros métodos em que ajustam os SNPs simultaneamente como efeitos aleatórios são capazes de explicar uma maior parte da variância genética (ONTERU et al., 2010; FAN et al., 2011) em relação aos que testam um marcador por vez. Métodos de regressão bayesiana, que combinam informações a priori sobre os efeitos de marcadores com os fenótipos de características de interesse, podem ser usados para estimar todos os efeitos dos marcadores ao mesmo tempo (MEUWISSEN et al., 2001).

Estudos de associação genômica ampla (GWAS) podem ser usados para detecção de QTL, construção de redes de genes e seleção genômica (EGGEN 2012; SNELLING et al., 2012). A identificação dos marcadores mais informativos associados a características de interesse, pode levar ao delineamento de painéis de marcadores de baixa densidade que explicam boa parte da variância genética para caracteres de interesse. Se comprovado um bom desempenho de predição, esses painéis podem se tornar soluções altamente desejáveis e de baixo custo, necessárias para a implementação da seleção genômica pela maioria dos setores de produção animal (VAN EENENNAAM et al., 2014).

O presente trabalho teve como objetivo a realização de um estudo de associação genômica ampla (GWAS), utilizando metodologia Bayesiana para

identificação de regiões genômicas e seleção de SNPs mais representativos associados com as características de peso ao nascer (PN), peso a desmama ajustado para os 205 dias (PD205), peso ao sobreano ajustado para os 550 dias (PS550) e o ganho de peso pós-desmama ajustado para os 345 dias (GPD345) em bovinos das raças Hereford e Braford. Outro objetivo foi verificar a acurácia de predição do painel de baixa densidade também foi estimado e comparado com os resultados do conjunto completo de marcadores SNPs.

7.2 MATERIAL E MÉTODOS

7.2.1 Arquivo de dados fenotípicos, genotípicos e pedigree

Foram utilizados registros fenotípicos de 126.290 animais das raças Hereford e Braford nascidos entre os anos de 1991 a 2012, pertencentes ao programa de melhoramento Conexão Delta G, localizados no estado do Rio Grande do Sul, Brasil. As características estudadas foram: peso ao nascer (PN), peso a desmama ajustado aos 205 dias de idade (PD205), peso ao sobreano ajustado aos 550 dias de idade (PS550) e o ganho de peso pós-desmama ajustado para aos 345 dias de idade (GPD345).

Foi utilizado um conjunto de 3.545 animais genotipados com o Illumina BovineSNP50 Bead Chip (50K; ILLUMINA, San Diego, CA) e 131 touros pais com o Illumina High-Density Bovine Bead Chip Array (HD; ILLUMINA). O controle de qualidade foi implementado usando R/SNPStats package (CLAYTON, 2014). Foram removidas amostras com taxa de genotipagem (Call Rate – CR) < 0,90, heterozigocidade com 3 desvios acima ou abaixo da média observada desse critério na população genotipada, e com erros de identificação de amostras. Esses erros foram constatados quando o sexo determinado através dos genótipos não correspondia ao sexo do indivíduo associado com tal amostra e quando duas amostras estavam associadas a indivíduos diferentes, mas tinham genótipos idênticos. Apenas SNPs mapeados para autossomas, com CR > 0,98, frequências alélicas menores (MAF) > 0,03, e desvio de Equilíbrio de Hardy-Weinberg não altamente significativo ($P > 10^{-7}$), foram consideradas para análises subsequentes. Além disso, apenas o SNP com maior MAF foi mantido quando havia SNPs na mesma posição ou os genótipos estavam altamente correlacionados ($r > 0,98$). Os dados do chip de alta densidade (HD) foram filtrados para selecionar somente SNPs que também estavam presentes no painel de 50K. Após a edição, um total de 41.045 marcadores SNPs e 3.592 amostras foram utilizadas nas análises, incluindo, 2934 animais da raça Braford e 658 da raça Hereford. Dados perdidos em genótipos (0,86% de todos os genótipos) foram imputados usando o software FImpute (SARGOLZAEI et al., 2011).

7.2.2 Análises e modelo estatístico

Primeiramente, foram estimados os componentes de (co)variâncias e parâmetros genéticos usando inferência bayesiana por amostragem de Gibbs, com o programa Gibbs1f90 (MISZTAL et al., 2002) em análises uni-características. Foram considerados 800.000 ciclos, com descarte inicial de 200.000 ciclos e intervalo de amostragem de 20. Posteriormente, os componentes de (co)variâncias estimados foram utilizados na predição dos valores genéticos, usando o software BLUPF90 (MISZTAL et al., 2002). O seguinte modelo estatístico, em notação matricial, foi utilizado nas análises:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{Wm} + \mathbf{Spe} + \mathbf{e} \quad [1]$$

em que: \mathbf{y} é o vetor de observações, \mathbf{X} , \mathbf{Z} , \mathbf{W} e \mathbf{S} são as matrizes de incidência que relacionam as observações aos vetores \mathbf{b} de efeitos fixos, \mathbf{u} de efeito aleatório genético aditivo direto, \mathbf{m} de efeito aleatório genético materno e \mathbf{pe} de efeito de ambiente permanente materno, e \mathbf{e} é o vetor dos efeitos aleatórios residuais.

Como efeitos fixos, foram considerados os efeitos de grupos contemporâneos (GC – mesma fazenda, sexo, ano e estação de nascimento, grupo de manejo e data da desmama ou sobreano) e classes de idade da vaca, os efeitos linear e quadrático da idade do animal na mensuração e os efeitos lineares aditivos e maternos de proporção de Nelore na composição racial e heterozigose individual e materna. Para as características medidas ao sobreano (PS550 e GPD345) não foram considerados os efeitos genéticos maternos e de ambiente permanente materno, efeitos maternos de proporção de Nelore e heterozigose materna. Foram excluídos GC com menos de cinco animais e medidas com 3,5 desvios padrão acima ou abaixo da média do GC ao qual pertenciam. A estrutura geral do arquivo de dados analisado está descrita na Tabela 7.

7.2.3 Estudo de associação genômica ampla (GWAS) bayesiano

Foi utilizada a abordagem proposta por GARRICK et al. (2009) para calcular os valores genéticos desregredidos (DEBV – deregression breeding values) e correspondentes ponderadores (\mathbf{w}) = $\{w_j\}$ para todos os animais, livre do efeito médio dos pais a partir dos valores genéticos e respectivas confiabilidades. Estes “pseudo fenótipos” ou DEBV foram então analisados em um modelo aleatório com efeitos de

substituição alélica de SNP utilizando o software GenSel versão 4.0 (FERNANDO & GARRICK, 2009).

Os métodos bayesianos BayesA e BayesB (MEUWISSEN et al., 2001) foram utilizados para analisar os dados de DEBV e genótipo simultaneamente. Os métodos BayesA e BayesB, consideram que cada SNP tem uma variância específica para cada locus, com uma distribuição de qui-quadrado invertida escalonada $X^{-2}(v, S)$ com graus de liberdade v e escala S determinados no GenSel por default (padrão). No entanto, a especificação para os efeitos de SNP em BayesB assume que uma proporção de marcadores tenha efeito nulo, com probabilidade fixa π , enquanto que os marcadores restantes têm efeitos normalmente distribuídos com variância de locus específica com probabilidade $1-\pi$. Por outro lado, no método BayesA, todos os SNP estão ajustados com $\pi = 0$ no modelo, para cada ciclo de Monte Carlo via Cadeias de Markov (MCMC). O seguinte modelo estatístico foi utilizado para as duas metodologias:

$$\mathbf{y} = \sum_{i=1}^{k=41,045} \delta_i \mathbf{z}_i \mathbf{a}_i + \mathbf{e}, \quad [2]$$

em que: \mathbf{y} é o vetor de fenótipos (DEBV); $K=41.045$ é o total de marcadores SNPs no painel; δ_i é o indicador de que o SNP i foi incluído ($\delta_i = 1$) ou excluído ($\delta_i = 0$) do modelo para uma dada iteração de MCMC; \mathbf{z}_i é o vetor de genótipos ajustado para cada i marcador, codificado como -10/0/10; \mathbf{a}_i é o vetor de efeito aleatório de substituição do marcador ajustado i com sua própria variância $\sigma_{a_i}^2$ e um efeito *a priori* nulo com probabilidade π ou não-nulo com probabilidade $1 - \pi$; e \mathbf{e} é o vetor aleatório residual com distribuição normal. Para o método BayesA, δ_i foi assumido ser igual a 1. Para BayesB o parâmetro π foi assumido ser igual a 0,995. Um alto valor de π foi escolhido para permitir que apenas as regiões com associações mais fortes sejam identificadas (ZARE et al., 2014). Um tamanho de cadeia de 42.000 amostras foi utilizado, em que as primeiras 2.000 foram descartadas como burn-in. A convergência de MCMC foi verificada pelo teste de GEWEKE (1991) usando o pacote “boa” do software R (SMITH, 1992).

Com base nas análises estatísticas e nos resultados dos componentes variância obtidos com a equação [1], estes foram utilizados como *prioris* para rodar as análises de GWAS bayesiano no Gensel.

Os animais genotipados para as características estudadas foram divididos em quatro ou cinco grupos (dependendo do número de animais para cada característica),

por duas estratégias de validação cruzada, usando o programa R (R CORE TEAM, 2013): agrupamento em função do parentesco genômico, pelo método K-médias, e de forma aleatória, de acordo com SAATCHI et al. (2011) com a finalidade de formar grupos de treinamento e validação independentes. Para selecionar subconjuntos de SNPs e testar a predição genômica, os GWAS foram executados usando o método de BayesB ($\pi = 0,995$) considerando todos os animais e também com os grupos separadamente.

7.2.4 Seleção de Janelas e SNPs (Tag SNPs)

Devido ao LD entre marcadores próximos de um QTL, o efeito de um QTL pode ser distribuído sobre marcadores vizinhos. Deste modo, os SNPs estavam distribuídos em 2.519 janelas não sobrepostas de 1 Mb contendo número variável de SNPs, usando um mapa físico ordenado de acordo com a montagem do genoma bovino UMD3.1 (ZIMIN et al., 2009). A variância genética explicada conjuntamente por cada subconjunto de SNP, considerada como a variância de janela, foi estimada e posteriormente convertida para a proporção da variância genética total explicada pela janela (SUN et al., 2011; WOLC et al., 2012).

As regiões do genoma que potencialmente contêm loci de características quantitativas (QTL) associados as características de crescimento, denominadas como janelas mais representativas (janelas top), foram identificadas a partir de um limiar estabelecido com base na contribuição da variância genética como descrito por SCHURINK et al., (2012). As janelas top foram identificadas no GWAS considerando todos os animais com genótipo e fenótipo para característica estudada, e também para cada um dos grupos separadamente. Assumindo contribuição igual de todas as regiões genômicas, a proporção esperada de variância genética explicada por cada uma das 2.519 janelas é de 0,04%. Consequentemente, as janelas de 1 Mb explicando pelo menos 0,2% da variância genética e, portanto, cinco vezes mais do que o esperado ($0,04\% * 5 = 0,2\%$), foram consideradas como regiões potencialmente associadas a QTL (ONTERU et al., 2013; ZARE et al., 2014) e selecionadas para análises posteriores.

Para identificar SNPs para construção de painéis de baixa densidade, foi testada uma estratégia para seleção de SNPs (Tag SNPs) nas janelas mais representativas, considerando os seguintes parâmetros: *model frequency* (MF), *t-like*

(TL), desequilíbrio de ligação (LD) e frequência do alelo menor (MAF). No GenSel, o parâmetro MF reflete a proporção de cadeias que incluíram determinado marcador no modelo, enquanto que o TL reflete a razão do efeito médio *a posteriori* sobre o desvio padrão das distribuições destes efeitos, considerando somente das cadeias que incluíram o efeito no modelo. O pacote R / snpStats (CLAYTON, 2012) foi utilizado para obter os valores de LD.

A estratégia de seleção de SNP foi a seguinte: 1) Identificação do maior valor de MF dentro de cada janela (MFtop); 2) Identificação do menor valor dentre os MFtop (minMFtop); 3) 1ª seleção - SNPs com MF > minMFtop; este passo teve como objetivo selecionar SNPs que não estavam no topo de suas próprias janelas, mas que tinham valor de MF suficientemente grande para exceder o valor de MFtop dos SNPs localizados em outras janelas selecionadas. Uma abordagem semelhante foi utilizada para avaliar a consistência do efeito do SNP considerando TL; 4) Identificação do SNP com menor TL dentre os selecionados pela 1ª seleção (TLmin) 5) 2ª seleção - SNPs com TL > TLmin; 6) Identificação de DL ($r^2 > 0,4$; BADKE et al., 2012) entre os SNPs selecionados e 7) 3ª seleção - nos SNPs em DL somente é selecionado o SNP com maior MAF.

7.2.5 Acurácia de predição dos painéis de SNPs selecionados

Para verificar a eficácia de escolher apenas os SNPs mais informativos para as características de crescimento, o método BayesA foi aplicado a cada um dos painéis SNP propostos numa estratégia de validação cruzada para cada grupo de forma independente, assumindo que todos os top SNPs possuem efeitos não nulos. Para os indivíduos dentro de um certo grupo de validação, os valores genômicos diretos (DGV) foram calculados com base nos SNPs selecionados e os correspondentes efeitos de substituição alélica, estimados a partir do conjunto de treinamento, consistindo em fenótipos e genótipos dos outros grupos, sendo que os dados dos animais do conjunto de validação não foram usado no treinamento. Para o *j*-ésimo indivíduo,

$$\widehat{DGV}_j = \sum_{i=1}^K z_{ji} \hat{a}_i,$$

em que os efeitos SNP estimados, \hat{a}_i , foram representados pelas suas médias *a posteriori* obtidos utilizando o método BayesA, e z_{ji} representa o genótipo para o *i*-ésimo SNP do total de marcadores K incluídos no painel de baixa densidade.

As acurácias de predição combinadas através de todos os grupos foram derivadas como a correlação genética entre os fenótipos observados (\mathbf{y}) e os valores genômicos diretos (DGV) para uma determinada característica, estimada em um modelo animal bicaracterística, utilizando uma matriz do numerador de parentesco (\mathbf{A}^*), em que as covariâncias entre indivíduos de diferentes grupos foram zeradas, conforme SAATCHI et al., (2013):

$$\mathbf{A}^* = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{A}_{33} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{A}_{44} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{A}_{55} \end{bmatrix}$$

Os componentes de (co)variâncias e correlações genéticas foram estimados por meio de inferência bayesiana com base no método de Monte Carlo via Cadeias de Markov, utilizando 100.000 ciclos e período de descarte inicial de 50.000 amostras usando o software Gibbs2f90 (MISZTAL et al., 2002).

Dentro de um dado grupo (g), as acurácias preditas foram alternativamente estimadas como a correlação ($\hat{r}_{a\hat{a}g}$) entre o predito ($\hat{\mathbf{a}}$) e o verdadeiro (\mathbf{a}) valor genético, como proposto por LEGARRA et al. (2008):

$$\hat{r}_{a\hat{a}g} = r(\mathbf{y}_g^*, \hat{\mathbf{y}}_g^*) / h$$

em que: \mathbf{y}_g^* é o vetor dos fenótipos dos animais pertencentes ao grupo g ajustados para os efeitos fixos como na equação [1] e $\hat{\mathbf{y}}_g^*$ são os DGV a partir da validação cruzada e h é a raiz quadrada da herdabilidade da característica.

Para comparação da acurácia de predição do DGV usando o painel de baixa densidade, esta foi também estimada para os mesmos agrupamentos de validação cruzada (k-médias e aleatório) com os métodos BayesB ($\pi=0,995$) e BayesA considerando todos os 41.045 marcadores.

7.2.6 Análise de Enriquecimento Funcional

A base de dados Medical Subject Headings (MeSH) foi utilizada para definir conjuntos funcionais de genes. O MeSH é um conjunto de descritores usados para artigos indexados no banco de dados Medline (COLETTI & BLEICH, 2001; NELSON et al., 2004). Esses descritores ou títulos representam tópicos-chave discutidos nos artigos. Em contraste com Gene Ontology (GO), onde as categorias funcionais são mapeadas para genes, os termos MeSH são atribuídos à literatura. Uma vez que os termos MeSH são atribuídos a publicações PubMed, um passo adicional é necessário para associar MeSH à genes (NAKAZATO et al., 2009, JANI et al., 2010). É importante observar que os genes atribuídos à mesma categoria funcional (ou seja, um determinado descritor MeSH) podem assim ser considerados membros de uma categoria de genes que estão mais intimamente relacionados em termos de algum aspecto biológico do que conjuntos aleatórios de genes.

A associação de um determinado descritor MeSH para as características de crescimento foram analisadas usando o teste exato de Fisher, o qual é um teste de proporções com base na distribuição hipergeométrica cumulativa (GAMBRA et al., 2013). Este teste foi realizado para pesquisa de uma super-representação de genes significativos em uma dada categoria funcional, entre todos os genes. Em particular, o *valor P* de g genes significativos observados foi calculado da seguinte forma:

$$P - value = 1 - \sum_{i=0}^{g-1} \frac{\binom{S}{i} \binom{N-S}{K-1}}{\binom{N}{K}}$$

Em que, S é o número total de genes que foram considerados significativamente associados com as características estudadas, N é o número total de genes analisados no estudo e K é o número total de genes na categoria funcional em consideração.

A análise de enriquecimento MeSH foi realizada usando o pacote “*MESHR*” (MOROTA et al., 2015, TSUYUZAKI et al., 2015), disponível no ambiente R (R CORE TEAM, 2013).

7.3 RESULTADOS E DISCUSSÃO

Na tabela 8, estão demonstrados os componentes de (co)variâncias estimados pelo pedigree para as características de crescimento.

Em relação aos métodos de agrupamento, o k-médias produziu grupos com número de animais desbalanceados para as características medidas a desmama e ao sobreano (já demonstrados nas Tabela 9 e 10), e também a média de parentesco genômico (G_{ij}) dentro de um grupo foram maiores do que a média entre grupos. Para o agrupamento de forma aleatória, os grupos tiveram número de animais balanceado, e não houve diferença nos valores de G_{ij} dentro e entre grupos, sendo estes próximos de zero (dados não mostrados). O objetivo foi formar grupos distintos para representar cenários com animais de parentesco mais distante e outro, com parentesco mais próximo, para as populações de treinamento e validação.

7.3.1 Estudo de associação genômica ampla (GWAS)

Foi utilizado para o método BayesB, incluindo todos os animais e marcadores, o parâmetro π de 0,995, o que corresponde a 0,5% dos SNPs são ajustados no modelo a cada iteração. As herdabilidades (h^2) estimadas para cada característica com a metodologia bayesiana foram de 0,22, 0,14, 0,28 e 0,16, para PN, PD205, PS550 e GPD345, respectivamente. Essas h^2 foram de menor proporção em relação a estimada pelo pedigree (Tabela 2) para PN e PD205 e de maior proporção para PS550 e GPD345. PETERS et al., (2012) analisando dados de bovinos da raça Brangus e utilizando o método BayesC ($\pi=0,999$) encontraram valores de h^2 menores do que o presente estudo, de 0,13, 0,04 e 0,19, para peso ao nascer, peso a desmama ajustado para os 205 dias e para peso ao ano ajustado para os 365 dias. De acordo com FERNANDO & GARRICK (2013) valores mais elevados de π podem às vezes ser mais discriminantes na identificação de maior QTL e, portanto, influenciar na proporção de variância genética explicada pelos marcadores.

A proporção da variância genética explicada pelas 2.519 janelas de 1Mb utilizando todos os marcadores (41.045 SNPs) ao longo do genoma para PN, PD205, PS550 e GPD345 estão representadas nas figuras 12, 13, 14 e 15, respectivamente.

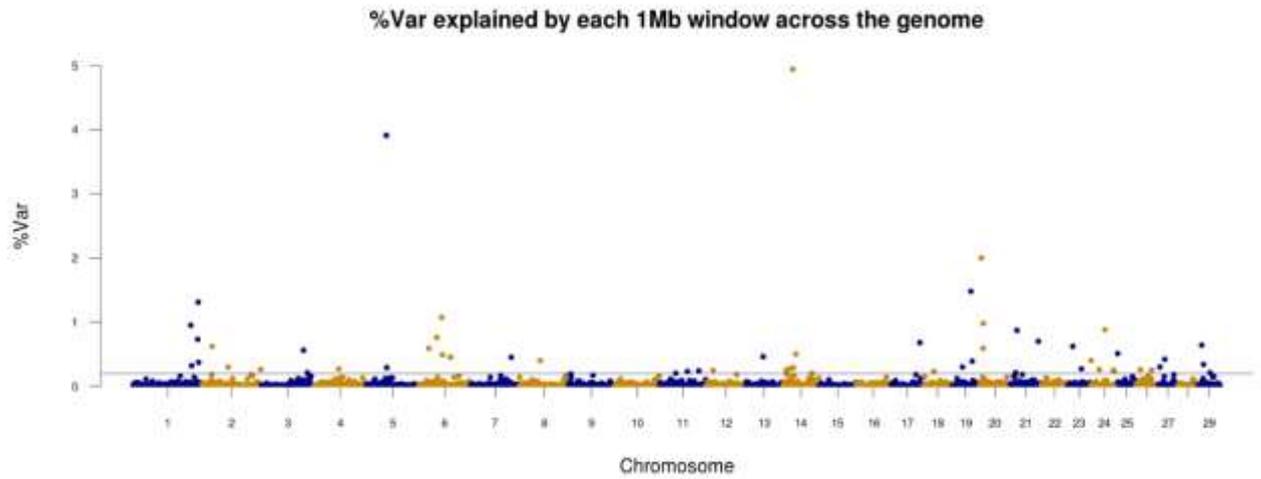


Figura 12. Gráfico de manhattan plot mostrando os resultados para o estudo de associação genômica ampla realizado pelo método Bayes B ($\pi=0.99$) para peso ao nascer. O eixo Y representa a proporção da variância genética explicada por janelas de 1Mb e o eixo X representa os cromossomos que as janelas estão localizadas.

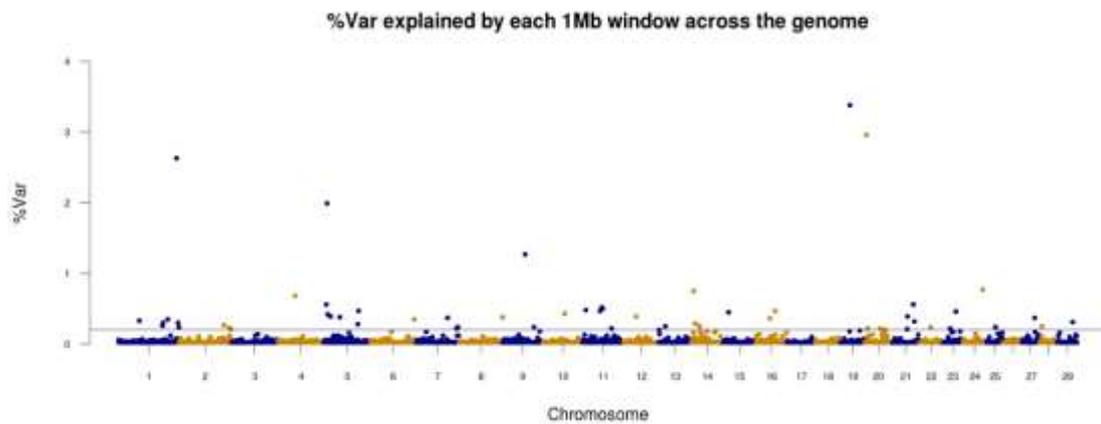


Figura 13. Gráfico de manhattan plot mostrando os resultados para o estudo de associação genômica ampla realizado pelo método Bayes B ($\pi=0.995$) para peso a desmama ajustado para os 205 dias. O eixo Y representa a proporção da variância genética explicada por janelas de 1Mb e o eixo X representa os cromossomos que as janelas estão localizadas.

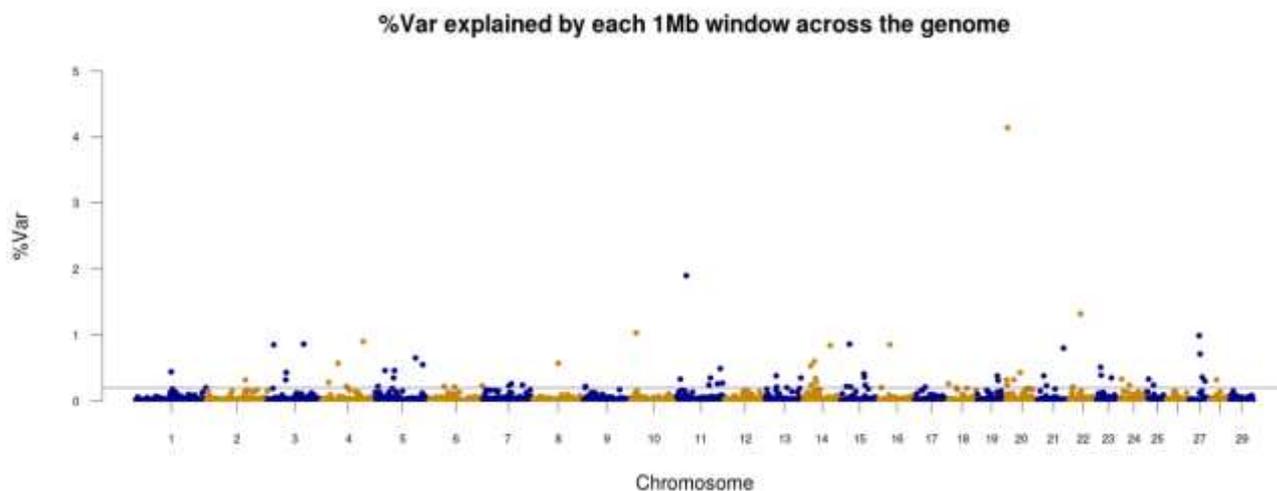


Figura 14. Gráfico de manhattan plot mostrando os resultados para o estudo de associação genômica ampla realizado pelo método Bayes B ($\pi=0.995$) para peso ao sobreano ajustado para os 550 dias. O eixo Y representa a proporção da variância genética explicada por janelas de 1Mb e o eixo X representa os cromossomos que as janelas estão localizadas.

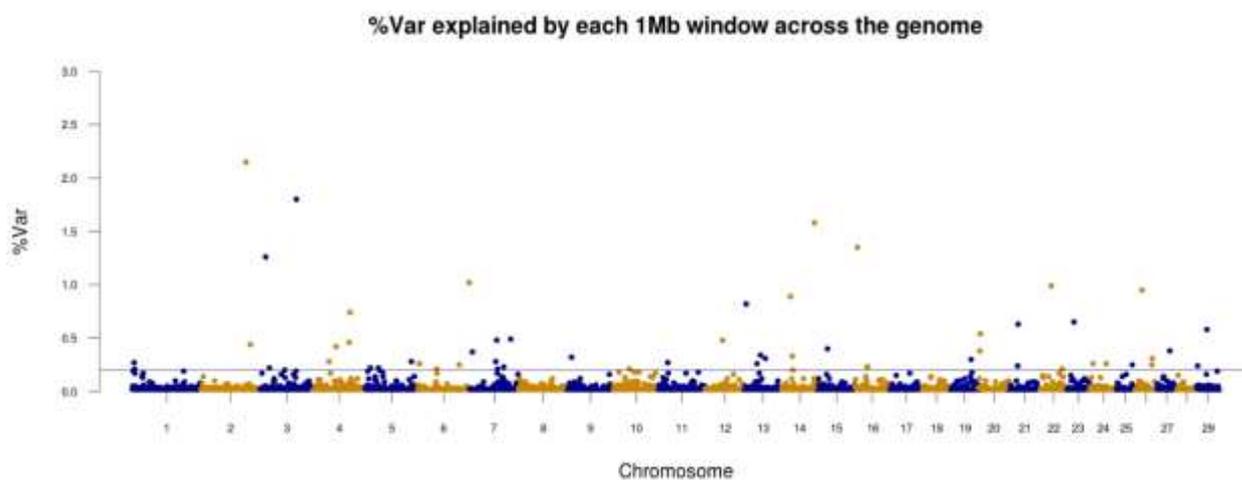


Figura 15. Gráfico de manhattan plot mostrando os resultados para o estudo de associação genômica ampla realizado pelo método Bayes B ($\pi=0.995$) para ganho de peso pós-desmama ajustado para os 345 dias. O eixo Y representa a proporção da variância genética explicada por janelas de 1Mb e o eixo X representa os cromossomos que as janelas estão localizadas.

Com GWAS usando o método BayesB ($\pi=0,995$), foi possível identificar as janelas mais representativas para as características estudadas, nas quais estas, foram consideradas contendo QTL. As janelas mais representativas foram selecionadas utilizando o limiar de 0,2% da variância genética explicada por cada janela.

Para PN, um total de 60 janelas representadas por um total de 1047 SNPs explicaram 36,66% da variância genética (Anexo 1). As Janelas mais representativas

podem ser observadas nos BTA14, 5, 20, 19, 1 e 6, representando aproximadamente 15% da variância genética, sendo que essas explicaram individualmente mais do que 1% da variância aditiva para PN. A janela que mais explicou foi a localizada no BTA14 (5%), indicando ser este um cromossomo de extrema importância e onde possivelmente estejam presentes os genes que exercem grande efeito sobre a característica. O número de SNPs em cada janela variou de 7 a 28 para PN. UTSONOMYIA et al., (2013) também encontraram SNPs no BTA14 que explicavam 4,62% da variância genética, analisando dados de peso ao nascer de bovinos Nelore. WENG et al., (2016) estudando animais da raça Brangus, relataram SNPs explicando mais do que 0,5% de variância aditiva nos BTA6, 5 e 20, semelhante as regiões encontradas no presente estudo, porém, também encontraram SNPs importantes no BTA7.

Para PD205, 57 janelas representadas por 1060 SNPs explicaram 30% da variância aditiva (Anexo 2). As janelas mais importantes foram observadas nos BTA19, 20, 1 e 5, responsáveis por aproximadamente 10% da variância aditiva para a característica. O número de SNPs para PD205 em cada janela variou de 5 até 28 marcadores. WENG et al., (2016) encontraram janelas mais representativas nos BTA7, 6, 20, 12, 5, 14 e 10 e PETERS et al., (2012) nas regiões do BTA3, BTA6, BTA10, BTA16 e BTA29, analisando dados de novilhas Brangus. No presente estudo, também foram encontrados SNPs representativos nos cromossomos BTA6, 7, 12, 10, 14 e 16, porém, a variância genética explicada pelas janelas presentes nessas regiões não foi muito expressiva.

Para as características medidas ao sobreano (Anexos 3 e 4), 74 janelas e 1369 SNPs foram responsáveis por explicar aproximadamente 37% da variância genética para PS550. As janelas mais representativas foram encontradas no BTA20, BTA11, BTA22, BTA10 e BTA27, responsáveis por explicar quase 10% da variância aditiva para PS550. Para a característica GPD345, um total de 58 janelas representadas por 1008 SNPs, explicaram quase 29% da variância aditiva. As janelas mais significativas foram encontradas nos BTA2, 3, 14, 16, 6 e 22. O número de SNPs para essas características variaram de 7 a 28 em cada janela. BUZANKAS et al., (2014) encontraram SNPs associados com peso ao sobreano ajustado para os 420 dias de idade em bovinos da raça Canchim nos BTA7, 22, 25 e 27. PETERS et al., (2012) encontraram janelas mais significativas nos BTA6, BTA7, BTA8 e BTA9 para peso ao

ano ajustado para os 365 dias de idade e nos BTA5, 6, 7, 8, 9 e 10 para ganho médio diário do nascimento aos 365 dias de idade em novilhas raça Brangus.

É importante destacar que em estudos de GWAS é comum a identificação de regiões cromossômicas associadas com a característica de interesse, que não tenham sido anteriormente citadas na literatura (TIZIOTO et al., 2013). TIZIOTO et al., (2013) ainda destaca que diferenças entre as frequências alélicas entre bovinos taurinos e zebuínos ou a extensão do desequilíbrio de ligação entre SNPs e as variantes causais devem resultar em diferentes efeitos de marcadores detectados em diferentes raças. O efeito de substituição alélica calculado de um SNP variará de um conjunto de dados para outro, enquanto que as associações podem ser significativas ou não, conforme o caso e também a quantidade de variância explicada pelo SNP pode diferir substancialmente, particularmente quando as raças ou tipos de bovinos utilizados em um estudo diferem das raças em outro (BOLORMAA et al., 2011).

7.3.2 Seleção de marcadores moleculares (Tag SNPs)

Após selecionar as janelas mais representativas, os SNPs foram também selecionados com base nos parâmetros MF e TL fornecidos pelo Gensel.

Dentre as janelas mais representativas para PN, no primeiro passo da seleção, utilizando o valor mínimo de MF de 0,044, foram selecionados 92 SNPs e, no segundo, através do valor mínimo de TL de 0,88, 150 SNPs. Do total de 150 SNPs que passaram pela seleção pelos parâmetros de MF e TL, 47 foram excluídos por estarem em DL com outro SNP selecionado e apresentarem menor valor de MAF, o que resultou em um painel de 103 SNPs selecionados como mais representativos para a característica de PN. Entre os SNPs selecionados, maior valor de MF foi observado ser de 0,6164 para rs109590980/ ARS-BFGL-NGS-73392 localizado no BTA5 dentro da janela de 1 Mb explicando grande parte da variância genética, o que indica que este particular SNP não tem efeito nulo em 62% das MCMC.

Para PD205, a seleção de SNPs utilizando o limiar de MF mínimo de 0,049, resultou em 75 SNPs pertencentes as janelas selecionadas anteriormente. Após a segunda seleção pelo valor mínimo de TL de 0,894, restaram um total de 97 SNPs, sendo que destes, 18 foram excluídos por estarem em DL e apresentarem menor valor de MAF, com isso 79 SNPs permaneceram para PD205. O SNP rs110133873/ARS-

BFGL-NGS-4904 localizado no BTA19 foi o que apresentou maior valor de MF de 0,7123.

A seleção de SNPs mais representativos para PS550 resultou em 112 selecionados pelo limiar de MF de 0,0583 e 120 SNPs por TL de 0,9170. Do total de 120 SNPs selecionados pelos dois parâmetros, 20 foram excluídos por apresentarem DL e menor MAF, o que resultou em um painel de 100 marcadores mais representativos para PS550. O SNP rs110245673/ARS-BFGL-BAC-11047, pertencente ao BTA11, apresentou maior valor de MF de 0,5452 entre os selecionados.

A seleção de SNPs para GPD345 com base nas janelas pré-selecionadas, resultou em 80 marcadores, com valor de MF de 0,0405 utilizado como limite mínimo para seleção de SNPs para esse parâmetro, e de 0,866 para TL originando 122 SNPs. Em virtude de apresentarem DL e menores valores de MAF, 23 SNPs foram excluídos, assim um total de 99 SNPs compôs o painel de marcadores para GPD345. O maior valor observado de MF entre os SNPs selecionados foi 0,6028 para rs41567611/Hapmap50215-BTA-111911 localizado no BTA2.

Os valores médios de MF para cada característica foram 0,12, 0,13, 0,17 e 0,11 para PN, PD205, PS550 e GPD, respectivamente, indicando que se existe LD entre SNPs pode que em cada ciclo um SNP diferente seja selecionado para representar o mesmo gene dentro de uma janela. Geralmente, os SNPs com MF mais alto nas janelas superiores estão potencialmente associados com o fenótipo em estudo (ZARE et al., (2014).

A correlação (r) entre MF e TL, utilizada para distinguir marcadores informativos com estimativas consistentes versus os de estimativas inconsistentes (FERNANDO & GARRICK, 2009), foi de magnitude moderada para todas as características, de 0,41, 0,43, 0,40 e 0,40 para PN, PD205, PS550 e GPD345, respectivamente.

Portanto, a finalidade da estratégia, foi selecionar os SNPs mais informativos ligados a QTLs relacionados com as características de crescimento em bovinos das raças Hereford e Braford. Dentro das janelas superiores, foram priorizados aqueles SNPs que foram incluídos no modelo mais frequentemente (maiores MF) e com efeitos mais consistentes (maiores TL), porém, evitando a redundância através do LD. Deste modo, o objetivo foi reter SNPs com capacidade de predição adequada para construir painéis de baixa densidade para a seleção genômica de características de crescimento com custo efetivo. Entre o total de SNPs, 381 para todas as

características, apenas 23 foram redundantes, ou seja, o mesmo SNP estava em diferentes painéis. De acordo com HABIÉR et al., (2009) uma desvantagem importante das abordagens de seleção de marcadores, é que provavelmente diferentes SNPs irão ser selecionados para distintas características, de modo que um painel de baixa densidade (LowDensity - LD-SNP) desenvolvido para uma característica, possivelmente não funcione tão bem para outra. Se cada característica econômica requer seu próprio painel LD-SNP, os custos para desenvolvimento e genotipagem não irão ser menores do que para o painel de média densidade, como por exemplo de 30K.

A mesma estratégia de seleção de SNPs foi aplicada aos grupos formados para validação cruzada. A partir dos métodos de agrupamento, foram realizados 8 ou 10 GWAS dependendo do número de grupos para cada característica, resultando em pequenas listas de SNPs (sub-listas) (Tabela 19).

7.3.3 Acurácia de predição dos painéis de SNPs selecionados

As herdabilidades *a posteriori* estimadas (h^2) com BayesA (Tabela 4) usando a lista de marcadores selecionados para cada grupo (k-médias e aleatório), foram maiores do que as h^2 utilizando o BayesB ($\pi=0,995$) com todos os 41.045 marcadores (Tabela 19). O número de SNPs selecionados variou entre todas as características estudadas, de 41 até 108 marcadores em cada grupo.

Várias abordagens, dentre elas as Bayesianas, têm sido recentemente propostas para predição de valores genéticos genômicos utilizando marcadores SNPs com painéis de alta densidade. Porém, esta estratégia pode não ser desejável na prática, pelo alto custo de genotipar todos os animais com estes painéis, tornando-se uma alternativa interessante, a utilização de painéis de baixa densidade (CLEVELAND et al., 2010).

A acurácia foi estimada através das correlações genéticas entre o fenótipo ajustado e o DGV obtidas em uma análise bicaracterística a partir da validação cruzada com o BayesA utilizando os SNPs selecionados (BayesA_TagSNPs), e utilizando todos os marcadores (41.045) também com o BayesA e BayesB ($\pi=0,995$) (Tabela 20). As acurácias foram maiores para todas as características para os grupos formados aleatoriamente em relação ao k-médias, indicando que quanto maior a distância da população geneticamente menor será a acurácia.

Tabela 19. Grupos de animais de acordo com a estratégia de validação cruzada, K-médias (cl) ou aleatório (rand), média a posteriori da proporção de variância explicada pelos marcadores (h^2) considerando o método BayesB ($\pi=0,995$) e BayesA, o número de janelas top (TopW), número de SNPs em cada janela (TopSNPs) e o número de SNPs selecionados (TagSNPs).

Caracteres ²	Parâmetros ¹	Grupos									
		cl1	cl2	cl3	cl4	cl5	rand1	rand2	rand3	rand4	rand5
PN	h^2_{BB995}	0,26	0,22	0,22	0,23	0,21	0,21	0,23	0,23	0,24	0,21
	h^2_{BATopW}	0,27	0,28	0,29	0,26	0,31	0,29	0,29	0,29	0,29	0,28
	Top_W	41	57	57	48	52	60	54	52	47	53
	TopSNPs	753	1057	1067	915	982	1110	985	945	844	1002
	TagSNPs	59	85	72	79	112	91	86	81	75	81
PD205	h^2_{BB995}	0,17	0,16	0,16	0,15	0,15	0,15	0,16	0,16	0,16	0,14
	h^2_{BATopW}	0,24	0,24	0,24	0,23	0,22	0,25	0,24	0,23	0,23	0,21
	TopW	46	46	41	54	52	54	53	46	41	42
	TopSNPs	810	895	762	984	980	995	1001	864	780	754
	TagSNPs	59	69	59	78	69	95	72	62	65	72
PS550	h^2_{BB995}	0,22	0,24	0,22	0,23	-	0,23	0,22	0,24	0,22	-
	h^2_{BATopW}	0,29	0,32	0,34	0,30	-	0,34	0,32	0,34	0,30	-
	TopW	57	54	55	42	-	60	55	56	58	-
	TopSNPs	1030	955	948	742	-	1045	1014	951	1068	-
	TagSNPs	87	84	88	67	-	108	73	80	86	-
GPD345	h^2_{BB995}	0,13	0,11	0,11	0,12	-	0,11	0,14	0,10	0,12	-
	h^2_{BATopW}	0,22	0,21	0,18	0,18	-	0,19	0,18	0,16	0,19	-
	TopW	48	40	31	38	-	39	33	36	42	-
	TopSNPs	891	725	538	661	-	739	575	647	758	-
	TagSNPs	83	91	55	53	-	67	44	49	48	-

¹TopW: representa o número de janelas que explicam acima de 0,2% da variância genética nas análises de GWAS usando BayesB ($\pi=0,995$) em cada grupo; TopSNPs: representa o número de SNPs incluídos em cada janela top (TopW); TagSNPs: representa os SNPs selecionados de acordo com os parâmetros de *model frequency* (MF) e *t-like* (TL), desequilíbrio de ligação (DL) e frequência do alelo menor (MAF).

²PN: peso ao nascer; PD205: peso a desmama ajustado para os 205 dias de idade; PS550: peso ao sobreano ajustado para os 550 dias de idade; GPD345: ganho pós-desmama ajustado para os 345 dias de idade.

Para as características medidas a desmama, utilizando BayesA TagSNPs o maior valor correlação genética 0,30 para PN com o agrupamento k-médias e de 0,65 com o aleatório para a mesma característica (Tabela 20). Para as características medidas ao sobreano, o maior valor foi para PS550, de 0,21 com os grupos formados pelo k-médias e 0,50 com o aleatório. Para todas as características foram encontrados maiores valores de acurácia utilizando todos os marcadores (41.045) do que selecionando os SNPs mais representativos. É esperado que a acurácia de predição

diminua a medida a densidade do painel é mais baixa, devido ao menor desequilíbrio de ligação esperado entre marcadores e QTLs altamente dispersos (HABIER et al., 2009) que afetam a característica de interesse. No entanto, alguns estudos têm demonstrado a superioridade dos métodos bayesianos para capturar este LD entre marcadores e QTL (HABIER et al., 2007; ZHONG et al., 2009).

Tabela 20. Correlações genéticas do fenótipo ajustado com o valor genômico direto (DGV) predito da validação cruzada usando diferentes métodos para características de crescimento medidas a desmama e ao sobreano.

Característica ¹	Grupo	Metodologia ²		
		BayesA_TagSNP	BayesA	BayesB
PN	K-médias	0,30±0,08	0,38±0,09	0,36±0,09
	Aleatório	0,65±0,20	0,77±0,17	0,74±0,16
PD205	K-médias	0,20±0,06	0,26±0,05	0,25±0,05
	Aleatório	0,30±0,08	0,47±0,13	0,42±0,10
PS550	K-médias	0,21±0,07	0,33±0,10	0,30±0,09
	Aleatório	0,50±0,23	0,71±0,17	0,67±0,15
GPD345	K-médias	0,13±0,09	0,23±0,11	0,20±0,11
	Aleatório	0,36±0,24	0,50±0,22	0,46±0,23

¹PN: peso ao nascer; PD205: peso a desmama ajustado para os 205 dias de idade; PS550: peso ao sobreano ajustado para os 550 dias de idade; GPD345: ganho pós-desmama ajustado para os 345 dias de idade.

²BayesA_TagSNP: modelo bayesiano com distribuição-t e probabilidade ($\pi=0$), usando somente os marcadores mais informativos; BayesA: modelo bayesiano com distribuição-t e probabilidade ($\pi=0$) utilizando todos os marcadores (41.045); BayesB: modelo de misture bayesiano com distribuição-t e ponto de massa em zero e probabilidade ($\pi=0,995$), utilizando todos os marcadores (41.045).

As acurácias dentro de grupos também foram estimadas, com base na equação [3], como proposto por LEGARRA et al., (2008) e para todas as características, foram mais baixas para o grupo composto na sua maior parte por animais da raça Hereford (Figuras 16, 17 e 18), pelo fato de ter menor número de animais genotipados para esta raça. Estes resultados já eram esperados, pois CARDOSO et al. (2015), trabalhando com o mesmo conjunto de dados, mas para a característica de resistência ao carrapato, utilizando a mesma estratégia para formação dos grupos, também encontraram menores acurácias para o grupo onde continham somente animais da raça Hereford, porém, utilizando as metodologias BayesB e BayesC com $\pi=0,95$.

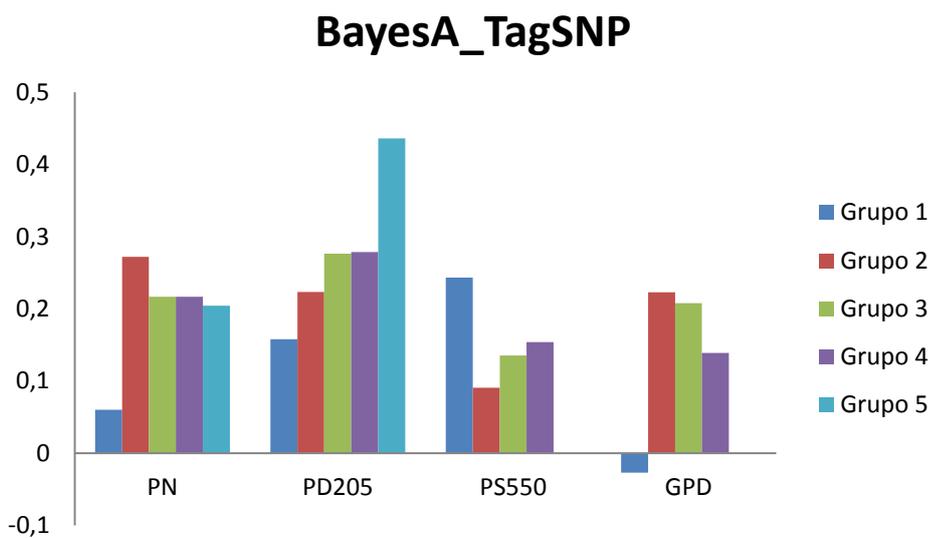


Figura 16. Acurácia dos valores genômicos diretos (DGV) preditos a partir da validação cruzada usando agrupamento k-médias com o método BayesA_TagSNP.

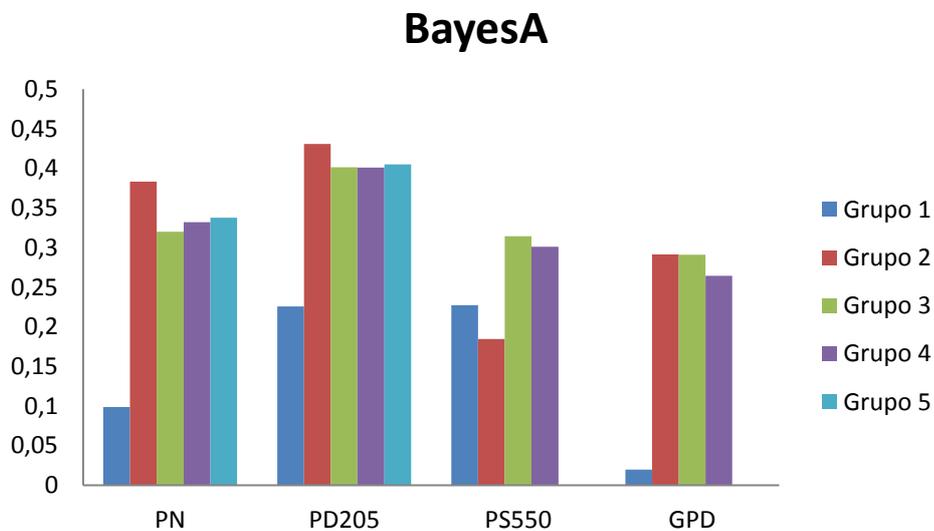


Figura 17. Acurácia dos valores genômicos diretos (DGV) preditos a partir da validação cruzada usando agrupamento k-médias com o método BayesA.

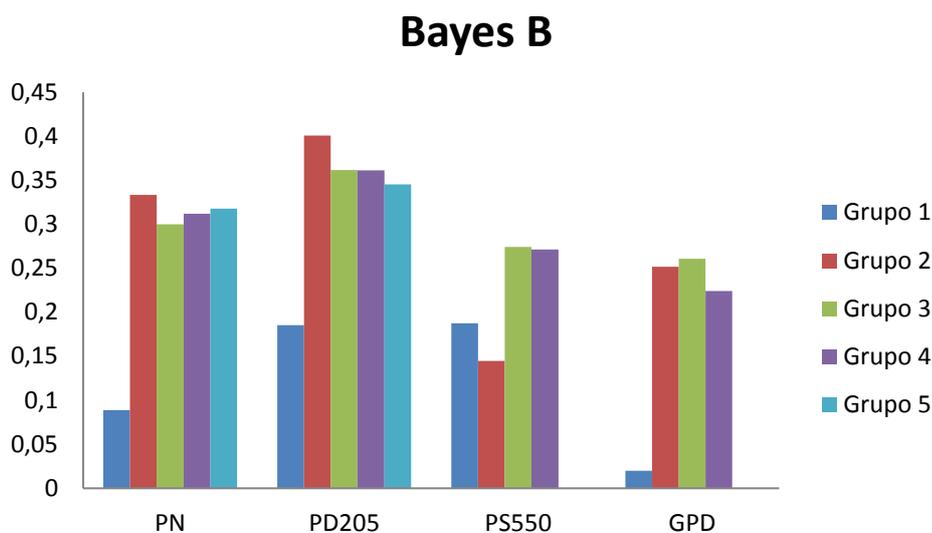


Figura 18. Acurácia dos valores genômicos diretos (DGV) preditos a partir da validação cruzada usando agrupamento k-médias com o método BayesB.

Comparando os resultados usando o painel de baixa densidade proposto através da seleção de SNPs, para PN as acurácias irão corresponder a pelo menos 79% se comparado com BayesA utilizando os 41.045 SNPs. Utilizando o BayesA Tag SNPs para PD205, a acurácia corresponde a 77% quando comparada ao BayesA utilizando painel de alta densidade. Para as características de sobreano, as acurácias usando o painel de baixa densidade proposto para cada característica, correspondem a 63% e 58% para PS550 e GPD345 quando comparado com painel utilizando a metodologia BayesA com todos os marcadores (41.045). Os resultados encontrados no presente estudo demonstram a possibilidade da seleção de marcadores SNPs mais informativos e ainda de se encontrar um painel de baixa densidade que possa reduzir o custo das predições genômicas em comparação com painéis de alta densidade.

As metodologias bayesianas geralmente usam procedimentos de encolhimento para considerar variações diferentes para SNPs individuais, nas quais a informação *a priori* sobre a distribuição de efeitos SNP é usada para forçar efeitos insignificantes para zero. No entanto, BayesA foi escolhido porque espera-se que painéis Tag SNP incluam apenas os marcadores mais significativos com efeitos detectáveis, permitindo que SNPs em cada cadeia tenham variâncias específicas (MEUWISSEN et al., 2001).

A aplicação com sucesso da seleção genômica no melhoramento genético de bovinos, tem como objetivo minimizar os custos de genotipagem e simultaneamente maximizar os ganhos genéticos e os lucros. Os dados de genótipos gerados a

menores custos, provenientes de pequenos subconjuntos de SNP altamente informativos, poderiam ser usados para genotipar a maioria dos animais em um rebanho e gerar previsões de valor genético genômico com base nos efeitos de SNP estimados a partir de dados de treinamento com painéis de alta densidade (HABIER et al., 2009) e ou se completar esse painel através da imputação dos genótipos (DASSONNEVILLE et al., 2011).

Além disso, para a implementação em nível nacional ou mundial da seleção genômica em bovinos de corte, é necessário o estabelecimento de estratégias de menor custo e desenvolvimento de populações de treinamento robustas para características de importância econômica. Deste modo, os painéis de baixa densidade podem representar uma alternativa viável para os rebanhos comerciais de bovinos de corte procurando obter ganhos genéticos e maior lucratividade com o uso da seleção genômica.

7.3.4 Enriquecimento Funcional

Nove termos MeSH classificados na categoria fenômenos e processos mostraram significativa super-representação de genes estatisticamente associados com PN (Tabela 21) e destes, os mais importantes foram encontrados nas categorias de componentes corporais e vazio gástrico. De forma geral, foram identificados como genes candidatos para PN, o SIRT2 (*Sirtium 2*), que está envolvido na regulação do crescimento e na qualidade da carne dos bovinos (GUI et al., 2014), os genes GHSR (*growth hormone secretagogue receptor*), localizado no BTA1, que está envolvido com secreção do hormônio do crescimento, regulação do apetite e outras funções importantes (SHIN et al. 2010; ALBARRAN-ZECKLER et al. 2011). O gene IGF1 (*insulin like growth factor 1*), no BTA19, também está relacionado a características de crescimento (GRASSI et al., 2015). O gene para Tiroglobina (TG) produz o precursor do hormônio da tireóide e foi proposto como um gene candidato para QTL com efeito na deposição de gordura, e, portanto, envolvido no crescimento, composição de carcaça e qualidade de carne (ZHANG et al., 2015). O gene FAS (*Fas cell surface death receptor*), no BTA26, classificado na categoria espermatogênese, de acordo com MEGGIOLARO et al., (2006) está ligado a subfertilidade ou infertilidade em bovinos.

O SNP no BTA14, gene mapeado para MOS (*V-Mos Moloney murine sarcoma viral oncogene homolog* (MOS), que explicou 4,94% da variância genética para PN, está relacionado com a estatura em bovinos e humanos (GUDBJARTSSON et al., 2008; PRYCE et al., 2011). UTSUNOMIYA et al., (2013), demonstraram que a mesma região de QTL no BTA14 afetam o peso ao nascer e o tamanho em bovinos Nelore (*Bos primigenius indicus*).

Tabela 21. Os termos MeSH significativamente sobre representados com genes estatisticamente associados com a característica de peso ao nascer (PN).

MeSH Term ID	MeSH Term Name	Nº Total Genes	Nº Genes significativos	P-value ¹
D013091	Spermatogenesis	25	3	0.011
D018900	Dinucleotide Repeats	10	2	0.014
D042002	Chromatin Assembly and Disassembly	10	2	0.014
D005746	Gastric Emptying	1	1	0.019
D005496	Follicular Atresia	12	2	0.02
D002886	Chromosomes, Human, Pair 17	13	2	0.024
D001823	Body Composition	37	3	0.031
D003241	Consanguinity	2	1	0.037
D012143	Respiratory Physiological Phenomena	2	1	0.037

¹Nível de significância (P<0,05).

Para a característica PD205, 10 termos MeSH foram classificados como associados com a característica estudada (Tabela 22). Da lista, apenas as categorias de regeneração e peso ao nascer foram identificados genes que estão envolvidos diretamente com o crescimento de bovinos de corte. O gene MSTN (*myostatin*), localizado no cromossomo 2, está associado com aumento da massa do músculo esquelético em humanos e outros mamíferos (IWASAKI et al., 2013; GAO et al., 2014). Os genes MYF5 e MYF6 no BTA 5, estão também envolvidos com aumento da massa do músculo esquelético (ROBAKOWSKA-HYZOREK et al., 2010). Alguns genes foram identificados estarem associados com o peso ao nascer e crescimento em *Bos Taurus*, como SLC44A5 (*solute carrier family 44 member 5*), GH1 (*growth hormone 1*), IGF1 (*insulin like growth factor 1*) e IGFBP3 (*insulin like growth factor binding protein 3*).

Tabela 22. Os termos MeSH significativamente sobre representados com genes estatisticamente associados com a característica de peso a desmama ajustado para os 205 dias (PD205).

MeSH Term ID	MeSH Term Name	Nº Total Genes	Nº Genes significativos	P-value ¹
D016385	TATA Box	6	2	0.0012
D012038	Regeneration	11	2	0.0043
D004357	Drug Synergism	15	2	0.0081
D013318	Stroke Volume	1	1	0.0092
D053140	Organelle Shape	1	1	0.0092
D012091	Repetitive Sequences, Nucleic Acid	53	3	0.013
D016897	Respiratory Burst	2	1	0.018
D000375	Aging	32	2	0.035
D001724	Birth Weight	4	1	0.036
D058507	Host Specificity	4	1	0.036

¹Nível de significância (P<0,05).

Para PS550, foram identificados 15 termos MeSH influenciando a característica estudada (Tabela 23). Apenas o gene *STC2* (*stanniocalcin 2*), foi identificado estar associado com características de crescimento em bovinos de corte. Este gene está envolvido com a deposição de gordura intramuscular e, portanto, influenciando as características de carcaça e qualidade de carne (RAMAYO-CALDAS et al., 2014). Foi identificado o gene *BSP1* (*binder of sperm 1*) ligado a fertilidade de touros (SOMASHEKAR et al., 2015). No entanto, foram identificados também genes ligados a características de importância econômica em bovinos de leite, como *LTF* (*lactotransferrin*) envolvido com a mastite, e o *LALBA* (*lactalbumin alpha*) e *PRL* (*prolactin*) com a produção de leite.

Tabela 23. Os termos MeSH significativamente sobre representados com genes estatisticamente associados com a característica de peso ao sobreano ajustado para os 550 dias (PS550).

MeSH Term ID	MeSH Term Name	Nº Total Genes	Nº Genes significativos	P-value ¹
D000327	Adsorption	8	2	0.006
D018919	Neovascularization, Physiologic	55	4	0.0096
D001794	Blood Pressure	11	2	0.011
D013318	Stroke Volume	1	1	0.015
D014664	Vasodilation	18	2	0.03
D002872	Chromosome Deletion	2	1	0.03
D005720	Gamma Rays	2	1	0.03
D009006	Monosomy	2	1	0.03
D016391	Genes, src	2	1	0.03
D016897	Respiratory Burst	2	1	0.03
D040681	Structural Homology, Protein	20	2	0.036
D007408	Intestinal Absorption	3	1	0.045
D017344	Genes, Insect	3	1	0.045
D020239	Receptor Cross-Talk	3	1	0.045
D006706	Homeostasis	23	2	0.047

¹Nível de significância (P<0,05).

Para GPD345 foram identificados 42 termos MeSH (Tabela 24) relacionados a característica de crescimento. Dentre a lista de genes, alguns foram identificados influenciar diretamente o crescimento de bovinos, como IGF1 no BTA5, MSTN no BTA2, LEP no BTA4, que já foram descritos para as outras características. Outros genes, ligados ao crescimento como o GAD1 (*glutamate decarboxylase 1*) no BTA2 e PPARGC1A (*PPARG coactivator 1 alpha*) no BTA6, que de acordo com LI et al., (2010) são genes candidatos que afetam características de crescimento em bovinos de corte. Os genes CSRP1 (*cysteine and glycine rich protein 1*), LPL (*lipoprotein lipase*), no BTA6 e FBX032 (*F-box protein 32*) (WANG et al., 2012; HE et al., 2013), estão associados com características de crescimento em bovinos de corte.

Tabela 24. Os termos MeSH significativamente sobre representados com genes estatisticamente associados com a característica de ganho pós-desmama ajustado para os 345 dias (GPD345).

MeSH Term ID	MeSH Term Name	Nº Total Genes	Nº Genes significativos	P-value ¹
D020871	RNA Stability	10	3	0.00043
D002472	Cell Transformation, Viral	5	2	0.0024
D016062	Porosity	5	2	0.0024
D024363	Transcription Initiation Site	7	2	0.0049
D019076	Transgenes	8	2	0.0065
D018389	Codon, Nonsense	9	2	0.0083
D005822	Genetic Vectors	53	4	0.0097
D012091	Repetitive Sequences, Nucleic Acid	53	4	0.0097
D020413	3' Untranslated Regions	53	4	0.0097
D025461	Feedback, Physiological	10	2	0.01
D004058	Diffusion	11	2	0.012
D020128	Inhibitory Concentration 50	11	2	0.012
D058767	Protein Unfolding	11	2	0.012
D013379	Substrate Specificity	196	8	0.013
D017434	Protein Structure, Tertiary	428	13	0.019
D000465	Algorithms	1399	32	0.022
D006579	Heterozygote	39	3	0.023
D010641	Phenotype	104	5	0.025
D014018	Tissue Distribution	270	9	0.028
D004905	Erythrocyte Aging	2	1	0.031
D005342	Fibrinolysis	2	1	0.031
D006790	Host-Parasite Interactions	2	1	0.031
D015334	Gene Rearrangement, gamma-Chain T-Cell Antigen Receptor	2	1	0.031
D016474	Weight-Bearing	2	1	0.031
D055261	Immunogenetic Phenomena	2	1	0.031
D010766	Phosphorylation	328	10	0.036
D020836	Protein Structure, Quaternary	79	4	0.036
D013499	Surface Properties	20	2	0.039
D000107	Acetylation	21	2	0.043
D017510	Protein Folding	121	5	0.043
D017433	Protein Structure, Secondary	249	8	0.044
D000705	Anaphase	3	1	0.047
D002468	Cell Physiological Phenomena	3	1	0.047
D004198	Disease Susceptibility	3	1	0.047
D007135	Immunoglobulin Variable Region	3	1	0.047
D018911	Trinucleotide Repeats	3	1	0.047
D011110	Polymorphism, Genetic	167	6	0.049

¹Nível de significância (P<0,05).

7.4 CONCLUSÕES

O método BayesB demonstrou ser adequado para a seleção de SNPs com base nos parâmetros *model frequency* (MF) e *t-like* (TL). Os painéis de baixa densidade propostos incluíram SNPs ligados a genes funcionais e que explicaram parte da variação genética para as características de crescimento.

As predições genômicas obtidas com os painéis de baixa densidade propostos usando BayesA apresentaram uma predição moderada para as características medidas a desmama e um pouco mais baixas que estas para as características medidas na fase de sobreano. Estes painéis podem ser úteis para estudos futuros relacionados ao mapeamento fino para a descoberta de variantes causais.

8. CONCLUSÃO GERAL

A estimação de parâmetros genéticos para os escores visuais com os modelos linear e de limiar foram muito semelhantes quando as características apresentaram distribuição normal dos dados, como para conformação, precocidade, musculatura e tamanho, indicando que pode ser usado qualquer um dos modelos na avaliação genética para esses escores. Por outro lado, para o escore de umbigo, em que não apresenta distribuição normal dos dados, o modelo de limiar se mostrou mais vantajoso na estimação da herdabilidade e também na análise de validação dos modelos.

A predição de valores genético utilizando informações dos marcadores moleculares pelos métodos GBLUP, BayesB e BayesC possibilitou a obtenção de valores com acurácia igual ou maior que o BLUP tradicional. Maiores valores foram obtidos com os métodos bayesianos. Um maior número de animais genotipados poderá incrementar a acurácia para todas as características estudadas. As acurácias de predição foram menores para os grupos que continham somente animais da raça Hereford (menor número de animais genotipados), demonstrando a necessidade de se genotipar um maior número de animais dessa raça. A utilização das metodologias que combinaram todas as fontes de informação, pedigree, fenótipo e genômica, e animais genotipados e não genotipados, permitiram obter altos ganhos em acurácia, com destaque para o método de passo único GBLUP (ssGBLUP) e indicando que deve ser usada como ferramenta para melhorar os ganhos genéticos.

O estudo de associação genômica ampla utilizando metodologia bayesiana permitiu a seleção de marcadores que representaram maior proporção da variância genética e também através dos parâmetros, modelo de frequência (MF), estatística do tipo t (TL), desequilíbrio de ligação (DL) e frequência do alelo menor (MAF). A seleção dos SNPs pelos parâmetros citados possibilitou o desenvolvimento de painéis de baixa densidade, porém, como já esperado os ganhos genéticos foram menores do que utilizando informação de todos os marcadores (painel de alta densidade). Maiores valores de acurácia foram obtidos para as características que tinham um maior número de animais com genótipos e fenótipos. Portanto, outros estudos devem ser realizados com um maior número de animais para propor o painel de baixa densidade em nível comercial. O uso destes painéis utilizando a metodologia de passo

único pode ser uma alternativa para implementação da seleção genômica de forma mais econômica e obtenção de ganhos genéticos consideráveis.

9. REFERÊNCIAS BIBLIOGRÁFICAS

- ABHB, **Raças Hereford e Braford**, 2017. <http://www.abhb.com.br> (Acessado em 06 de Março de 2017).
- ABIEC, **Relatório anual, perfil da pecuária no Brasil**, 2016. <http://abiec.siteoficial.ws/images/upload/sumario-pt-010217.pdf>. (Acessado em 20 de dezembro de 2016).
- ABO-SMAIL, M.K.; VOORT, G.V.; SQUIRES, J.J.; SWANSON, K.C.; MANDELL, I.B.; LIAO, X.; STOTHARD, P.; MOORE, S.; PLASTOW, G.; MILLER, S.P. Single nucleotide polymorphisms for feed efficiency and performance in crossbred beef cattle. **BMC Genetics**, v.15, n.14, 2014.
- ALBARRAN-ZECKLER, R. G.; SUN, Y.; SMITH, R. G. Physiological roles revealed by ghrelin and ghrelin receptor deficient mice. **Peptides**, v. 32, n. 11, p. 2229-2235, 2011.
- AGUILAR, I.; MISZTAL, I.; JOHNSON, D. L.; LEGARRA, A.; TSURUTA, S.; LAWLOR, T. J. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. **Journal of Dairy Science**, v. 93, p. 743–752, 2010.
- AGUILAR, I.; MISZTAL, I. 2014. PreGSF90. <http://nce.ads.uga.edu/wiki/doku.php?id=readme.seekparentf90>. (Acessado em 30 de Junho de 2014).
- ASBIA, **Relatório anual de importação, exportação e comercialização de sêmen**, 2015. <http://www.asbia.org.br/novo/home/>. (Acessado em 15 de dezembro de 2016).
- BADKE, Y.M.; BATES, R.O.; ERNST, C.W.; SCHWAB, C.; STEIBEL, J.P. **Estimation of linkage disequilibrium in four US pig breeds**. *BMC Genomics*, v. 13, n. 1, p. 24, 2012.
- BIF. **Guidelines for uniform improvement programs**. BEEF IMPROVEMENT FEDERATION, Ninth edition, 2010.
- BIGNARDI, A. B.; GORDO, D.G.M.; ALBUQUERQUE, L.G.; SESANAI, J.C. Parâmetros genéticos de escore visual do umbigo em bovinos da raça Nelore. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, v. 63, p. 941-947, 2011.
- BODDHIREDDY, P.; KELLY, M. J.; NORTHCUTT, S.; PRAYAGA, K. C.; RUMPH, J.; DENISE, S. Genomic predictions in Angus cattle: Comparisons of sample size, response variables, and clustering methods for cross-validation. **Journal of Animal Science**, v.92, p. 485–497, 2014.
- BOLDMAN, K. G.; VAN VLECK, L. D.; GREGORY, K. E.; CUNDIFF, L. V. Estimates of direct and maternal parameters for 200 d weight in purebred and composite lines of beef cattle [abstract]. **Journal of Animal Science**, v. 69, (Suppl 1), p. 203, 1991.

BOLIGON, A. A.; BALDI, F.; ALBUQUERQUE, L. G. D. Genetic parameters and relationships between growth traits and scrotal circumference measured at different ages in Nelore cattle. **Genetics and molecular biology**, v. 34, n. 2, p. 225-230, 2011.

BOLIGON, A. A.; MERCADANTE, M. E. Z.; ALBUQUERQUE, L. G. D. Genetic associations of conformation, finishing precocity and muscling visual scores with mature weight in Nelore cattle. **Livestock Science**, v. 135, n. 2, p. 238-243, 2011.

BOLIGON, A. A.; AYRES, D.R.; PEREIRA, R.J.; MOROTTI, N.P.; ALBUQUERQUE, L.G. Genetic associations of visual scores with subsequent rebreeding and days to first calving in Nelore cattle. **Journal of Animal Breeding and Genetics**, v. 129, n. 6, p. 448-456, 2012.

BOLIGON, A. A.; CARVALHEIRO, R.; ALBUQUERQUE, L. G. Evaluation of mature cow weight: Genetic correlations with traits used in selection indices, correlated responses, and genetic trends in Nelore cattle. **Journal of animal science**, v. 91, n. 1, p. 20-28, 2013.

BOLIGON, A. A., DE VARGAS, L., SILVEIRA, D. D., ROSO, V. M., CAMPOS, G. S., VAZ, R. Z.; SOUZA, F. R. P. Genetic models for breed quality and navel development scores and its associations with growth traits in beef cattle. **Tropical animal health and production**, v. 48, n. 8, p. 1679-1684, 2016.

BOLORMAA, S.; NETO, L. R.; ZHANG, Y. D.; BUNCH, R. J.; HARRISON, B. E.; GODDARD, M. E.; BARENDSE, W. A genome-wide association study of meat and carcass traits in Australian cattle. **Journal of Animal Science**, v. 89, n. 8, p. 2297-2309, 2011.

BOLORMAA, S.; PRYCE, J.E.; KEMPER, K.E.; HAYES, B.J.; ZHANG, Y.; TIER, B. BARENDSE, W.; REVERTER, A.; GODDARD, M.E. Detection of quantitative trait loci in *Bos indicus* and *Bos taurus* cattle using genome-wide association studies. **Genetics Selection Evolution**, v. 45, n.43, 2013.

BOURDON, R. M. Bovine nirvana-from the perspective of a modeler and purebred breeder. **Journal of Animal Science**, v. 66, n. 8, p. 1892-1898, 1988.

BOVINE GENOME SEQUENCING AND ANALYSIS CONSORTIUM. The genome sequence of taurine cattle: a window to ruminant biology and evolution. **Science**, v.324, n. 5926, p. 522-528, 2009.

BUZANSKAS, M. E.; GROSSI, D. A.; VENTURA, R. V.; SCHENKEL, F. S.; SARGOLZAEI, M.; MEIRELLES, S. L.; MOKRY, F.B.; HIGA, R. H.; MADADU, M. A.; SILVA, M. V. G. B. D.; NICIURA, S. C.; TORRES JÚNIOR, R. A. A.; ALENCAR, M. A.; REGITANO, L. C. A.; MUNARI, D. P. Genome-wide association for growth traits in canchim beef cattle. **PloS one**, v. 9, n. 4, p. e94802, 2014.

CAETANO, A.R. Marcadores SNP: conceitos básicos, aplicações no manejo e no melhoramento animal e perspectivas para o futuro. **Revista Brasileira de Zootecnia**, v. 38, n. 8, p. 64-71, 2009 (supl. especial).

CAETANO, A. R.; ARAÚJO, R. O. DE; BIEGELMEYER, P.; CARDOSO, F. F. Ferramentas genômicas aplicadas no melhoramento genético animal: progressos recentes e perspectivas futuras. In: REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE ZOOTECNIA, 49., 2012, Brasília. A produção animal no mundo em transformação: anais. Brasília, DF: SBZ, 2012. 1 CD-ROM.

CALUS, M. P. L. Genomic breeding value prediction: methods and procedures. **Animal**, v. 4, n. 2, p 157-164, 2010.

CAMPOS, L. T. PROMEBO® – Programa de Melhoramento de Bovinos de Carne. In: Curso de Melhoramento da Raça Aberdeen Angus – Formação de Avaliadores PROMEBO/ANGUS, 5, 2011, Porto Alegre. p.2-15, 2011.

CARDOSO, F. F.; CARDELLINO, R. A.; CAMPOS, L. T. Componentes de (co) variância e parâmetros genéticos para caracteres produtivos à desmama de bezerros Angus criados no estado do Rio Grande do Sul. **Revista Brasileira de Zootecnia**, v. 30(1), p. 41-48, 2001.

CARDOSO, F. F.; CARDELLINO, R. A.; CAMPOS, L. T. Componentes de (co) variância e parâmetros genéticos de caracteres pós-desmama em bovinos da raça Angus. **Revista Brasileira de Zootecnia**, v. 33, n. 2, p. 313-319, 2004.

CARDOSO, F. F.; TEMPELMAN, R. J. Hierarchical Bayes multiple-breed inference with an application to genetic evaluation of a Nelore-Hereford population. **Journal of Animal Science**, v. 82, p. 1589-1601, 2004.

CARDOSO, F.F; GULIAS GOMES, C. C.; OLIVEIRA, M.M. et al. Predição da resistência genética ao carrapato de bovinos Braford e Hereford a partir de um painel denso de marcadores moleculares. Embrapa Pecuária Sul, Circular Técnica 41, 2011.

CARDOSO, F. F.; LOPA, T. B. P.; BIEGELMEYER, P.; CAETANO, A. R. Avanços nos programas de avaliação genética aplicados à pecuária de corte. In: CANOZZI, M. E. A.; BREMM, B.; COSTA JÚNIOR, J. B. G.; BARCELLOS, J. O. J. (Org.). VIII Jornada NESPRO. Porto Alegre: UFRGS, p. 179-196, 2013.

CARDOSO, F. F.; GOMES, C. C. G.; SOLLERO, B. P.; OLIVEIRA, M. M.; ROSO, V. M.; PICCOLI, M. L.; HIGA, R. H.; YOKOO, M. J.; CAETANO, A. R.; AGUILAR, I. Genomic prediction for tick resistance in Braford and Hereford cattle. **Journal of Animal Science**, v. 93, n. 6, p. 2693-2705, 2015.

CDCB, **Council on dairy cattle breeding**, 2017. <https://www.uscdcb.com> (Acessado em 06 de Março de 2017).

CHEN, C. Y.; MISZTAL, I.; AGUILAR, I.; TSURUTA, S.; AGGREY, S. E.; WING, T.; MUIR, W. M. Genome-wide marker-assisted selection combining all pedigree phenotypic information with genotypic data in one step: an example using broiler chickens. **Journal of Animal Science**, v. 89, n. 1, p. 23-28, 2011.

CHRISTENSEN, O.F.; MOGENS S.L. Genomic prediction when some animals are not genotyped. **Genetics Selection Evolution**, v. 42, n. 1, p. 1-8, 2010.

CLAYTON, D. 2014. snpStats: SnpMatrix and XSnpmatrix classes and methods. R package version 1.18.0.

<http://www.bioconductor.org/packages/release/bioc/html/snpStats.html>

(Acessado em 30 de Novembro de 2015).

CLARK, S. A.; HICKEY, J. M.; WERF, JULIUS H. J. Different models of genetic variation and their effect on genomic evaluation. **Genetics Selection Evolution**, v. 43, n.18, 2011.

CLEVELAND, M. A.; FORNI, S.; DEEB, N.; MALTECCA, C. Genomic breeding value prediction using three Bayesian methods and application to reduced density marker panels. In: **BMC proceedings**. BioMed Central, p. S6, 2010.

COLETTI, M.H.; BLEICH, H.L. Medical subject headings used to search the biomedical literature. **Journal of the American Medical Informatics Association**, v. 8, p. 317–23, 2001.

CUNDIFF, L. V.; GREGORY, K. E.; LONG, C. R. Genetic variation among and within herds of Angus and Hereford cattle. **Journal Animal Science**, v. 41, p. 1270-1280, 1975.

DASSONNEVILLE, R.; BRONDUM, R. F.; DRUET, T.; FRITZ, S.; GUILLAUME, F.; GULDBRANDTSEN, B.; LUND, M. S.; DUCROCQ, V.; SU, G. Effect of imputing markers from a low-density chip on the reliability of genomic breeding values in Holstein populations. **Journal of Dairy Science**, v. 94, p. 3679–3686, 2011.

DAETWYLER, H. D. et al. The impact of genetic architecture on genome-wide evaluation methods. **Genetics**, v. 185, p. 1021–1031, 2010.

DE MATTOS, D.; MISZTAL, I.; BERTRAND, J. K. Variance and covariance components for weaning weight for Herefords in three countries. **Journal Animal Science**, v. 78, p. 33-37, 2000.

DAETWYLER, H. D.; PONG-WONG, R.; VILLANUEVA, B.; WOOLLIAMS, J. A. The impact of genetic architecture on genome-wide evaluation methods. **Genetics**, v. 185, p. 1021–1031, 2010.

DE LOS CAMPOS, G.; NAYA, H.; GIANOLA, D.; CROSSA, J.; LEGARRA, A.; MANFREDI, E.; WEIGEL, K; COTES, J. M. Predicting quantitative traits with regression models for dense molecular markers and pedigree. **Genetics**, v. 182, n. 5, p. 375-385, 2009.

DE LOS CAMPOS, G. HICKEY, J. M. PONG-WONG, R. DAETWYLER, H. D. CALUS, M. P. L. Whole-Genome Regression and Prediction Methods Applied to Plant and Animal Breeding. **Genetics**, v. 193, p. 327–345, 2013.

DE MATTOS, D.; MISZTAL, I.; BERTRAND, J. K. Variance and covariance components for weaning weight for Herefords in three countries. **Journal Animal Science**, v. 78, p. 33-37, 2000.

DE ROOS A. P. W.; HAYES, B. J.; SPELMAN, R.; GODDARD, M. E. Linkage disequilibrium and persistence of phase in Holstein Friesian, Jersey and Angus cattle. **Genetics**, v. 179, n. 3, p. 1503-1512, 2008.

DE ROOS, A. P. W.; HAYES, B. J.; SPELMAN, R.; GODDARD, M. E. Reliability of genomic predictions across multiple populations. **Genetics**, v. 183, p. 1545-1553, 2009.

DEKKERS, J. C. M.; HOSPITAL, F. The use of molecular genetics in improvement of agricultural populations. **Nature Reviews Genetics**, v. 3, n. 1, p. 22-32, 2002.

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

EGGEN, A. The development and application of genomic selection as a new breeding paradigm. **Animal frontiers**, v. 2, n. 1, p. 10-15, 2012.

EVERLING, D. M.; RORATO, P. R. N.; ARAUJO, R. O. D.; BOLIGON, A. A.; BRESOLIN, T.; DORNELLES, M. D. A.; WEBER, T.; SANTANA, P. P.; CAMPOS, L. T. Associação genética de escores de conformação com características de crescimento em bovinos da raça Angus. **Pesquisa Agropecuária Brasileira**, v. 47, n. 10, p. 1524-1532, 2012.

EVERLING, D. M.; BRESOLIN, T.; RORATO, P. R. N.; ARAÚJO, R. O.; BOLIGON, A. A.; WEBER, T.; DORNELLES, M. A.; Campos, L. T. Finishing precocity visual score and genetic associations with growth traits in Angus beef cattle. **Genetics and Molecular Research**, v. 13, n. 3, p. 7757-7765, 2014.

FAN B.; ONTERU, S. K.; DU, Z. Q.; GARRICK, D. J.; STALDER K. J.; ROTHSCHILD, M. F. Genome wide association study identifies loci for body composition and structural soundness traits in pigs. **PLoS One**, v. 6, n. 2, p. e14726, 2011.

FARIA, C. U. D.; MAGNABOSCO, C. U.; ALBUQUERQUE, L. G.; DE LOS REYES, A.; BEZERRA, L. A. F.; LOBO, R. B. Análise genética de escores de avaliação visual de bovinos com modelos bayesianos de limiar e linear. **Pesquisa Agropecuária Brasileira**, v. 43, n. 7, p. 835-841, 2008.

FARIA, C.U. DE; KOURY FILHO, W.; MAGNABOSCO, C.U.; ALBUQUERQUE, L.G. de; BEZERRA, L.A.F.; LÔBO, R.B. Bayesian inference in genetic parameter estimation of visual scores in Nellore beef-cattle. **Genetics and Molecular Biology**, v.32, p.753-760, 2009.

FERNANDO, R. L.; GARRICK, D. J. 2009. GenSel: User manual for a portfolio of genomic selection related analyses.
<http://www.biomedcentral.com/content/supplementary/1471-2105-12-186-S1.pdf>.
(Acessado em 5 de Agosto de 2014).

FERNANDO, R. L.; GARRICK D. J. **GenSel-user manual**. Mapp genes complex Trait Domest Anim their use Breed Program, 2009.

FERNANDO, R. L.; GARRICK, D. J. Bayesian methods applied to GWAS. In: **Genome-Wide Association Studies and Genomic Prediction**. Springer; p. 237–274, 2013.

FERNANDO, R.L.; DEKKERS, J.C.M.; GARRICK, D.J. A class of Bayesian methods to combine large numbers of genotyped and non-genotyped animals for whole-genome analyses. **Genetics Selection Evolution**, v. 46, p. 50, 2014.

FORNI, S.; AGUILAR, I.; MISZTAL, I. Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. **Genetics Selection Evolution**, v.43, n.1, p.1-7, 2011.

GAMBRA, R.; PEÑAGARICANO, F.; KROPP, J.; KHATEEB, K.; WEIGEL, K.; LUCEY, J.; KHATIB, H. Genomic architecture of bovine j-casein and b-lactoglobulin. **Journal of Dairy Science**, v. 96, p. 5333–43, 2013.

GAO, F.; SUN, B.; XING, S.; YU, X.; LU, C.; LI, A.; YANG, R. The effect of leader peptide mutations on the biological function of bovine myostatin gene. **Gene**, v. 540, n. 2, p. 171-177, 2014.

GARRICK, D. J.; TAYLOR, J.F; FERNANDO, R.L. Deregressing estimated breeding values and weighting information for genomic regression analyses. **Genetics Selection Evolution**, v. 41, p. 55, 2009.

GARRICK, D. J. The nature, scope and impact of some whole-genome analyses in beef cattle. In: **Proceedings of the 9th World Congress on Genetics Applied to Livestock Production**: 1-6 August 2010; Leipzig (p. 0023).

GELFAND, A.E.; SMITH, A.F.M. Sampling-based approaches to calculating marginal densities. **Journal of the American Statistics Association**, v. 85, p. 398–409, 1990.

GEWEKE, J. Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments. In: Bernardo, J.M., Berger, J.O., Dawid, A.P., Smith, A.F.M. (Eds.), **Bayesian Statistics 4**. Oxford University Press, New York, p. 625–631, 1992.

GIANOLA, D. Theory and analysis of threshold characters. **Journal of Animal Sciences**, v. 54, n. 5, p. 1079-1096, 1982.

GIANOLA, D.; FOULLEY, J.L. Sire evaluation for ordered categorical data with a threshold model. **Genetics Selection Evolution**, v. 15, p. 201–224, 1983.

GIANOLA D.; SORENSEN, D. Likelihood, bayesian, and MCMC methods in quantitative genetics. 2.ed. **Statistics for Biology and Health**: Springer, p. 740, 2002.

GIANOLA, D.; PEREZ-ENCISO, M.; TORO, M.A. On Marker-Assisted Prediction of Genetic Value: Beyond the Ridge. **Genetics**, v.163, n.1, p.347-365, 2003.

GIANOLA, D.; DE LOS CAMPOS, G.; HILL, W. G.; MANFREDI, E.; FERNANDO, R. Additive genetic variability and the Bayesian alphabet. **Genetics**, v. 183, n. 1, p. 347-363, 2009.

GODDARD, M.E.; HAYES, B.J. Genomic selection. **Journal of Animal Breeding and Genetics**, v.124, p.331-341, 2007.

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

GODDARD, M. Genomic selection: prediction of accuracy and maximisation of long term response. **Genetica**, v. 136, n. 2, p. 245-257, 2009.

GORDO, D. G. M.; BALDI, F.; LÔBO, R. B.; SAINZ, R. D.; ALBUQUERQUE, L. G. Genetic association between body composition measured by ultrasound and visual scores in Brazilian Nelore cattle. **Journal of Animal Science**, v. 90, n. 12, p. 4223-4229, 2012.

GORDO, D. G. M.; ESPIGOLAN, R.; TONUSSI, R. L.; JÚNIOR, G. A. F.; BRESOLIN, T., MAGALHÃES, A. F.; Feitosa, F. L.; BALDI, F.; CARVALHEIRO, R.; TONHATI, H.; OLIVEIRA, H. N.; CHARDULO, L. A. L.; ALBUQUERQUE, L. G. Genetic parameter estimates for carcass traits and visual scores including or not genomic information. **Journal of Animal Science**, v. 94, n. 5, p. 1821-1826, 2016.

GROSSI, D. D. A.; BUZANSKAS, M. E.; GRUPIONI, N. V.; DE PAZ, C. C. P.; DE ALMEIDA REGITANO, L. C.; DE ALENCAR, M. M.; MUNARI, D. P. Effect of IGF1, GH, and PIT1 markers on the genetic parameters of growth and reproduction traits in Canchim cattle. **Molecular biology reports**, v. 42, n. 1, p. 245-251, 2015.

GROSSO, J. L. B. M.; OLIVEIRA Jr, G. A.; MENEZES, I. R.; MATTOS, E. C.; FERRAZ, J. B. S.; ELER, J. P. Genetic parameters for body conformation scores and carcass traits measured by real-time ultrasound in Nelore cattle. In: **Proceedings 10th World Congress on Genetics Applied to Livestock Production**, Vancouver, British Columbia, Canada, p. 718, 2014.

GUDBJARTSSON, D. F.; WALTERS, G. B.; THORLEIFSSON, G.; STEFANSSON, H.; HALLDORSSON, B. V.; ZUSMANOVICH, P. et al. Many sequence variants affecting diversity of adult human height. **Nature genetics**, v. 40, n. 5, p. 609-615, 2008.

GUI, L. S.; YANG, W. C.; ZHAO, C. P.; WEI, S. J.; ZHAO, Z. D.; ZAN, L. S. Association of SIRT2 gene polymorphisms with body measurement and growth traits of Qinchuan cattle. **Genetics and Molecular Research**, v. 13, n. 4, p. 8834-8844, 2014.

HABIER D.; FERNANDO, R. L.; DEKKERS, J. C. M. The impact of genetic relationship information on genome-assisted breeding values. **Genetics**, v. 177, p. 2389-2397, 2007.

HABIER, D.; FERNANDO, R. L.; DEKKERS, J. C. M. Genomic selection using low-density marker panels. **Genetics**, v. 182, p. 343–353, 2009.

HABIER, D.; FERNANDO, R. L.; KIZILKAYA, K.; GARRICK, D. J. Extension of the Bayesian alphabet for genomic selection. **BMC Bioinformatics**, v. 12, p. 186, 2011.

HAYES, B.J.; BOWMAN, P.J.; CHAMBERLAIN, A.J.; GODDARD, M. E. Invited review: Genomic selection in dairy cattle: Progress and challenges. **Journal of Dairy Science**, v. 92, n. 3, p. 433–443, 2009.

HE, H.; LIU, X. L.; ZHANG, H. L.; YANG, J.; NIU, F. B.; LI, Z. X.; CHEN, L. SNV and haplotype analysis reveals new CSRP1 variants associated with growth and carcass traits. **Gene**, v. 522, n. 2, p. 206-213, 2013.

HEFFNER, E. L.; JANNINK, J.; SORRELLS, M. E. 2011. Genomic Selection Accuracy using Multifamily Prediction Models in a Wheat Breeding Program. **Plant Genome**, v. 4, p. 65-75, 2011.

HENDERSON, C.R. Best linear unbiased estimation and prediction under a selection model. **Biometrics**, v.31, p.423, 1975.

HIRSCHHORN, J. N.; DALY, M. J. Genome-wide association studies for common diseases and complex traits. **Nature**, v. 6, 2005.

ILLUMINA. 2006. “TOP/BOT” strand and “A/B” allele: A guide to Illumina’s method for determining strand and allele for the GoldenGate and Infinium assays. Technical note. http://res.illumina.com/documents/products/technotes/technote_topbot.pdf (Acessado em 20 de Novembro de 2014).

IWASAKI, S.; MIYAKE, M.; HAYASHI, S.; WATANABE, H.; NAGASAWA, Y.; TERADA, S.; ASO, H. Effect of myostatin on chemokine expression in regenerating skeletal muscle cells. **Cells Tissues Organs**, v. 198, n. 1, p. 66-74, 2013.

JANI, S.D.; ARGRAVES, G.L.; BARTH, J.L.; ARGRAVES, W.S. GENEMESH: a web-based microarray analysis tool for relating differentially expressed genes to MeSH terms. **BMC Bioinformatics**, v. 11, p. 166, 2010.

JOHNSTON, D. J.; TIER, B.; GRASER, H. U. Beef cattle breeding in Australia with genomics: opportunities and needs. **Animal Production Science**, v. 52, n. 3, p. 100-106, 2012.

KOHAVI, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: **Ijcai**, v. 14, n. 2, p. 1137-1145, 1995.

KOOTS, K.R.; GIBSON, J.P.; WILTON, J.W. Analyses of published genetic parameters estimates for beef production traits. 1- Heritability. **Animal Breeding Abstracts**, v.62, n. 5, p.309-338, 1994.

KOURY FILHO, W.; ALBUQUERQUE, L. G. D.; FORNI, S.; SILVA, J. A. I.; YOKOO, M. J.; ALENCAR, M. M. D. Estimativas de parâmetros genéticos para os escores visuais e suas associações com peso corporal em bovinos de corte. **Revista Brasileira de Zootecnia**, v.35, n. 2, p. 1015-1022, 2010.

LANDE, R.; THOMPSON, R. Efficiency of marker-assisted selection in the improvement of quantitative traits. **Genetics**, v. 124, n.3, p. 743-756, 1990.

LEGARRA, A.; ROBERT-GRANIÉ, C.; MANFREDI, E.; ELSEN, J. M. Performance of genomic selection in mice. **Genetics**, v. 180, p. 611–618, 2008.

LEGARRA, A.; AGUILAR, I.; MISZTAL, I. A relationship matrix including full pedigree and genomic information. **Journal of Dairy Science**, v. 92, p. 4656–4663, 2009.

LI, F.; CHEN, H.; LEI, C. Z.; REN, G.; WANG, J.; LI, Z. J.; WANG, J. Q. Novel SNPs of the bovine GAD1/gad67 gene and their association with growth traits in three native Chinese cattle breeds. **Molecular biology reports**, v. 37, n. 1, p. 501-505, 2010.

LOURENCO, D. A. L.; TSURUTA, S.; FRAGOMENI, B. D.; MASUDA, Y.; POCRNIC, I.; AGUILAR, I.; MISZTAL, I. Issues in commercial application of single-step genomic BLUP for genetic evaluation in American Angus. **Journal of Animal Science**, v. 94 (supplement5), p. 144-145, 2016.

LOUVIERE, J. J.; HENSHER, D. A.; SWAIT, J. D. Stated choice methods: analysis and applications. Cambridge University Press, 2000.

LUND, M. S.; SAHANA, G.; DE KONING, D. J.; SU, G.; CARLBORG, Ö. Comparison of analyses of the QTLMAS XII common dataset. I: Genomic selection. In: **BMC proceedings**, BioMed Central, v. 3, n. 1, p. 1, 2009.

MACNEIL, M. D.; NKUMAH, J. D.; WOODWARD, B. W.; NORTHCUTT, S. L. Genetic evaluation of Angus cattle for carcass marbling using ultrasound and genomic indicators. **Journal of Animal Science**, v. 88, p. 517–522, 2010.

MARCONDES, C. R.; PANETO, J. C. C.; SILVA, J. A.; OLIVEIRA, H. N. D.; LÔBO, R. B. Comparação entre análises para permanência no rebanho de vacas Nelore utilizando modelo linear e modelo de limiar. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, v. 57, p. 234-240, 2005.

McFADDEN, D. Conditional logit analysis of qualitative choice behavior. In *Frontiers in Econometrics*, P. Zarembka (Ed.), p. 105-142, Academic Press, New York, 1970.

McHUGH, N.; CROMIE, A. R.; EVANS, R. D.; Berry, D. P. Validation of national genetic evaluations for maternal beef cattle traits using Irish field data. **Journal of animal science**, v. 92, n. 4, p. 1423-1432, 2014.

MEGGIOLARO, D.; PORCELLI, F.; CARNEVALI, A.; CREPALDI, P.; SAVARESE, E.; FERRANDI, B. A possible role of Fas antigen in ejaculated spermatozoa of fertile bulls: an immunocytochemical quantitative approach. **Acta histochemica**, v. 107, n.6, p. 463-468, 2006.

MEUWISSEN, T. H. E.; HAYES, B. J.; GODDARD, M. E. Prediction of total genetic value using genome-wide dense marker maps. **Genetics**, v. 157, p. 1819–1829, 2001.

MEYER, K. Covariance matrices for growth traits of Australian Polled Hereford cattle. **Animal Production Science**, v. 57, p. 37-45, 1993.

MEYER, K.; JOHNSTON, D. J.; GRASER, H. U. Estimates of the complete genetic covariance matrix for traits in multi-trait genetic evaluation of Australian Hereford cattle. **Australian Journal of Agricultural Research**, v. 55, p. 195-210, 2004.

MISZTAL, I.; TSURUTA, S.; STRABEL, T.; AUVRAY, B.; DRUET, T.; LEE, D. H. BLUPF90 and related programs (BGF90) In: 7th World Congress on Genetics Applied to Livestock Production, Montpellier, France. CD-ROM. Communication number 28-07, 2002.

MISZTAL, I.; LEGARRA, A.; AGUILAR, I. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. **Journal of Dairy Science**, v. 92, p. 4648–4655, 2009.

MOROTA, G.; PEÑAGARICANO, F.; PETERSEN, J.L.; CIOBANU, D.C.; TSUYUZAKI, K.; NIKAIIDO, I. An application of MeSH enrichment analysis in livestock. **Animal Genetics**, v. 46, p. 381–387, 2015.

- MWANSA, P.B.; CREWS, D.H.J.; WILTON, J.W.; KEMP, R.A. Multiple trait selection for maternal productivity in beef cattle. **Journal of Animal Breeding and Genetics**, v. 119, p. 391-399, 2002.
- NAKAZATO, T.; BONO, H.; MATSUDA, H.; TAKAGI, T. GENDOO: functional profiling of gene and disease features using MeSH vocabulary. **Nucleic Acids Research**, 37, p. W166–9, 2009.
- NELSON, S.J.; SCHOPEN, M.; SAVAGE, A.G.; SCHULMAN, J. L.; ARLUK, N. The MeSH translation maintenance system: structure, interface design, and implementation. **Medinfo** v. 11, p. 67–9, 2004.
- NEVES, H. H.; CARVALHEIRO, R.; O'BRIEN, A. M. P.; UTSUNOMIYA, Y. T.; DO CARMO, A. S.; SCHENKEL, F. S.; DA SILVA, M. V. Accuracy of genomic predictions in *Bos indicus* (Nelore) cattle. **Genetics Selection Evolution**, v. 46, p. 17, 2014.
- NISHIMURA, S.; WATANABE, T.; MIZOSHITA, K.; TATSUDA, K.; FUJITA, T.; WATANABE, N.; SUGIMOTO, Y.; TAKASUGA, A. Genome-wide association study identified three major QTL for carcass weight including the PLAG1-CHCHD7 QTN for stature in Japanese Black cattle. **BMC genetics**, v. 13, n. 1, p. 40, 2012.
- O'BRIEN, A. M. P.; UTSUNOMIYA, Y. T.; MÉSZÁROS, G.; BICKHART, D. M.; LIU, G. E.; VAN TASSELL, C. P.; SÖLKNER, J. Assessing signatures of selection through variation in linkage disequilibrium between taurine and indicine cattle. **Genetics Selection Evolution**, v. 46, p. 19, 2014.
- OLIVEIRA, M. M.; CARDOSO, F. F.; OSÓRIO, J. C. S. Componentes de variância e parâmetros genéticos em uma população multirracial Nelore-Angus sob enfoque Bayesiano. **Revista Brasileira de Zootecnia**. v.39, n.11, p.2426-2433, 2010.
- ONTERU, S.K.; FAN, B.; NIKKILÄ, M. T.; GARRICK, D. J.; STALDER, K. J.; ROTHSCCHILD, M. F. Whole-genome association analyses for lifetime reproductive traits in the pig. **Journal of Animal Science**, v. 89, n. 4, p. 988-995, 2010.
- ONTERU, S. K.; GORBACH, D. M.; YOUNG, J. M.; GARRICK, D. J.; DEKKERS, J. C.; ROTHSCCHILD, M. F. Whole genome association studies of residual feed intake and related traits in the pig. **PloS one**, v. 8, n. 6, p. e61756, 2013.
- PARK, T.; CASELLA, G. The Bayesian LASSO. **Journal of the American Statistical Association**, v. 103, n. 482, p.681-686, 2008.
- PAUSCH, H.; FLISIKOWSKI, K.; JUNG, S.; EMMERLING, R.; EDEL, C.; GÖTZ, K. U.; FRIES, R. Genome-wide association study identifies two major loci affecting calving ease and growth-related traits in cattle. **Genetics**, v. 187, n. 1, p. 289-297, 2011.
- PEDROSA, V. B.; ELER, J. P.; FERRAZ, J. B. S.; SILVA, J. A. I.; RIBEIRO, S.; SILVA, M. R.; PINTO, L. F. B. Parâmetros genéticos do peso adulto e características

de desenvolvimento ponderal na raça Nelore. **Revista Brasileira de Saúde e Produção Animal**, v. 11, n. 1, p.104-113, 2010.

PÉREZ-CABAL, M. A.; CHARFEDDINE, N. Models for genetic evaluations of claw health traits in Spanish dairy cattle. **Journal of dairy science**, v. 98, n. 11, p. 8186-8194, 2015.

PEREIRA, J.C.C. **Melhoramento Genético Aplicado à Produção Animal**. FEPE-MVZ, Belo Horizonte, 5ª edição, 618 p., 2008.

PETERS, S. O.; KIZILKAYA, K.; GARRICK, D. J.; FERNANDO, R. L.; REECY, J. M.; WEABER, R. L.; SILVER, G. A.; THOMAS, M. G. Heritability and Bayesian genome-wide association study of first service conception and pregnancy in Brangus heifers. **Journal of animal science**, v. 91, n. 2, p. 605-612, 2013.

PIRES, B. C.; FARIA, C. U.; VIU, M. A. D. O.; TERRA, J. P.; LOPES, D. T.; MAGNABOSCO, C. U.; LÔBO, R. B. Modelos bayesianos de limiar e linear na estimação de parâmetros genéticos para características morfológicas de bovinos da raça Nelore. **Revista Brasileira de Saúde e Produção Animal**, v. 11, n. 3, p. 651-661, 2010.

PRYCE, J. E.; HAYES, B. J.; BOLORMAA, S.; GODDARD, M. E. Polymorphic regions affecting human height also control stature in cattle. **Genetics**, v. 187, p. 981–984, 2011.

R CORE TEAM. 2013. The R Project for Statistical Computing. <http://www.R-project.org/>. (Acessado em 15 de Maio de 2014).

RAMAYO-CALDAS, Y.; FORTES, M. R. S.; HUDSON, N. J.; PORTO-NETO, L. R.; BOLORMAA, S.; BARENDSE, W.; REVERTER, A. A marker-derived gene network reveals the regulatory role of, and in intramuscular fat deposition of beef cattle. **Journal of animal science**, v. 92, n. 7, p. 2832-2845, 2014.

RESENDE, M. D. V.; LOPES, P. S.; SILVA, R. L.; PIRES, I. E. Seleção genômica ampla (GWS) e maximização da eficiência do melhoramento genético. **Pesquisa Florestal Brasileira**, v. 56, p. 63-77, 2008.

RESENDE, M.D.V.; RESENDE JR., M.F.R.; AGUIAR, A.M.; ABAD, J.I.M.; MISSIAGGIA, A.A.; SANSALONI, C.; PETROLI, C.; GRATTAPAGLIA, D. Computação da Seleção Genômica Ampla (GWS). Colombo, PR: Embrapa, Documentos 210, 2010. 79p.

RINCON, G.; WEBER, K. L.; VAN EENENNAAM, A. L. V.; GOLDEN, B. L.; MEDRANO, J. F. Hot topic: Performance of bovine high-density genotyping platforms in Holsteins and Jerseys. **American Dairy Science Association**, v. 94, n. 12, 2011.

ROBAKOWSKA-HYŻOREK, D.; OPRZĄDEK, J.; ŻELAZOWSKA, B.; OLBROMSKI, R.; ZWIERSZCHOWSKI, L. Effect of the g.-723G→ T Polymorphism in the Bovine

Myogenic Factor 5 (Myf5) Gene Promoter Region on Gene Transcript Level in the Longissimus Dorsi Muscle and on Meat Traits of Polish Holstein-Friesian Cattle. **Biochemical genetics**, v. 48, p. 450-464, 2010.

SAATCHI, M.; MCCLURE, M. C.; MCKAY, S. D.; ROLF, M. M.; KIM, J.; DECKER, J. E.; TAXIS, T. M.; CHAPPLE, R. H.; RAMEY, H. R.; NORTHCUTT, S. L.; BAUCK, S.; WOODWARD, B.; DEKKERS, J. C.; FERNANDO, R. H.; SCHNABEL, R. D.; GARRICK, D. J.; TAYLOR, J. F. Accuracies of genomic breeding values in American Angus beef cattle using K-means clustering for cross-validation. **Genetics Selection Evolution**, v. 43, p. 40, 2011.

SAATCHI, M.; SCHNABEL, R. D.; ROLF, M. M.; TAYLOR, J. F.; GARRICK, D. J. Accuracy of direct genomic breeding values for nationally evaluated traits in US Limousin and Simmental beef cattle. **Genetics Selection Evolution**, v. 44, p. 38, 2012.

SAATCHI, M.; WARD, J.; GARRICK, D. J. Accuracies of direct genomic breeding values in Hereford beef cattle using national or international training populations. **Journal of Animal Science**, v. 91, n. 4, p. 1538–1551, 2013.

SAATCHI, M.; SCHNABEL, R. D.; TAYLOR, J. F.; GARRICK, D. J. Large-effect pleiotropic or closely linked QTL segregate within and across ten US cattle breeds. **BMC genomics**, v. 15, n. 1, p. 1, 2014.

SANTANA, M. H. A.; UTSUNOMIYA, Y. T.; NEVES, H. H. R.; GOMES, R. C.; GARCIA, J. F.; FUKUMASU, H.; SILVA, S. L.; LEME, P. R.; COUTINHO, L. L.; ELER, J. P.; FERRAZ, J. B. S. Genome-wide association study for feedlot average daily gain in Nelore cattle (*Bos indicus*). **Journal of Animal Breeding and Genetics**, v. 131, n. 3, p. 210-216, 2014.

SARGOLZAEI, M.; CHESNAIS, J. P.; SCHENKEL, F. S. FImpute – An efficient imputation algorithm for dairy cattle populations. **Journal of Animal Science**, v. 89, p. 333, 2011.

SCHURINK, A.; WOLC, A.; DUCRO, B. J.; FRANKENA, K.; GARRICK, D. J.; DEKKERS, J. C. M.; VAN ARENDONK, J. A. M. Genome-wide association study of insect bite hypersensitivity in two horse populations in the Netherlands. **Genetics Selection Evolution**, v. 44, n.1, p. 31, 2012.

SHIN, Y. K.; MARTIN, B.; KIM, W.; WHITE, C. M.; JI, S.; SUN, Y.; SMITH, R. G.; SÉVIGNY, J.; TSCHÖP, M. H.; MAUDSLEY, S.; EGAN, J. M. Ghrelin is produced in taste cells and ghrelin receptor null mice show reduced taste responsivity to salty (NaCl) and sour (citric acid) tastants. **PLoS One**, v. 5, n. 9, p. e12729, 2010.

SILVA, J.A.V.; DIAS, L.T.; ALBUQUERQUE, L.G. Estudo genético da precocidade sexual de novilhas em um rebanho Nelore. **Revista Brasileira de Zootecnia**, v. 34, n. 5, p.1568-1572, 2005.

SILVA, J. A. II V.; MARCELO, E. T.; RIBEIRO, C. B.; MAIORANO, A. M.; CURI, R. A.; OLIVEIRA, H. N.; MOTA, M. D. S. Análise genética de características de

crescimento e perímetro escrotal em bovinos da raça Brangus. **Pesquisa Agropecuária Brasileira**, v.47, n.8, p.1166-1173, 2012.

SMITH, B.J. BOA: an R package for MCMC output convergence assessment and posterior inference. *Journal Statistical Software*, v. 21, n. 11, p. 1–37, 1997.

SNELLING, W. M.; MACNEIL, M. D.; KRESS, D. D.; ANDERSON, D. C.; TESS, M. W.. Factors influencing genetic evaluations of linebred Hereford cattle in diverse environments. **Journal of Animal Science**, v. 74, p. 1499-1510, 1996.

SNELLING, W. M.; ALLAN, M. F.; KEELE, J. W.; KUEHN, L. A.; MCDANELD, T.; SMITH, T. P.; SOSTEGARD, T. S.; THALLMAN, R. M.; BENNETT, G. L. Genome-wide association study of growth in crossbred beef cattle. **Journal of Animal Science**, v. 88, n. 3, p. 837-848, 2010.

SNELLING, W. M.; CUSHMAN, R. A; FORTES, M. R. S.; REVERTER, A.; BENNETT, G. L.; KEELE, J. W.; KUEHN, L. A.; MCDANELD, T. G.; THALLMAN, R. M.; THOMAS, M. G. How SNP chips will advance our knowledge of factors controlling puberty and aid in selecting replacement females. **Journal of Animal Science**, v. 90, p. 1152-1165, 2012.

SOMASHEKAR, L.; SELVARAJU, S.; PARTHIPAN, S.; RAVINDRA, J. P. Profiling of sperm proteins and association of sperm PDC-109 with bull fertility. **Systems biology in reproductive medicine**, v. 61, n. 6, p. 376-387, 2015.

SORENSEN, D.A.; ANDERSEN, S.; GIANOLA, D.; KORSGAARD, I. Bayesian inference in threshold models using Gibbs sampling. **Genetics Selection Evolution**, v. 27, p. 229-249, 1995.

STRANDEN, I.; GARRICK, D. J. Derivation of equivalent computing algorithms for genomic predictions and reliabilities of animal merit. **Journal of Dairy Science**, v. 92, p. 2971–2975, 2009.

SU, G.; GULDBRANDTSEN, B.; GREGERSEN, V. R.; LUND, M. S. Preliminary investigation on reliability of genomic estimated breeding values in the Danish Holstein population. **Journal of dairy science**, v. 93(3), p. 1175-1183, 2010.

SUN, X.; HABIER, D.; FERNANDO, R. L.; GARRICK, D. J.; DEKKERS, J. C. Genomic breeding value prediction and QTL mapping of QTLMAS2010 data using Bayesian methods. In: **BMC proceedings**, v. 5, n. 3, p. 1, BioMed Central, 2011.

TIZIOTO, P.C.; DECKER, J.E.; TAYLOR, J.F.; SCHNABEL, R.D.; MUDADU, M.A.; SILVA, F.L.; MOURÃO, G.B.; COUTINHO, L.L.; THOLON, P.; SONSTEGARD, T.S.; ROSA, A.N.; ALENCAR, M.M.; TULLIO, R.R.; MEDEIROS, S.R.; NASSU, R.T.; FEIJÓ, G.L.D.; SILVA, L.O.C.; TORRES, R.A.; SIQUEIRA, F.; HIGA, R.H.; REGITANO, L.C.A. Genome scan for meat quality traits in Nelore beef cattle. **Physiological Genomics**, v.45, p.1012-1020, 2013.

TOOSI, A.; FERNANDO, R. L.; DEKKERS, J. C. M. Genomic selection in admixed and crossbred populations¹. **Journal of Animal Science**, v. 88, p. 32-46, 2010.

TORAL, F. L. B. et al. Genetic parameters and response to selection for post-weaning weight gain, visual scores and carcass traits in Hereford and Herefordx Nellore cattle, **Livestock Science**, v. 137(1), p. 231-237, 2011.

TORRES-VAZQUEZ, J. A.; SPANGLER, M. L. Genetic parameters for docility, weaning weight, yearling weight, and intramuscular fat percentage in Hereford cattle. **Journal of animal science**, v. 94(1), p. 21-27, 2016.

TSUYUZAKI, K.; MOROTA, G.; ISHII, M.; NAKAZATO, T.; MIYAZAKI S.; NIKAIDO, I. MeSH ORA framework: R/Bioconductor packages to support MeSH over-representation analysis. **BMC Bioinformatics**, v. 16, n. 1, p. 45, 2015.

UTSUNOMIYA, Y. T., DO CARMO, A. S., CARVALHEIRO, R., NEVES, H. H., MATOS, M. C., ZAVAREZ, L. B.; O'BRIEN, A. M. P.; SÖLKNER, J.; MCEWAN, J. C.; COLE, J. B.; VAN TASSELL, C. P.; SCHENKEL, F. S.; DA SILVA, M. V. G. B.; PORTO NETO, L. R.; SONSTEGARD, T. S.; JOSÉ F GARCIA, J. F. Genome-wide association study for birth weight in Nellore cattle points to previously described orthologous genes affecting human and bovine height. **BMC genetics**, v. 14, n. 1, p. 52, 2013.

VAN EENENNAAM, A. L.; DRAKE, D. J. Where in the beef-cattle supply chain might DNA tests generate value? **Animal production science**, v. 52, n. 3, p. 185-196, 2012.

VAN EENENNAAM, A. L.; WEIGEL, K. A.; YOUNG, A. E.; CLEVELAND, M. A.; DEKKERS, J. C. Applied Animal Genomics: Results from the Field. **Annual Review of Animal Biosciences**, v. 2, n. 1, p. 105–139, 2014.

VAN TASSEL, C. P.; VAN VLECK, L.D. Multiple-trait Gibbs sampler for animal models; flexible programs for Bayesian and likelihood-based (co)variance components inference. **Journal of Animal Science**, v. 74, p. 2586–2597, 1996.

VAN TASSEL, C.P. The use of SNP chips for selection of dairy cattle. In: Applied Genomics for Sustainable Livestock Breeding Meeting, 1., 2011, Melborn. Proceedings... Melborn: 2011. (CD-ROM).

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

VANRADEN, P. M.; VAN TASSELL, C. P.; WIGGANS, G. R.; SONSTEGARD, T. S.; SCHNABEL, R. D.; TAYLOR, J. F.; SCHENKEL, F. S. Invited review: Reliability of genomic predictions for North American Holstein bulls. **Journal of Dairy Science**, v. 92, p. 16–24, 2009.

VARONA, L.; MIZTAL, I.; BERTRAND, J.K. Threshold-linear versus linear-linear analysis of birth weight and calving ease using an animal model: I. Variance Component Estimation. **Journal of Animal Science**, v. 77, p. 1994-2002, 1999.

VESELÁ, Z.; VOSTRÝ, L.; ŠAFUS, P. Linear and linear-threshold model for genetic parameters for SEUROP carcass traits in Czech beef cattle. **Czech Journal of Animal Science**, v. 56, n. 9, p. 414-425, 2011.

WANG, H.; MISZTAL, I.; AGUILAR, I.; LEGARRA, A.; MUIR, W. M. Genome-wide association mapping including phenotypes from relatives without genotypes. **Genetics Research**, v. 94, n. 2 p. 73-83, 2012.

WANG, H.; MISZTAL, I.; AGUILAR, I.; LEGARRA, A.; FERNANDO, R.L.; VITEZICA, Z.; OKIMOTO, R.; WING, T.; HAWKEN, R.; MUIR, W.M. Genome-wide association mapping including phenotypes from relatives without genotypes in a single-step (ssGWAS) for 6-week body weight in broiler chickens. **Frontiers in Genetics**, v. 5, n. 134, 2014.

WENG, Z.; SU, H.; SAATCHI, M.; LEE, J.; THOMAS, M. G.; DUNKELBERGER, J. R.; GARRICK, D. J. Genome-wide association study of growth and body composition traits in Brangus beef cattle. **Livestock Science**, v. 183, p. 4-11, 2016.

WHITTAKER, J.C.; THOMPSON, R.; DENHAM, M.C. Marker-assisted selection using ridge regression. **Genetics Research**, v. 75, n. 4, p. 249-252, 2000.

WILLHAM, R. L. The role of maternal effects in animal breeding: III. Biometrical aspects of maternal effects in animals. **Journal of Animal Science**, v. 35, n. 6, p. 1288-1293, 1972.

WOLC, A.; ARANGO, J.; SETTAR, P.; FULTON, J. E.; O'SULLIVAN, N. P.; PREISINGER, R.; DEKKERS, J. C. Persistence of accuracy of genomic estimated breeding values over generations in layer chickens. **Genetics Selection Evolution**, v. 43, p. 23, 2011.

WOLC, A.; ARANGO J.; SETTAR, P.; FULTON, J. E.; O'SULLIVAN, N. P.; PREISINGER, R.; HABIER, D.; FERNANDO, R.; GARRICK, D. J.; HILL, W. G.; DEKKERS, J. C. M. Genome-wide association analysis and genetic architecture of egg weight and egg uniformity in layer chickens. **Animal Genetics**, v. 43 (SUPPL. 1), p. 87-96, 2012.

XU, S. An empirical bayes method for estimating epistatic effects of quantitative trait loci. **Biometrics**, v.63, n.2, p.513-521, 2007.

YOKOO, M. J. I.; WERNECK, J. N.; PEREIRA, M. C.; ALBUQUERQUE, L. G. D.; KOURY FILHO, W.; SAINZ, R. D.; LÔBO, R. B.; ARAUJO, F. R. D. C. Correlações genéticas entre escores visuais e características de carcaça medidas por ultrassom em bovinos de corte. **Pesquisa Agropecuária Brasileira**, v. 44, p. 197-202, 2009.

ZARE, Y.; SHOOK, G. E.; COLLINS, M. T.; KIRKPATRICK, B. W. Genome-wide association analysis and genomic prediction of *Mycobacterium avium* subspecies paratuberculosis infection in US Jersey cattle. **PLoS One**, v. 9, n. 2, p. e88380, 2014.

ZHONG, S.; DEKKERS, J. C. M.; FERNANDO, R. L.; JANNINK, J. L. Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: A barley case study. **Genetics**, v. 182, p. 355–364, 2009.

ZIMIN, A. V.; DELCHER, A. L.; FLOREA, L.; KELLEY, D. R.; SCHATZ, M. C.; PUIU, D.; HANRAHAN, F.; PERTEA, G.; VAN TASSELL, C. P.; SONSTEGARD, T. S. A whole-genome assembly of the domestic cow, *Bos taurus*. **Genome biology**, v. 10, n. 4, p. R42, 2009.

ANEXOS

Anexo 1. Janelas de maior representatividade da variância genética (% Var) juntamente com seu SNP de maior *Model Frequency* e *t-like* para a característica de peso ao nascer em bovinos Hereford e Braford.

Obs	Window	Start SNP Name	End SNP Name	N SNPs	% Var	Cum.Var	chr_Mb	Top SNP Name	Ref Top SNP	ModelFreq	t.like
1	1523	UA-IFASA-7482	Hapmap40120-BTA-34288	15	4.94	4.94	14_24	Hapmap40120-BTA-34288	rs41627956	0.4211	1.051
2	585	ARS-BFGL-NGS-58789	Hapmap50837-BTA-98392	10	3.91	8.85	5_46	ARS-BFGL-NGS-73392	rs109590980	0.6164	1.198
3	1961	Hapmap23517-BTC-058008	BTB-01988115	23	2.0	10.85	20_4	BTB-00144202	rs43349772	0.4232	1.045
4	1936	Hapmap60860-rs29018127	ARS-BFGL-NGS-109970	20	1.48	12.33	19_44	BTA-45551-no-rs	rs41603040	0.4746	1.064
6	149	ARS-BFGL-NGS-70354	Hapmap25017-BTA-124913	17	1.31	13.64	1_149	ARS-BFGL-NGS-103594	rs41612865	0.4257	1.043
7	713	ARS-BFGL-NGS-14410	BTB-00257125	23	1.07	14.72	6_52	BTB-00257125	rs43466396	0.4046	1.028
8	1966	ARS-BFGL-NGS-118231	Hapmap54493-rs29018712	20	0.98	15.7	20_9	BTA-51491-no-rs	rs41638183	0.2747	0.978
9	132	ARS-BFGL-NGS-116637	ARS-BFGL-BAC-11042	18	0.95	16.66	1_132	Hapmap38506-BTA-52316	rs41602583	0.3719	1.015
10	2249	ARS-BFGL-NGS-22009	ARS-BFGL-NGS-54587	23	0.88	17.53	24_33	ARS-BFGL-NGS-100144	rs42934321	0.2338	0.971
11	2044	ARS-BFGL-BAC-15488	Hapmap40073-BTA-53493	19	0.87	18.4	21_15	Hapmap32600-BTA-149891	rs109870665	0.3332	0.996
12	702	ARS-BFGL-NGS-101499	BTA-75905-no-rs	28	0.76	19.16	6_41	Hapmap32713-BTC-036432	rs43460584	0.27	0.977
13	2093	ARS-BFGL-NGS-14107	ARS-BFGL-NGS-63564	12	0.7	20.59	21_64	Hapmap39756-BTA-52915	rs41637735	0.3273	0.991
14	1817	ARS-BFGL-NGS-68160	ARS-BFGL-NGS-5584	24	0.68	21.27	17_66	ARS-BFGL-NGS-111681	rs110879614	0.2964	0.988
15	2476	ARS-BFGL-NGS-79157	ARS-BFGL-NGS-51329	21	0.64	21.92	29_9	BTA-03099-rs29010866	rs29010866	0.3014	0.988
16	2174	Hapmap49122-BTA-32454	BTA-55470-no-rs	15	0.62	22.54	23_11	ARS-BFGL-NGS-118921	rs109115482	0.2822	0.984
17	182	ARS-BFGL-NGS-107035	Hapmap40436-BTA-46804	20	0.62	23.15	2_23	ARS-BFGL-NGS-14141	rs110441014	0.1667	0.95
18	684	ARS-BFGL-NGS-91446	BTB-00247050	15	0.59	23.74	6_23	Hapmap49541-BTA-24412	rs41633767	0.2064	0.96
19	1965	Hapmap32801-BTA-133738	BTA-51494-no-rs	23	0.59	24.33	20_8	BTB-00771463	rs41930731	0.041	0.904
20	393	ARS-BFGL-NGS-35251	BTB-00151079	23	0.56	24.9	3_97	ARS-BFGL-NGS-114466	rs109451073	0.2362	0.968
21	2279	ARS-BFGL-NGS-100100	ARS-BFGL-NGS-4009	18	0.51	25.41	25_0	Hapmap40362-BTA-59892	rs41649720	0.1014	0.935
22	1530	BTB-00043033	ARS-BFGL-NGS-93027	20	0.5	25.91	14_31	ARS-BFGL-NGS-107639	rs109292290	0.1342	0.944
23	715	BTB-01280026	ARS-BFGL-NGS-112970	16	0.49	26.39	6_54	BTB-01206363	rs42365512	0.0952	0.926
24	1455	ARS-BFGL-NGS-116606	ARS-BFGL-NGS-58346	21	0.46	26.86	13_40	ARS-BFGL-NGS-27815	rs42761643	0.1641	0.948
25	874	ARS-BFGL-NGS-103594	Hapmap30843-BTA-148402	7	0.45	28.21	7_93	ARS-BFGL-NGS-101886	rs41625563	0.1989	0.959
26	734	BES3_Contig393_647	Hapmap50775-BTA-76767	20	0.45	27.31	6_73	Hapmap53308-rs29024839	rs29024839	0.1001	0.931
27	733	BTB-01468535	BTB-01899352	22	0.45	27.76	6_72	Hapmap43934-BTA-76802	rs41605955	0.0305	0.898
28	2389	BTA-117811-no-rs	BTB-00952213	21	0.42	28.63	27_15	ARS-BFGL-NGS-97489	rs42111498	0.089	0.935
29	942	BTA-34427-no-rs	BTB-00392243	16	0.4	29.43	8_48	BTB-00392243	rs43598289	0.1554	0.947
30	2217	ARS-BFGL-BAC-3741	ARS-BFGL-NGS-11247	22	0.4	29.03	24_1	SCAFFOLD222971_1761	rs29020832	0.1223	0.941
31	1939	ARS-BFGL-NGS-20225	UA-IFASA-8219	20	0.39	29.82	19_47	BTB-00755336	rs41917898	0.0751	0.93
32	150	BTA-98387-no-rs	ARS-BFGL-NGS-14751	18	0.37	30.18	1_150	ARS-BFGL-NGS-41125	rs109981319	0.0954	0.935

33	2480	ARS-BFGL-NGS-86900	BTB-01941352	11	0.34	30.52	29_13	BTB-01941352	rs43046837	0.1605	0.95
34	134	ARS-BFGL-NGS-20128	ARS-BFGL-NGS-54112	15	0.32	30.84	1_134	ARS-BFGL-NGS-100206	rs109963113	0.1099	0.94
35	2378	Hapmap59709-rs29021868	ARS-BFGL-NGS-116642	16	0.3	31.74	27_4	ARS-BFGL-NGS-116642	rs109394114	0.1713	0.954
36	219	ARS-BFGL-NGS-95965	Hapmap38373-BTA-107037	12	0.3	31.14	2_60	ARS-BFGL-NGS-13806	rs109602190	0.1631	0.949
37	1917	ARS-BFGL-NGS-103247	ARS-BFGL-NGS-93006	27	0.3	31.44	19_25	ARS-BFGL-NGS-106007	rs110687354	0.0984	0.935
38	2194	ARS-BFGL-NGS-11408	ARS-BFGL-NGS-101938	14	0.27	32.61	23_31	ARS-BFGL-NGS-112824	rs110277740	0.1467	0.948
39	1516	Hapmap50014-BTA-51389	Hapmap31084-BTA-128979	17	0.27	33.14	14_17	ARS-BFGL-NGS-103654	rs109114543	0.0933	0.919
40	475	Hapmap25843-BTA-146186	Hapmap60763-rs29014471	23	0.27	32.88	4_57	BTB-01403737	rs42527853	0.0436	0.915
41	1516	ARS-BFGL-NGS-114087	Hapmap31084-BTA-128979	17	0.27	33.14	14_17	BTA-35990-no-rs	rs41627794	0.0249	0.89
42	2236	Hapmap47669-BTA-59022	BTA-57516-no-rs	17	0.26	33.66	24_20	ARS-BFGL-NGS-24496	rs109621296	0.1395	0.931
43	294	ARS-BFGL-NGS-117865	ARS-BFGL-NGS-63882	24	0.26	33.92	2_135	ARS-BFGL-NGS-111732	rs110513892	0.1091	0.935
44	2332	ARS-BFGL-NGS-5696	ARS-BFGL-NGS-22284	23	0.26	33.4	26_10	BTB-00924013	rs42083869	0.0924	0.932
45	294	ARS-BFGL-NGS-84618	ARS-BFGL-NGS-63882	24	0.26	33.92	2_135	ARS-BFGL-NGS-36460	rs43327919	0.0633	0.912
46	294	ARS-BFGL-NGS-86508	ARS-BFGL-NGS-63882	24	0.26	33.92	2_135	ARS-BFGL-NGS-2645	rs109601200	0.0162	0.885
47	2358	ARS-BFGL-BAC-12987	BTB-01314304	21	0.25	34.91	26_36	BTA-77888-no-rs	rs41590962	0.1551	0.948
48	2268	Hapmap39886-BTA-112449	Hapmap55314-rs29026474	16	0.25	34.42	24_52	ARS-BFGL-NGS-93055	rs110791437	0.1306	0.916
49	1342	Hapmap59127-rs29023698	ARS-BFGL-NGS-77196	22	0.25	34.17	12_16	ARS-BFGL-NGS-114512	rs110889743	0.1234	0.94
50	1508	ARS-BFGL-NGS-104268	Hapmap24767-BTC-058058	23	0.25	34.66	14_9	ARS-BFGL-BAC-2036	rs110188741	0.0802	0.924
51	1308	ARS-BFGL-NGS-82122	ARS-BFGL-NGS-30661	14	0.24	35.15	11_90	Hapmap39716-BTA-112467	rs41614197	0.0596	0.909
52	2270	Hapmap40932-BTA-27105	ARS-BFGL-NGS-85334	9	0.23	35.38	24_54	ARS-BFGL-NGS-116424	rs110740974	0.1344	0.944
53	1282	BTB-00100191	ARS-BFGL-NGS-33664	16	0.23	35.83	11_64	ARS-BFGL-NGS-33664	rs110138103	0.1301	0.943
54	1850	ARS-BFGL-NGS-110121	Hapmap52077-ss46526963	14	0.23	35.6	18_24	BTA-42865-no-rs	rs41600953	0.1169	0.939
55	2270	BTB-01693574	ARS-BFGL-NGS-85334	9	0.23	35.38	24_54	ARS-BFGL-NGS-113579	rs109835683	0.02	0.887
			Hapmap34873-					Hapmap34873-			
56	2041	UA-IFASA-6120	BES4_Contig386_675	20	0.21	36.45	21_12	BES4_Contig386_675	rs43708444	0.084	0.934
57	402	ARS-BFGL-NGS-20740	ARS-BFGL-NGS-40609	17	0.21	36.04	3_106	ARS-BFGL-NGS-40609	rs110237486	0.0802	0.918
58	2495	ARS-BFGL-NGS-102626	ARS-BFGL-NGS-79220	16	0.21	36.25	29_28	ARS-BFGL-NGS-79220	rs110881630	0.0606	0.885
59	402	ARS-BFGL-NGS-75486	ARS-BFGL-NGS-40609	17	0.21	36.04	3_106	BTB-01676489	rs42791325	0.02	0.893
60	1255	ARS-BFGL-NGS-32105	Hapmap46632-BTA-89801	15	0.2	36.66	11_37	Hapmap46632-BTA-89801	rs41662162	0.1158	0.937

Anexo 2. Janelas de maior representatividade da variância genética (% Var) juntamente com seu SNP de maior *Model Frequency* e *t-like* para a característica de peso a desmama ajustado para os 205 dias em bovinos Hereford e Braford.

Obs	Window	Start SNP Name	End SNP Name	N SNPs	% Var	Cum.Var	chr_Mb	Top SNP Name	Ref Top SNP	ModelFreq	t.like
1	1917	ARS-BFGL-NGS-84618	ARS-BFGL-NGS-93006	27	3.38	3.38	19_25	ARS-BFGL-NGS-4904	rs110133873	0.7123	1.313
2	1961	ARS-BFGL-NGS-20128	BTB-01988115	23	2.96	6.33	20_4	ARS-BFGL-NGS-25266	rs41932505	0.4493	1.054
3	151	ARS-BFGL-NGS-70523	BTA-16239-no-rs	19	2.63	8.96	1_151	ARS-BFGL-NGS-105623	rs109817439	0.6364	1.202
4	547	BTB-01507733	BTB-00219231	26	1.99	10.95	5_8	BTA-75037-no-rs	rs41655514	0.4934	1.077
5	1067	BES9_Contig573_967	BTA-83827-no-rs	28	1.27	12.22	9_59	ARS-BFGL-NGS-93427	rs110721265	0.3586	1.007
6	2268	ARS-BFGL-NGS-102626	rs29026474	16	0.77	12.99	24_52	ARS-BFGL-NGS-116083	rs110808470	0.2576	0.969
7	1506	Hapmap27091-BTC-048823	ARS-BFGL-NGS-20781	27	0.75	13.74	14_7	ARS-BFGL-BAC-8730	rs110165168	0.1761	0.951
8	463	BTB-00181882	ARS-BFGL-NGS-108242	17	0.68	14.42	4_45	BTB-00614426	rs41780457	0.1353	0.937
9	545	ARS-BFGL-NGS-43284	ARS-BFGL-NGS-39595	21	0.56	15.54	5_6	UA-IFASA-4428	rs29023001	0.1132	0.922
10	2084	BTB-01610675	ARS-BFGL-NGS-116152	13	0.56	14.98	21_55	ARS-BFGL-NGS-105670	rs110510910	0.1865	0.955
11	1268	BTA-88741-no-rs	ARS-BFGL-NGS-66467	14	0.52	16.06	11_50	BTB-01828679	rs42939302	0.1945	0.953
12	1270	Hapmap33275-BTA-98050	ARS-BFGL-NGS-117912	7	0.5	16.56	11_52	Hapmap33275-BTA-98050	rs41608892	0.1855	0.952
13	1226	BTB-00455305	ARS-BFGL-NGS-110878	21	0.48	17.04	11_8	BTB-00455305	rs43657228	0.1357	0.939
14	1263	ARS-BFGL-NGS-32643	ARS-BFGL-NGS-116676	19	0.47	18.45	11_45	ARS-BFGL-NGS-112743	rs109383510	0.1787	0.952
15	1719	Hapmap57710-rs29018089	ARS-BFGL-NGS-119467	18	0.47	17.98	16_50	ARS-BFGL-NGS-96028	rs109660893	0.1321	0.938
16	629	BTA-95130-no-rs	ARS-BFGL-NGS-365	14	0.47	17.51	5_90	BTA-74541-no-rs	rs41653536	0.1343	0.941
17	2197	ARS-BFGL-NGS-82538	BTA-106796-no-rs	8	0.46	18.91	23_34	ARS-BFGL-NGS-61632	rs109984909	0.1865	0.953
18	1598	Hapmap45702-BTA-93884	BTA-93194-no-rs	17	0.45	19.37	15_14	ARS-BFGL-NGS-117603	rs41746121	0.1133	0.931
19	1170	ARS-BFGL-NGS-43120	ARS-BFGL-NGS-44942	17	0.43	19.79	10_57	BTB-01125630	rs42286839	0.1498	0.943
20	549	ARS-BFGL-NGS-16627	22406	23	0.42	20.22	5_10	Hapmap54948-rs29009927	rs29009927	0.0861	0.931
21	556	ARS-BFGL-NGS-57294	ARS-BFGL-NGS-108063	5	0.39	21.39	5_17	ARS-BFGL-NGS-104042	rs109623851	0.145	0.941
22	2069	Hapmap48378-BTA-52242	BTB-01383788	15	0.39	21	21_40	Hapmap32050-BTA-135951	rs109402967	0.0782	0.912
23	1358	ARS-BFGL-NGS-104042	BTA-20647-no-rs	9	0.39	20.61	12_32	ARS-BFGL-NGS-39585	rs109758875	0.1764	0.942
24	1007	ARS-BFGL-NGS-19252	BTB-00377449	11	0.38	22.15	8_113	ARS-BFGL-NGS-106379	rs42390350	0.1521	0.945
25	580	ARS-BFGL-NGS-107896	ARS-BFGL-NGS-28318	13	0.38	21.77	5_41	BTA-97739-no-rs	rs43745968	0.0719	0.921
26	863	ARS-BFGL-NGS-72140	ARS-BFGL-NGS-11872	19	0.37	22.89	7_82	BES10_Contig573_944	rs43527554	0.1239	0.938
27	2406	Hapmap43360-BTA-79989	BTB-00967319	23	0.37	22.52	27_32	ARS-BFGL-NGS-30959	rs110773328	0.1124	0.92
28	1706	BTA-38746-no-rs	ARS-BFGL-NGS-110100	21	0.36	23.25	16_37	BTB-02086174	rs43185446	0.09	0.925
29	128	Hapmap59539-rs29025538	ARS-BFGL-NGS-31046	19	0.35	23.95	1_128	ARS-BFGL-NGS-92777	rs43265232	0.1224	0.937

30	775	ARS-BFGL-BAC-16161	BTB-00282326	25	0.35	23.6	6_114	BTB-01906617	rs43015979	0.1358	0.941
31	54	Hapmap48975-BTA-99363	ARS-BFGL-NGS-86970	23	0.33	24.28	1_54	ARS-BFGL-NGS-111929	rs109010613	0.0935	0.919
33	2087	ARS-BFGL-NGS-24785	ARS-BFGL-NGS-115017	12	0.32	24.6	21_58	ARS-BFGL-NGS-54269	rs109345757	0.0703	0.92
34	155	ARS-BFGL-NGS-115406	BTB-00023937	25	0.3	25.5	1_155	Hapmap47183-BTA-113360	rs41618669	0.1339	0.937
35	116	BTB-01191390	BTB-02083452	16	0.3	25.21	1_116	BTB-00051291	rs43258621	0.1131	0.932
36	1510	Hapmap57477-rs29022938	Hapmap30171-BTC- 063976	25	0.29	25.8	14_11	Hapmap25965-BTC-009503	rs41581271	0.0617	0.896
37	627	ARS-BFGL-BAC-31572	Hapmap59389- rs29023212	21	0.28	26.08	5_88	BTB-00234696	rs43438620	0.1113	0.921
38	275	Hapmap51627-BTA-74367	BTA-103078-no-rs	21	0.27	26.35	2_116	ARS-BFGL-NGS-117340	rs109164337	0.1042	0.933
39	114	BTB-00821810	ARS-BFGL-NGS-11580	14	0.26	26.61	1_114	ARS-BFGL-NGS-2027	rs109026214	0.1004	0.93
40	1432	Hapmap55381-rs29025399	Hapmap39401-BTA- 34229	19	0.25	27.36	13_17	ARS-BFGL-NGS-104014	rs110812783	0.1092	0.929
41	2426	UA-IFASA-7482	Hapmap47440-BTA- 90993	14	0.25	27.12	28_6	Hapmap43150-BTA-110538	rs41614882	0.1089	0.935
42	1521	ARS-BFGL-NGS-103093	Hapmap40063-BTA- 34263	24	0.25	26.86	14_22	ARS-BFGL-NGS-66553	rs42018872	0.0572	0.914
43	2302	ARS-BFGL-NGS-66728	Hapmap29724-BTA- 138409	14	0.24	28.55	25_23	Hapmap57584-rs29014865	rs29014865	0.07	0.91
44	2130	BTB-00401246	ARS-BFGL-NGS-20317	13	0.24	28.32	22_29	ARS-BFGL-NGS-20317	rs109201435	0.1342	0.938
45	157	ARS-BFGL-NGS-107132	BTA-15355-no-rs	26	0.24	28.08	1_157	ARS-BFGL-NGS-5053	rs110655223	0.0852	0.927
46	890	ARS-BFGL-NGS-24919	BTB-00329502	18	0.24	27.84	7_109	BTB-01842929	rs42952749	0.0815	0.922
47	1090	ARS-BFGL-NGS-4227	BTB-00400244	20	0.24	27.61	9_82	BTA-84397-no-rs	rs41594197	0.0763	0.926
48	289	ARS-BFGL-NGS-81742	ARS-BFGL-NGS-43912	15	0.23	28.78	2_130	ARS-BFGL-NGS-114524	rs110692481	0.0492	0.915
49	886	Hapmap39569-BTA-49765	ARS-BFGL-NGS-104100	19	0.22	29.89	7_105	BTB-00956573	rs42112860	0.0669	0.925
50	1997	ARS-BFGL-NGS-19413	ARS-BFGL-NGS-68808	19	0.22	29.67	20_40	ARS-BFGL-NGS-20739	rs109516416	0.0645	0.916
51	1964	Hapmap38275-BTA-101455	Hapmap25537-BTA- 134848	24	0.22	29.45	20_7	BTB-01197460	rs42356467	0.0868	0.93
52	2182	ARS-BFGL-NGS-14107	BTA-55724-no-rs Hapmap25799-BTA-	22	0.22	29.23	23_19	ARS-BFGL-NGS-101622	rs109011792	0.0633	0.909
53	1293	ARS-BFGL-NGS-103261	126762	17	0.22	29	11_75	ARS-BFGL-NGS-35610	rs109213266	0.0493	0.908
54	294	BTA-63505-no-rs	ARS-BFGL-NGS-63882	24	0.21	30.31	2_135	ARS-BFGL-NGS-110452	rs110708447	0.0999	0.93
55	2068	Hapmap48377-BTA-52225	BTB-00819205	18	0.21	30.1	21_39	Hapmap33736-BTA-163613	rs109870937	0.0739	0.925
56	2010	BTA-52274-no-rs	BTB-01435010	24	0.2	30.71	20_53	BTB-00788635	rs41951674	0.0803	0.922
57	1416	ARS-BFGL-NGS-71458	UA-IFASA-5150	21	0.2	30.51	13_1	BTB-01721718	rs42832792	0.0758	0.922

Anexo 3. Janelas de maior representatividade da variância genética (% Var) juntamente com seu SNP de maior *Model Frequency* e *t-like* para a característica de peso ao sobreano ajustado para os 550 dias em bovinos Hereford e Braford.

Obs	Window	Start SNP Name	End SNP Name	N SNPs	% Var	Cum.Var	chr_Mb	Top SNP Name	Ref Top SNP	ModelFreq	t.like
1	1961	ARS-BFGL-NGS-20128	BTB-01988115	23	4.14	4.14	20_4	ARS-BFGL-NGS-102165	rs41932453	0.3443	1.013
2	1239	Hapmap43962-BTA-86597	Hapmap42711-BTA-87541	24	1.9	6.05	11_21	ARS-BFGL-BAC-11047	rs110245673	0.5254	1.106
3	2126	Hapmap50387-BTA-53912	Hapmap32083-BTA-147240	12	1.32	7.37	22_25	BTA-53914-no-rs	rs41583553	0.3289	1.004
4	1127	ARS-BFGL-NGS-113038	BTB-00410322	24	1.03	8.39	10_13	ARS-BFGL-NGS-71024	rs43622359	0.4079	1.035
5	2396	ARS-BFGL-NGS-20982	ARS-BFGL-NGS-113093	22	0.99	9.38	27_22	ARS-BFGL-NGS-26366	rs43334569	0.2921	0.991
6	511	Hapmap41207-BTA-22424	BTB-00202925	17	0.9	10.28	4_93	BTB-00203359	rs43411498	0.3314	1.0
////7	377	Hapmap51621-BTA-68487	ARS-BFGL-NGS-117782	26	0.86	11.15	3_81	ARS-BFGL-NGS-103424	rs110909637	0.3592	1.013
8	1604	ARS-BFGL-NGS-12339	ARS-BFGL-NGS-57433	14	0.86	12.0	15_20	ARS-BFGL-NGS-97051	rs42357017	0.4145	1.036
9	1694	ARS-BFGL-NGS-87462	Hapmap53517-rs29024619	10	0.85	12.85	16_25	BTA-38177-no-rs	rs41633929	0.2819	0.983
10	310	UA-IFASA-8925	ARS-BFGL-NGS-102327	14	0.85	13.7	3_14	Hapmap53344-ss46526153	rs41255296	0.2996	0.99
11	1560	ARS-BFGL-BAC-21623	ARS-BFGL-NGS-10025	13	0.84	14.54	14_61	BTB-00574951	rs41742577	0.4544	1.057
12	2087	ARS-BFGL-NGS-24785	ARS-BFGL-NGS-115017	12	0.8	15.34	21_58	ARS-BFGL-NGS-23598	rs109377027	0.2141	0.966
13	2398	ARS-BFGL-NGS-1516	Hapmap41317-BTA-62644	18	0.71	16.05	27_24	BTA-62507-no-rs	rs41584803	0.3175	0.998
14	629	ARS-BFGL-NGS-32643	ARS-BFGL-NGS-365	14	0.65	16.7	5_90	ARS-BFGL-NGS-113558	rs110024250	0.2897	0.987
15	1524	Hapmap41234-BTA-34285	ARS-BFGL-NGS-8308	18	0.6	17.3	14_25	BTB-01779799	rs42892557	0.147	0.948
16	455	Hapmap47073-BTA-66626	Hapmap44397-BTA-70124	20	0.57	17.87	4_37	BTA-70113-no-rs	rs41586860	0.3233	0.99
17	951	BTB-00349358	BTB-00529060	20	0.57	18.44	8_57	ARS-BFGL-NGS-33376	rs109592129	0.2515	0.971
18	645	ARS-BFGL-NGS-118977	ARS-BFGL-NGS-2465	16	0.55	18.99	5_106	ARS-BFGL-NGS-39379	rs110421124	0.1678	0.949
19	1516	BTA-108463-no-rs	Hapmap31084-BTA-128979	17	0.53	19.52	14_17	ARS-BFGL-NGS-82859	rs110506037	0.1935	0.961
20	2172	ARS-BFGL-NGS-55885	ARS-BFGL-NGS-118658	20	0.51	20.03	23_9	BTA-98024-no-rs	rs110291685	0.2249	0.968
21	1315	ARS-BFGL-NGS-21607	ARS-BFGL-NGS-25573	18	0.49	20.52	11_97	ARS-BFGL-NGS-118218	rs108984228	0.2657	0.977
22	582	Hapmap58370-rs29014396	BTB-00226603	19	0.46	20.98	5_43	BTB-01477513	rs42600489	0.1146	0.94
23	561	BTA-72912-no-rs	BTA-72966-no-rs	15	0.46	21.44	5_22	Hapmap23146-BTA-123179	rs43710144	0.2426	0.967
24	79	ARS-BFGL-BAC-33415	ARS-BFGL-NGS-111888	19	0.44	21.88	1_79	ARS-BFGL-NGS-111725	rs109121314	0.1523	0.949
25	1989	ARS-BFGL-NGS-118998	ARS-BFGL-NGS-18037	8	0.43	22.31	20_32	ARS-BFGL-NGS-18037	rs109621511	0.2459	0.969
26	338	ARS-BFGL-NGS-115249	ARS-BFGL-NGS-4915	10	0.43	22.75	3_42	Hapmap57990-rs29018907	rs29018907	0.2142	0.965
27	1636	BTB-00604291	Hapmap41245-BTA-36880	20	0.41	23.16	15_52	ARS-BFGL-NGS-5740	rs109781293	0.1753	0.953
28	2173	ARS-BFGL-NGS-44232	ARS-BFGL-NGS-40677	15	0.39	23.56	23_10	ARS-BFGL-NGS-94682	rs109474253	0.2038	0.956
29	1937	BTB-00754259	ARS-BFGL-NGS-32338	24	0.38	23.94	19_45	ARS-BFGL-NGS-36780	rs109772260	0.132	0.947
30	2043	ARS-BFGL-NGS-101699	ARS-BFGL-NGS-31918	14	0.38	24.31	21_14	Hapmap39722-BTA-119704	rs41619364	0.1935	0.96
31	1439	ARS-BFGL-NGS-79011	ARS-BFGL-NGS-113296	15	0.38	24.69	13_24	ARS-BFGL-NGS-79011	rs109161961	0.185	0.952
32	2403	ARS-BFGL-NGS-58109	ARS-BFGL-NGS-40430	10	0.36	25.05	27_29	Hapmap27965-BTA-148364	rs109809156	0.2206	0.968

											Hapmap35547-	
33	1637	ARS-BFGL-NGS-59460	BTA-37420-no-rs	11	0.36	25.41	15_53	SCAFFOLD316738_25851	rs29017039	0.1972	0.959	
34	580	ARS-BFGL-NGS-19252	ARS-BFGL-NGS-28318	13	0.35	25.76	5_41	BTA-118386-no-rs	rs110853898	0.1607	0.951	
35	2196	UA-IFASA-8452	ARS-BFGL-NGS-107881	16	0.35	26.11	23_33	ARS-BFGL-NGS-107881	rs109129418	0.176	0.954	
36	1494	ARS-BFGL-NGS-22500	ARS-BFGL-NGS-25698	22	0.35	26.47	13_79	ARS-BFGL-NGS-112386	rs110710210	0.2045	0.961	
37	1293	ARS-BFGL-NGS-19413	Hapmap25799-BTA-126762	17	0.35	26.81	11_75	ARS-BFGL-NGS-110868	rs110616006	0.1549	0.95	
38	1527	ARS-BFGL-NGS-19050	BTB-00561102	21	0.34	27.15	14_28	UA-IFASA-8135	rs41579042	0.1254	0.942	
39	2281	Hapmap24734-BTC-016033	ARS-BFGL-NGS-10250	19	0.33	27.48	25_2	Hapmap24734-BTC-016033	rs110546422	0.0964	0.929	
40	2220	Hapmap33715-BTA-137579	BTB-01921336	19	0.33	27.82	24_4	BTB-00878429	rs42042153	0.1592	0.947	
41	1226	BTB-00455305	ARS-BFGL-NGS-110878	21	0.33	28.15	11_8	ARS-BFGL-NGS-25292	rs43659393	0.1128	0.939	
42	1226	BTB-00774691	ARS-BFGL-NGS-110878	21	0.33	28.15	11_8	Hapmap44582-BTA-107322	rs41616683	0.0699	0.927	
43	1975	BTA-00128-rs29013292	BTA-110421-no-rs	22	0.32	28.47	20_18	BTA-110421-no-rs	rs43711314	0.1149	0.937	
44	246	ARS-BFGL-NGS-55270	Hapmap41164-BTA-116565	11	0.32	29.12	2_87	ARS-BFGL-NGS-91673	rs110192736	0.1415	0.945	
45	2436	Hapmap51476-BTA-63589	BTA-20879-no-rs	23	0.32	29.44	28_16	Hapmap42306-BTA-73982	rs41649017	0.1385	0.946	
											Hapmap36588-	
46	1959	ARS-BFGL-NGS-105473	ARS-BFGL-BAC-27938	21	0.32	29.75	20_2	SCAFFOLD90561_9460	rs29023669	0.1731	0.953	
47	2409	ARS-BFGL-NGS-116606	ARS-BFGL-NGS-9378	17	0.3	30.37	27_35	Hapmap46767-BTA-117224	rs41570143	0.1672	0.953	
48	434	Hapmap42802-BTA-109751	Hapmap60832-rs29020531	18	0.28	30.65	4_16	Hapmap60832-rs29020531	rs29020531	0.1545	0.946	
49	1528	Hapmap33770-BTA-104911	ARS-BFGL-NGS-2627	18	0.28	30.93	14_29	Hapmap49073-BTA-18180	rs41570273	0.0804	0.93	
50	1320	Hapmap49554-BTA-28269	ARS-BFGL-NGS-119318	22	0.27	31.19	11_102	ARS-BFGL-NGS-119318	rs110996839	0.1601	0.943	
51	845	ARS-BFGL-NGS-76019	BTB-00316348	21	0.26	31.45	7_64	BTB-00316348	rs43523199	0.1294	0.946	
52	1309	ARS-BFGL-NGS-109962	BTA-109943-no-rs	16	0.26	31.71	11_91	ARS-BFGL-BAC-14322	rs110573334	0.1509	0.95	
53	1827	ARS-BFGL-NGS-30162	ARS-BFGL-BAC-36904	13	0.26	31.97	18_1	ARS-BFGL-BAC-36904	rs110491665	0.1208	0.943	
54	1290	ARS-BFGL-NGS-57978	ARS-BFGL-NGS-74702	15	0.24	32.21	11_72	Hapmap42771-BTA-104230	rs41610337	0.0583	0.917	
55	2292	Hapmap42771-BTA-104230	Hapmap29990-BTA-138311	13	0.24	32.45	25_13	Hapmap29990-BTA-138311	rs109142685	0.0691	0.926	
56	2237	ARS-BFGL-BAC-2989	ARS-BFGL-NGS-10227	19	0.24	32.69	24_21	ARS-BFGL-NGS-14384	rs109827879	0.1489	0.943	
57	870	Hapmap42845-BTA-121426	Hapmap49872-BTA-115580	19	0.24	32.93	7_89	ARS-BFGL-NGS-28876	rs109938490	0.1307	0.946	
58	1517	ARS-BFGL-NGS-116915	ARS-BFGL-NGS-93803	19	0.24	33.4	14_18	ARS-BFGL-NGS-103663	rs109480189	0.1118	0.937	
59	1960	ARS-BFGL-NGS-117833	ARS-BFGL-NGS-116843	17	0.24	33.64	20_3	ARS-BFGL-NGS-116843	rs110627799	0.1072	0.929	
60	840	BTB-00553442	Hapmap23996-BTA-145153	21	0.23	33.87	7_59	ARS-BFGL-NGS-113745	rs110059512	0.1195	0.942	
61	779	ARS-BFGL-NGS-57239	BTB-00288047	18	0.23	34.1	6_118	ARS-BFGL-NGS-119881	rs109238890	0.1302	0.943	
62	2049	Hapmap48553-BTA-10000	Hapmap44836-BTA-51861	14	0.23	34.33	21_20	ARS-BFGL-NGS-7759	rs109222589	0.1013	0.938	
63	474	ARS-BFGL-NGS-17985	BTA-70690-no-rs	17	0.22	34.55	4_56	Hapmap54543-rs29011993	rs29011993	0.1125	0.94	
64	1597	ARS-BFGL-NGS-7759	Hapmap38637-BTA-88156	15	0.22	34.77	15_13	Hapmap38637-BTA-88156	rs41596665	0.1446	0.947	
65	1013	BTB-00186413	ARS-BFGL-NGS-105604	19	0.22	34.99	9_5	BTB-00379074	rs43588162	0.1011	0.937	

			Hapmap36813-									
66	694	BTB-00581380	SCAFFOLD50174_9004	17	0.22	35.21	6_33	Hapmap32447-BTC-033214	rs110349550	0.1379	0.941	
67	545	BTB-00379161	ARS-BFGL-NGS-39595	21	0.22	35.43	5_6	BTB-00218064	rs43425658	0.0899	0.931	
68	717	Hapmap33286-BTC-032290	BTA-26162-no-rs	20	0.21	35.64	6_56	BTB-01132969	rs42287981	0.0903	0.933	
69	1774	BTB-01610675	ARS-BFGL-NGS-25840	7	0.21	36.06	17_23	Hapmap51950-BTA-40650	rs41584487	0.1417	0.948	
70	2109	ARS-BFGL-NGS-98343	BTA-09886-rs29026042	18	0.21	36.26	22_8	BTB-01743103	rs42857922	0.0876	0.923	
71	1607	Hapmap50656-BTA-31918	Hapmap27194-BTA-153449	28	0.21	36.47	15_23	ARS-BFGL-NGS-89413	rs110552464	0.06	0.927	
72	1012	BTA-40659-no-rs	BTB-01696182	18	0.2	36.67	9_4	BTB-02033050	rs43140139	0.0798	0.924	
73	1845	Hapmap42824-BTA-116676	BTA-94677-no-rs	16	0.2	36.88	18_19	ARS-BFGL-NGS-12314	rs110424623	0.1166	0.942	
74	1463	ARS-BFGL-NGS-73400	BTB-01186854	14	0.2	37.08	13_48	ARS-BFGL-NGS-43431	rs110844879	0.0583	0.924	

Anexo 4. Janelas de maior representatividade da variância genética (% Var) juntamente com seu SNP de maior *Model Frequency* e *t-like* para a característica de ganho de peso pós-desmama ajustado para os 345 dias em bovinos Hereford e Braford.

Obs	Window	Start SNP Name	End SNP Name	N SNPs	% Var	Cum.Var	chr_Mb	Top SNP Name	Ref Top SNP	ModelFreq	t.like
1	262	ARS-BFGL-NGS-17875	BTA-48084-no-rs	11	2.15	2.15	2_103	Hapmap50215-BTA-111911	rs41567611	0.6028	1.159
2	379	BTB-00139975	Hapmap49270-BTA-68543	19	1.8	3.95	3_83	BTA-68532-no-rs	rs43709424	0.4885	1.071
3	1575	UA-IFASA-7973	Hapmap43459-BTA-111191	13	1.58	5.53	14_76	UA-IFASA-7973	rs41580472	0.5025	1.082
4	1676	BTB-01975868	Hapmap33722-BTA-155362	9	1.35	6.88	16_6	BTB-02010595	rs43116743	0.4293	1.026
5	308	ARS-BFGL-NGS-36635	ARS-BFGL-NGS-20009	8	1.26	8.14	3_12	Hapmap60323-rs29014816	rs29014816	0.3262	0.99
6	779	ARS-BFGL-NGS-17985	BTB-00288047	18	1.02	9.17	6_118	ARS-BFGL-NGS-119881	rs109238890	0.3463	1.001
7	2126	Hapmap50387-BTA-53912	Hapmap32083-BTA-147240	12	0.99	10.15	22_25	BTA-53914-no-rs	rs41583553	0.2382	0.967
8	2338	BTB-01619101	ARS-BFGL-NGS-48955	15	0.95	11.1	26_16	BTA-121474-no-rs	rs41621632	0.3142	0.987
9	1520	ARS-BFGL-NGS-6136	Hapmap27177-BTA-147405	20	0.89	11.99	14_21	ARS-BFGL-NGS-37279	rs109065397	0.303	0.981
		Hapmap35931-									
10	1418	SCAFFOLD200024_14429	ARS-BFGL-NGS-105636	15	0.82	12.8	13_3	ARS-BFGL-NGS-62490	rs110099559	0.2369	0.961
11	503	ARS-BFGL-NGS-107560	BTB-01257567	20	0.74	13.55	4_85	BTB-01739497	rs42850841	0.1299	0.938
12	2179	ARS-BFGL-NGS-76846	ARS-BFGL-NGS-114979	15	0.65	14.19	23_16	Hapmap46306-BTA-55660	rs41640779	0.1444	0.94
13	2049	ARS-BFGL-NGS-7759	Hapmap44836-BTA-51861	14	0.63	14.82	21_20	ARS-BFGL-NGS-7759	rs109222589	0.1555	0.943
14	2490	ARS-BFGL-NGS-60509	BTB-00935033	16	0.58	15.4	29_23	BTB-01012992	rs42175089	0.2534	0.959
15	1961	ARS-BFGL-NGS-20128	BTB-01988115	23	0.54	15.94	20_4	BTB-00144185	rs43349755	0.0812	0.919
16	875	Hapmap40829-BTA-88244	Hapmap43946-BTA-80204	12	0.49	16.43	7_94	BTA-80207-no-rs	rs41656976	0.2319	0.961
17	843	ARS-BFGL-NGS-113598	ARS-BFGL-NGS-26618	20	0.48	16.91	7_62	ARS-BFGL-NGS-113598	rs42809616	0.1544	0.939
18	1366	ARS-BFGL-NGS-110832	Hapmap51894-BTA-29431	8	0.48	17.38	12_40	Hapmap57051-rs29018418	rs29018418	0.1954	0.953
19	501	ARS-BFGL-NGS-117077	Hapmap60682-rs29013346	19	0.46	17.84	4_83	BTB-01261088	rs42233703	0.11	0.927
20	272	ARS-BFGL-NGS-12225	Hapmap48541-BTA-97822	17	0.44	18.28	2_113	Hapmap48541-BTA-97822	rs41666711	0.1451	0.94
21	471	BTB-01065788	BTB-00188617	22	0.42	18.71	4_53	ARS-BFGL-NGS-30090	rs109476769	0.1457	0.938
22	1606	ARS-BFGL-NGS-104925	BTB-00583912	21	0.4	19.11	15_22	BTB-01512525	rs42632232	0.1546	0.922
23	2403	ARS-BFGL-NGS-58109	ARS-BFGL-NGS-40430	10	0.38	19.49	27_29	Hapmap27965-BTA-148364	rs109809156	0.179	0.947
24	1960	ARS-BFGL-NGS-57239	ARS-BFGL-NGS-116843	17	0.38	19.87	20_3	ARS-BFGL-NGS-116843	rs110627799	0.1573	0.935
25	787	Hapmap49762-BTA-80425	ARS-BFGL-NGS-10823	20	0.37	20.24	7_6	ARS-BFGL-NGS-91225	rs110274027	0.1358	0.936
26	1451	BTA-32100-no-rs	ARS-BFGL-NGS-35710	20	0.34	20.57	13_36	ARS-BFGL-NGS-14452	rs110768183	0.1266	0.935
27	1524	Hapmap41234-BTA-34285	ARS-BFGL-NGS-8308	18	0.33	20.9	14_25	BTB-01144089	rs42299127	0.1189	0.932
28	1017	ARS-BFGL-NGS-117804	ARS-BFGL-NGS-5339	9	0.32	21.23	9_9	ARS-BFGL-NGS-117804	rs110649736	0.1554	0.943
29	1462	ARS-BFGL-NGS-32053	ARS-BFGL-NGS-322	21	0.31	21.54	13_47	Hapmap49571-BTA-32781	rs41585336	0.1436	0.935
30	2362	Hapmap24081-BTA-139000	ARS-BFGL-NGS-118736	9	0.31	21.85	26_40	DPI-59	rs110935890	0.1227	0.93
31	1939	ARS-BFGL-NGS-116606	UA-IFASA-8219	20	0.3	22.15	19_47	ARS-BFGL-NGS-117004	rs109932054	0.0524	0.91
32	455	Hapmap47073-BTA-66626	Hapmap44397-BTA-70124	20	0.28	22.43	4_37	BTA-70113-no-rs	rs41586860	0.1116	0.92

33	841	BTB-00313206	Hapmap22864-BTA-145165	18	0.28	22.71	7_60	BTB-00312832	rs43516489	0.0946	0.928
34	645	ARS-BFGL-NGS-118977	ARS-BFGL-NGS-2465	16	0.28	22.99	5_106	ARS-BFGL-NGS-39379	rs110421124	0.0893	0.917
35	4	BTA-34568-no-rs	BTB-01159977	14	0.27	23.26	1_4	Hapmap60540-rs29021339	rs29021339	0.1292	0.938
36	1239	Hapmap43962-BTA-86597	Hapmap42711-BTA-87541	24	0.27	23.54	11_21	ARS-BFGL-BAC-11047	rs110245673	0.0667	0.91
37	2224	BTB-01414148	ARS-BFGL-NGS-100431	14	0.26	23.79	24_8	BTB-01910403	rs43015221	0.1187	0.932
38	2254	ARS-BFGL-NGS-112380	BTB-01621818	22	0.26	24.05	24_38	BTB-00888968	rs42050817	0.1349	0.935
39	664	Hapmap30828-BTA-143720	Hapmap30881-BTA-159706	23	0.26	24.31	6_3	Hapmap53408-rs29014449	rs29014449	0.1088	0.93
40	1443	ARS-BFGL-NGS-30195	ARS-BFGL-NGS-108308	19	0.26	24.57	13_28	ARS-BFGL-NGS-107316	rs109262940	0.0712	0.917
41	2361	ARS-BFGL-NGS-1051	ARS-BFGL-NGS-103038	21	0.25	24.82	26_39	ARS-BFGL-NGS-115002	rs110093934	0.0932	0.916
42	2315	ARS-BFGL-NGS-10694	BTA-60128-no-rs	23	0.25	25.07	25_36	ARS-BFGL-NGS-4979	rs109618153	0.0945	0.922
43	756	BTA-01561-rs29012522	Hapmap41966-BTA-77379	24	0.25	25.32	6_95	BTB-00753045	rs41911450	0.0755	0.917
44	2048	BTB-00806795	Hapmap47651-BTA-51810	27	0.24	25.56	21_19	Hapmap47651-BTA-51810	rs41640647	0.0775	0.919
45	2469	ARS-BFGL-NGS-39483	BTB-00998633	23	0.24	25.8	29_2	ARS-BFGL-NGS-119674	rs42156592	0.0474	0.907
46	859	Hapmap43998-BTA-107878	ARS-BFGL-NGS-9914	14	0.23	26.03	7_78	BTA-23130-no-rs	rs41578538	0.1121	0.931
47	1697	Hapmap60671-rs29023031	ARS-BFGL-BAC-33348	12	0.23	26.27	16_28	Hapmap40450-BTA-61265	rs41606763	0.1318	0.922
48	1698	Hapmap38469-BTA-12042 Hapmap34759-	BTA-23170-no-rs	10	0.23	26.5	16_29	Hapmap44646-BTA-38361	rs41635148	0.0713	0.919
49	569	BES10_Contig780_1565	ARS-BFGL-NGS-23612	13	0.22	26.72	5_30	ARS-BFGL-NGS-23612	rs110445443	0.0523	0.909
50	551	ARS-BFGL-NGS-26310	Hapmap38799-BTA-101880	10	0.22	26.94	5_12	ARS-BFGL-NGS-29426	rs110162221	0.0726	0.919
51	317	ARS-BFGL-NGS-97101	Hapmap23289-BTA-66836	17	0.22	27.16	3_21	Hapmap59143-ss46526687	rs41257141	0.0637	0.912
52	845	ARS-BFGL-NGS-109962	BTB-00316348	21	0.21	27.37	7_64	BTB-00316348	rs43523199	0.1074	0.931
53	1150	ARS-BFGL-NGS-4688	BTA-01320-no-rs	15	0.21	27.58	10_37	ARS-BFGL-NGS-95080	rs110141019	0.0875	0.924
54	5	Hapmap44178-BTA-101248	BTB-01489336	23	0.21	27.79	1_5	BTB-00003652	rs43210840	0.0683	0.917
55	704	BTB-01468535	BTB-01893222	23	0.21	28	6_43	ARS-BFGL-NGS-58762	rs110635310	0.0725	0.919
56	2151	ARS-BFGL-NGS-31703	BTB-00851968	12	0.21	28.21	22_50	ARS-BFGL-NGS-15965	rs110485899	0.0983	0.916
57	572	ARS-BFGL-NGS-110548	BTB-01495858	25	0.2	28.41	5_33	BTB-01226666	rs42384792	0.0522	0.89
58	352	BTA-118287-no-rs	ARS-BFGL-NGS-119366	10	0.2	28.61	3_56	Hapmap39892-BTA-118279	rs41571698	0.084	0.909