

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

**METODOLOGIA DE ATRIBUIÇÃO DOS ESCORES  
VISUAIS E ESTUDO DE ASSOCIAÇÃO GENÔMICA  
PARA OS ESCORES DE CONFORMAÇÃO,  
PRECOCIDADE E MUSCULOSIDADE EM BOVINOS  
NELORE**

**Luis Orlando Duitama Carreño**

Zootecnista

**2015**



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

**METODOLOGIA DE ATRIBUIÇÃO DOS ESCORES  
VISUAIS E ESTUDO DE ASSOCIAÇÃO GENÔMICA  
PARA OS ESCORES DE CONFORMAÇÃO,  
PRECOCIDADE E MUSCULOSIDADE EM BOVINOS  
NELORE**

**Luis Orlando Duitama Carreño**

**Orientador: Prof. Dr. Ricardo da Fonseca**

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

**2015**

Duitama-Carreño, Luis Orlando  
D875m Metodologia de atribuição dos escores visuais e estudo de associação genômica para os escores de conformação, precocidade e musculosidade em bovinos Nelore / Luis Orlando Duitama Carreño. -- Jaboticabal, 2015  
xi, 51 p.; 28cm

Tese (Doutorado) – Universidade Estadual Paulista, Faculdade de Ciências Agrárias e Veterinárias, 2015.

Orientador: Ricardo da Fonseca

Banca examinadora: Henrique Nunes de Oliveira, Aldrin Viera Pires, Roberto Carvalheiro, Francisco Ribeiro de Araujo Neto.

Bibliografia

1. BayesC. 2. GWAS. 3. LASSO Bayesiano. 4. Modelo de limiar. 5. Simulação. I. Título. II. Jaboticabal-Faculdade de Ciências Agrárias e Veterinárias.

CDU 636.082:636.2

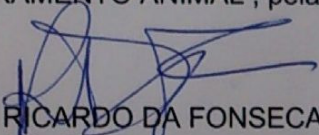
CERTIFICADO DE APROVAÇÃO

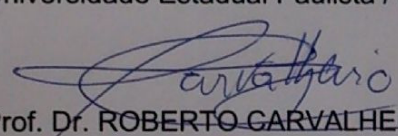
**TÍTULO:** METODOLOGIA DE ATRIBUIÇÃO DOS ESCORES VISUAIS E ESTUDO DE ASSOCIAÇÃO GENÔMICA PARA OS ESCORES DE CONFORMAÇÃO, PRECOCIDADE E MUSCULOSIDADE EM BOVINOS NELORE

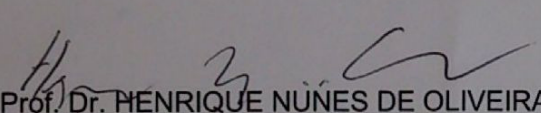
**AUTOR:** LUIS ORLANDO DUITAMA CARREÑO

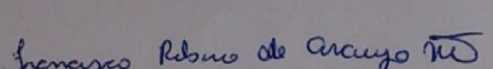
**ORIENTADOR:** Prof. Dr. RICARDO DA FONSECA

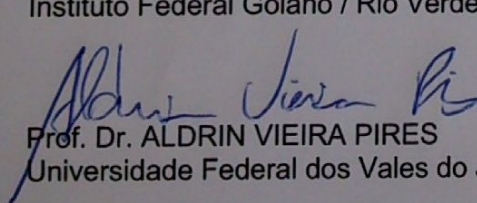
Aprovado como parte das exigências para obtenção do Título de DOUTOR EM GENÉTICA E MELHORAMENTO ANIMAL, pela Comissão Examinadora:

  
Prof. Dr. RICARDO DA FONSECA  
Universidade Estadual Paulista / Dracena/SP

  
Prof. Dr. ROBERTO CARVALHEIRO  
Departamento de Zootecnia / Faculdade de Ciências Agrárias e Veterinárias de Jaboticabal

  
Prof. Dr. HENRIQUE NUNES DE OLIVEIRA  
Departamento de Zootecnia / Faculdade de Ciências Agrárias e Veterinárias de Jaboticabal

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

  
Prof. Dr. ALDRIN VIEIRA PIRES  
Universidade Federal dos Vales do Jequitinhonha e Mucuri / Diamantina/MG

Data da realização: 12 de março de 2015.



## **DADOS CURRICULARES DO AUTOR**

Luis Orlando Duitama Carreño, nascido em 08 de fevereiro de 1982, na cidade de Bogotá – Colômbia, filho de Luis Alberto Duitama e Ana Isabel Carreño. Iniciou em fevereiro de 2000 o curso de graduação em Zootecnia na “Universidad Nacional de Colombia”, obtendo o título de Zootecnista em agosto de 2005. Em março de 2009 ingressou no Programa de Pós-graduação em Genética e Melhoramento Animal na Faculdade de Ciências Agrárias e Veterinárias, UNESP, Campus de Jaboticabal – SP, obtendo o grau de mestre em fevereiro de 2011, sob orientação do prof. Dr. Ricardo da Fonseca. Em março de 2011, iniciou o curso de Doutorado no programa de Pós-Graduação em Genética e Melhoramento Animal na Faculdade de Ciências Agrárias e Veterinárias (UNESP), Jaboticabal - SP, sob orientação da prof. Dr. Ricardo da Fonseca. De setembro de 2014 à fevereiro de 2015, realizou estágio de Doutorado no exterior na “Universidad Politécnica de Valencia”, sob orientação do prof. Dr. Agustín Blasco.





*El secreto de la felicidad no esta en hacer siempre lo que se quiere, sino en querer siempre lo que se hace*

*León Tolstói*



*Aos meus queridos pais, Ana Isabel e Luis Alberto.  
A meus irmãos Milena, Omar, Teresa, Ana, Cecilia e Joaquin.*

*Dedico e Ofereço.*



## **AGRADECIMENTOS**

A Deus, por iluminar a minha caminhada.

Aos meus pais, pela educação, amor e exemplo de vida.

A minha família, especialmente a meus irmãos Omar e Milena pelo apoio, amor e conselhos.

Ao Prof. Dr. Ricardo da Fonseca, pela orientação, amizade e confiança.

Aos membros da banca examinadora, Prof. Dr. Henrique Nunes de Oliveira, Prof. Dr. Aldrin Viera Pire, Prof. Dr. Roberto Carvalheiro e Prof. Dr. Francisco Ribeiro de Araujo Neto pelos importantes questionamentos e sugestões.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela concessão da bolsa de doutorado, processo 5819/10-5.

Ao Prof. Dr. Agustín Blasco, pelos ensinamentos e conselhos durante o estágio na Universidade Politécnica de Valencia.

Ao programa BABEL *Erasmus Mundus* pela concessão da bolsa de doutorado sanduíche.

Aos companheiros e amigos do LuCCA-Z, Rafael Keith, Adam, Michel, Tássia, Matilde, Diogo, Michele e Ligia pela amizade, aprendizado e momentos de descontração.

Aos companheiros da república “K-baret” em Dracena, pelos momentos de descontração e amizade.

À UNESP e ao Programa de Pós-graduação em Genética e Melhoramento Animal, pela oportunidade de crescimento científico que me proporcionaram.

Aos meus compatriotas que moram ou moraram em Jaboticabal, Anny, Luis Gabriel, Javier, Astrid, Donicer, Diana, Edna, Yury, .... pela amizade e apoio nos momentos difíceis.

E a todos aqueles que de alguma forma contribuíram para minha formação pessoal e profissional.



## SUMÁRIO

SUMÁRIO . . . . .	ix
CAPÍTULO 1 CONSIDERAÇÕES GERAIS . . . . .	1
1.1 Escores visuais . . . . .	1
1.2 Estudos de associação genômica GWAS . . . . .	3
1.2.1 Regressão por marcador . . . . .	4
1.2.2 Regressão por haplótipos . . . . .	5
1.2.3 Regressão múltipla para os marcadores . . . . .	6
1.2.3.1 Regressão “stepwise” . . . . .	6
1.2.3.2 Regressão Bayesiana . . . . .	7
1.2.3.3 Machine learning . . . . .	8
1.3 Objetivos . . . . .	9
REFERÊNCIA . . . . .	10
CAPÍTULO 2 ESTIMAÇÃO DE PARÂMETROS E VALORES GENÉTICOS EM CARACTERÍSTICAS DE ESCORES VISUAIS: ESTUDO DE SI- MULAÇÃO . . . . .	17
2.1 Introdução . . . . .	18
2.2 Material e métodos . . . . .	19
2.3 Resultados e discussão . . . . .	23
2.4 Conclusões . . . . .	28
REFERÊNCIA . . . . .	29
CAPÍTULO 3 ESTUDO DE ASSOCIAÇÃO GENÔMICA PARA ESCORES VI- SUAIS MENSURADOS À DESMAMA EM BOVINOS NELORE . . . . .	33
3.1 Introdução . . . . .	34
3.2 Material e métodos . . . . .	35
3.2.1 Dados genotípicos . . . . .	36
3.2.2 Análises de associação . . . . .	37
3.2.3 Critério de associação . . . . .	38
3.2.4 Identificação dos genes . . . . .	39
3.3 Resultados e discussão . . . . .	39
3.4 Conclusões . . . . .	46
REFERÊNCIAS . . . . .	47

## **METODOLOGIA DE ATRIBUIÇÃO DOS ESCORES VISUAIS E ESTUDO DE ASSOCIAÇÃO GENÔMICA PARA OS ESCORES DE CONFORMAÇÃO, PRECOCIDADE E MUSCULOSIDADE EM BOVINOS NELORE**

**RESUMO-** Os escores visuais são usados como critério de seleção com o objetivo de melhorar as características de carcaça. Estas características têm a particularidade de serem atribuídas com base em uma referência relativa, e não absoluta, o que pode trazer dificuldades na estimação de parâmetros e valores genéticos. Dada a importância dos escores, dois estudos foram conduzidos com os seguintes objetivos: 1) determinar se a forma de atribuir os escores, tem consequências na estimação de parâmetros e valores genéticos; e 2) identificar as regiões do genoma associadas com os escores de conformação, precocidade e musculosidade em bovinos da raça Nelore. Para o primeiro objetivo foi realizado um estudo de simulação, considerando populações com e sem seleção e três tamanhos de grupo de atribuição dos escores (10, 40 e 100 animais), as populações foram acasaladas ao longo de 12 gerações sobrepostas. A cada geração parâmetros e valores genéticos foram estimados usando um modelo de limiar. Os resultados indicaram que as estimativas de parâmetros e valores genéticos não são afetadas pela forma de atribuir os escores, entretanto os valores genéticos estimados são afetados pelo tamanho do grupo, sendo desejável a formação de grupos de avaliação superiores a 40 animais. Para o segundo objetivo foi realizado um GWAS usando dois modelos, BayesC e LASSO Bayesiano com duas classes de variáveis dependentes, o fenótipo dos escores e os valores genéticos desregredidos (dEBV). Nas análises foram utilizados 2873 registros de machos e fêmeas Nelore. Os animais foram genotipados com o painel BovineHD da Illumina, após o controle de qualidade restaram 309.865 SNPs. O critério de associação foi a porcentagem de variância genética explicada por janelas de 1 Mb de comprimento. Após as análises, o modelo BayesC foi o que se ajustou melhor aos dados, porque explicou uma maior proporção de variância fenotípica para os três escores nas duas variáveis dependentes. Os resultados mostraram que os escores são de natureza poligênica, uma vez que não houve janelas explicando grandes porcentagens de variância genética, no entanto, nos cromossomos 1, 3, 5, 7, 14, 15, 16, 20 e 23 foram identificadas as regiões que explicaram a maior porcentagem de variância genética para conformação, precocidade e musculosidade.

**Palavras-chave:** BayesC, GWAS, LASSO Bayesiano, modelo de limiar, parâmetros genéticos, simulação



## ATTRIBUTION METHODOLOGY OF VISUAL SCORES AND WHOLE-GENOME ASSOCIATION STUDY FOR CONFORMATION, FINISHING PRECOCITY AND MUSCLING VISUAL SCORES IN NELLORE CATTLE

**ABSTRACT-** The visual scores are used as selection criteria in Brazilian cattle and aiming to improve the carcass traits. These traits have the particularity of being attributed depending on relative data and not on absolute references. This property could lead to difficulties in the estimation of genetic parameters and breeding values as well. Considering the importance of the visual scores, two studies were performed with the following objectives: 1) determine if the forms in which these scores are attributed have consequences in the genetic parameters and breeding values estimation, and, 2) identify regions of the genome that are associated with the scores of conformation, finishing precocity and muscling in Nelore cattle. For the first objective, a simulation study was conducted considering populations with and without selection, and three sizes of group attribution scores (10, 40 and 100 animals), these populations were mated for 12 overlapping generations. Genetic parameters and breeding values for each generation were estimated using a threshold model. The results indicate that the estimates of genetic parameters and breeding values are not affected by how the scores are attributed; however, the estimated breeding values are affected by the group size, being recommended the formation of evaluation groups more than 40 animals. For the second objective, a GWAS was conducted using BayesC and Bayesian LASSO models with two types of dependent variable. These variables were the phenotype and the deregressed breeding values (dEBV). In the analyses, a total of 2873 records of males and females from Nelore cattle were used. The animals were genotyped using the chip BovineHD of Illumina, after the quality control, a total of 309,865 SNPs were maintained. The association criterion was the proportion of genetic variance explained by 1 Mb genome windows. After the analyses, the BayesC model was the best fitted the data, as it explained a greater proportion of the phenotypic variance for the three scores in the two dependent variables. The results show that the visual scores have a polygenic nature, given that there were not genome windows explaining large percentages of the genetic variance. However, there were identified regions on the 1, 3, 5, 7, 14, 15, 16, 20 and 23 chromosomes that may explain the highest proportion of genetic variance for conformation, precocity and muscling.

**keywords:** BayesC, Bayesian LASSO, genetic parameters, GWAS, threshold model, simulation



## **CAPÍTULO 1 - CONSIDERAÇÕES GERAIS**

### **1.1 Escores visuais**

Os primeiros esforços em melhoramento genético na bovinocultura começaram com a avaliação visual, tentando associar biótipos morfológicos com animais mais produtivos, considerando-se o olho humano como a ferramenta capaz de identificar os animais superiores para seu posterior acasalamento (BERG; BUTTERFIELD, 1976). Nos dias atuais, a avaliação visual continua sendo utilizada, porém de uma forma mais padronizada, mediante a implementação de metodologias de avaliação visual nos programas de melhoramento genético animal.

No Brasil, as avaliações visuais em programas de melhoramento genético de bovinos, começaram a ser formalmente utilizadas a partir de 1974 no PROMEBO (programa de melhoramento de bovinos) (FRIES, 1996), que exigia, além das pesagens às idades padrão, a avaliação visual dos animais com base em duas metodologias: escore de conformação (EC) do USDA (Departamento de Agricultura dos Estados Unidos) (BIF, 1974), e o sistema de avaliação Ankony (LONG, 1971).

A incorporação de medidas de escores visuais nos programas de melhoramento genético buscam auxiliar a seleção de animais mais eficientes, sem que seja limitada somente às medidas de peso em diferentes idades (MAIWASHE et al., 2002; AFOLAYAN et al., 2007; KOURY FILHO et al., 2010). Os escores objetivam estimar a proporção dos diferentes tecidos da carcaça (músculo, osso e gordura) dos animais no momento do abate, bem como identificar os animais de acabamento precoce (quantidade de gordura mínima necessária para o abate), ajudando a melhorar as características de carcaça (JÚNIOR et al., 2001).

Para atingir o objetivo, diferentes metodologias de avaliação dos escores visuais têm sido propostas. Inicialmente, eram atribuídos valores inteiros com base em uma referência absoluta, como é o caso das características lineares de tipo em gado de leite (ZINK; ŠTÍPKOVÁ; LASSEN, 2011). Posteriormente, Fries (1996) propôs que os escores fossem atribuídos com base em uma referência relativa, neste caso o grupo de animais avaliados seria o referencial, sendo que o escore atribuído a cada animal dependeria da média do grupo, e não mais de uma referência absoluta. Esse procedimento objetiva aumentar a variabilidade fenotípica nas avaliações, visto que com a metodologia anterior a distribuição dos fenótipos era concentrada em torno de poucos escores.

A avaliação das características de carcaça por meio de escores visuais apresenta várias vantagens: facilidade, ao permitir avaliar um grande número de animais

em pouco tempo; baixo custo, porque é feito por avaliadores treinados sem o uso de equipamentos; e finalmente, não é invasiva nem deixa resíduos no animal, viabilizando a inclusão dos escores nos programas de melhoramento genético. Por outro lado, também apresenta desvantagens que deixam dúvida, se realmente ajudam a melhorar a qualidade de carcaça. A principal desvantagem é a baixa correlação com características de qualidade de carcaça. A maioria dos trabalhos mostram baixas correlações fenotípicas (BERG; BUTTERFIELD, 1976; CONROY et al., 2009), principalmente quando os escores são avaliados em idades que não são próximas do abate. Para as correlações genéticas, os resultados têm sido semelhantes (FARIA et al., 2009; YOKOO et al., 2009; GORDO et al., 2012; GORDO, 2014). Outra crítica importante é a subjetividade das avaliações, uma vez que ao serem medidas atribuídas através da interpretação do olho humano, erros podem ser produzidos.

Parâmetros genéticos para escores visuais e correlações genéticas com outras características de importância econômica, foram estimados em várias populações de bovinos no Brasil (FARIA et al., 2008b; FARIA et al., 2008a; KOURY FILHO et al., 2009; SHIOTSUKI et al., 2009; YOKOO et al., 2009; BARICHELLO et al., 2010; KOURY FILHO et al., 2010; GORDO et al., 2012). Os resultados mostram que existe suficiente variabilidade genética que permite progressos genéticos por seleção, uma vez que apresentam estimativas de herdabilidade moderadas, variando de 0,1 até 0,65. As correlações genéticas com características de crescimento apresentam-se altas, especialmente com os pesos corporais, isto porque, a avaliação visual leva em consideração o tamanho do animal e o desenvolvimento das massas musculares. O objetivo principal dos escores visuais é melhorar as características de carcaça, no entanto, as correlações genéticas se apresentam baixas a moderadas, principalmente para a espessura de gordura. Para a área de olho de lombo as correlações genéticas foram maiores e com grande variabilidade nos estudos, variando de 0,11 até 0,91 (FARIA et al., 2009; YOKOO et al., 2009; GORDO et al., 2012). Apesar da literatura indicar que nem sempre os escores são preditores adequados da qualidade de carcaça, eles continuam sendo incluídos nos programas de melhoramento genético, porque auxiliam a seleção de animais mais harmoniosos e produtivos.

A natureza do fenótipo dos escores visuais é categórico, variando no número de níveis dependendo da metodologia empregada na avaliação. Na maioria dos trabalhos essa natureza não é levada em consideração e é tratada como uma característica contínua (KOURY FILHO et al., 2009; YOKOO et al., 2009; BARICHELLO et al., 2010; KOURY FILHO et al., 2010; GORDO et al., 2012), o que pode trazer erros de estimação, porque algumas propriedades do BLUP não são mantidas, como a habilidade de maximizar o ranqueamento dos animais (FERNANDO; BILLINGSLEY; GIANOLA,

1983), também pode trazer problemas na estimação da variância residual (GIANOLA, 1982). Faria et al. (2008a) compararam os modelos de limiar e linear na estimação de parâmetros genéticos e componentes de variância para escores visuais e concluíram que os dois modelos foram próximos nas estimativas, no entanto, as propriedades estatísticas do modelo de limiar permitem estimações mais robustas.

Como descrito por Koury Filho et al. (2009) os escores são atribuídos com base em uma referência relativa, que neste caso é a média do grupo de animais avaliado, o que não permite que animais com o mesmo escore, pertencentes a diferentes grupos de avaliação, possam ser comparados fenotipicamente, o que pode trazer dificuldades na estimação de parâmetros e valores genéticos. Duitama et al. (2014) em um estudo de simulação, estimaram parâmetros e valores genéticos para escores visuais e encontraram subestimação em populações sob seleção.

## **1.2 Estudos de associação genômica GWAS**

Os estudos de associação genômica ampla (GWAS) ganharam importância rapidamente nos últimos anos, sendo utilizados nos diferentes campos das ciências biológicas, tornando-se a principal ferramenta para a identificação de regiões do genoma associadas com características complexas ou associadas às doenças nos humanos (ZIEGLER; KÖNIG; THOMPSON, 2008; CRUCHAGA et al., 2013; YUNG, 2013) e de características de interesse econômico nas espécies de produção zootécnica (FORTES et al., 2012; UTSUNOMIYA et al., 2013; UTSUNOMIYA et al., 2014; BUZANSKAS et al., 2014).

Os GWAS exploram o desequilíbrio de ligação (LD), que é o nível de associação não aleatória entre os marcadores e as variantes causais (“quantitative trait loci-QTL”). Esta associação ocorre, porque, segmentos de cromossomos são herdados desde um ancestral comum, mantendo-se idênticos por descendência, indicando que a taxa de recombinação é baixa. Se existem QTLs e marcadores no segmento, estes estarão em LD, sendo possível identificar o QTL através dos marcadores (GONDRO; WERF; HAYES, 2013).

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

Inicialmente, os estudos de associação genômica eram feitos considerando marcadores moleculares tipo microssatélites ou RFLP, usados em baixo número por seu alto custo, limitando-se a cobrir regiões específicas do genoma. Delineamentos experimentais específicos eram desenvolvidos para reduzir o impacto do pequeno número de marcadores sobre o poder dos testes estatísticos (DAETWYLER et al., 2008). Com o sequenciamento do genoma bovino (ELSIK et al., 2009) e o desenvolvimento de painéis contendo milhares de SNPs, novas oportunidades e desafios surgiram. Com os novos marcadores foi possível aumentar a densidade e cobrir a totalidade do genoma, assim como aumentar o número de animais utilizados, dado o menor custo quando comparado com os outros tipos de marcadores. No entanto, o grande volume de dados e de variáveis preditoras trouxeram desafios computacionais e estatísticos que precisaram ser superados (CANTOR; LANGE; SINSHEIMER, 2010; ZIEGLER; KÖNIG; THOMPSON, 2008).

Muitas metodologias estatísticas foram propostas para abordar a nova informação e contornar o problema da dimensionalidade (número de variáveis preditoras maior do que observações  $p > n$ ), quando considerados todos os SNPs na análise. As metodologias usadas com maior frequência são:

### 1.2.1 Regressão por marcador

Esta metodologia utiliza uma regressão simples para cada marcador, sendo necessário fazer tantas regressões quanto o número de SNPs a testar, e depois definir quais SNPs foram significativos (associados com o fenótipo). A associação entre o marcador e o fenótipo pode ser testada seguindo o modelo:

$$Y = Wb + Xg + e$$

em que: **Y** é o vetor de fenótipos; **W** é uma matriz que associa os efeitos fixos com as observações; **b** é o vetor de soluções para os efeitos fixos (sexo, manejo, estrutura da população etc.); **X** é a matriz que armazena o genótipo do marcador para todos os animais (codificado como 0 para o primeiro homocigoto, 1 para o heterocigoto e 2 para o segundo homocigoto); **g** contém o efeito do marcador, e **e** é um vetor aleatório de desvios  $e \sim N(0, \sigma_e^2)$ , onde  $\sigma_e^2$  é a variância do resíduo. Neste modelo, o efeito do marcador é tratado como fixo. A pressuposição é que o marcador só afetará a característica se estiver em LD com o QTL (CANTOR; LANGE; SINSHEIMER, 2010), sendo que a hipótese nula ( $H_0$ ) indica que o marcador não tem efeito sobre a característica, enquanto que a hipótese alternativa ( $H_1$ ) indica que existe um efeito do marcador associado a característica.

Uma questão importante é definir o nível de significância (p-value) dos múltiplos testes, dado que o nível definido para um único teste não é o mesmo para o experimento inteiro, sendo necessários ajustes do nível de significância dos testes individuais. Uma opção é usar o ajuste de Bonferroni para obter o nível de significância desejado no experimento inteiro, porém, a correção assume que os testes são independentes o que não é verdade, dado que os marcadores estão em LD (DUD-BRIDGE; GUSNANTO, 2008), outras possíveis correções são a técnica de permutação (CHURCHILL; DOERGE, 1994), e a taxa de descoberta de falsos positivos (FDR) que monitora o número de falsos positivos relativo ao número de resultados positivos (FERNANDO et al., 2004).

### 1.2.2 Regressão por haplótipos

As análises de GWAS também podem ser feitas utilizando haplótipos ao invés de marcadores. Nesta abordagem blocos de vários SNPs são testados para saber se estão associados com o fenótipo. A justificativa para usar haplótipos é que eles estão em maior LD com os QTLs, o que aumenta o poder dos testes estatísticos (ZHAO; FERNANDO; DEKKERS, 2007; BARENDSE, 2011).

O modelo para testar os haplótipos em um estudo de associação pode ser como o modelo anterior.

$$Y = Wb + Xg + e$$

neste caso,  $\mathbf{g}$  é agora um vetor de soluções para os haplótipos e não dos marcadores. Os haplótipos podem ser tratados como aleatórios uma vez que, provavelmente, existem muitos deles e alguns aparecem em baixa frequência. Assim,  $\mathbf{g}$  segue uma distribuição normal  $\mathbf{g} \sim N(0, \mathbf{I}\sigma_h^2)$ , onde  $\mathbf{I}$  é uma matriz identidade e  $\sigma_h^2$  é a variância do efeito do haplótipo.  $\mathbf{X}$  é uma matriz que contém o número de cópias de cada haplótipo em cada animal. Neste modelo a estimação dos parâmetros não é feita por mínimos quadrados, e sim por máxima verossimilhança, que por sua vez não permite fazer teste F para testar a hipóteses e, conseqüentemente, usa-se o teste da razão de verossimilhança.

Apesar da abordagem de haplótipos parecer atrativa, por oferecer um maior poder estatístico, há vários fatores que limitam seu uso em favor da abordagem de genótipos por SNPs. Entre eles, a necessidade de classificar os genótipos em haplótipos o que pode não ser uma tarefa fácil; outro fator limitante é o número de efeitos que tem que ser estimados, uma vez que cada bloco de haplótipos tem múltiplas variantes, as quais incrementam-se na medida que o número de SNPs no haplótipo é maior (PRYCE et al., 2010).

### 1.2.3 Regressão múltipla para os marcadores

Os métodos anteriormente descritos consideram um marcador ou um bloco de haplótipo por vez, não sendo possível considerar toda a informação disponível no modelo, o que se transforma em desvantagem. As principais desvantagens desses modelos são: 1) alta taxa de falsos positivos, porque os múltiplos testes são dependentes por causa do LD existente entre os marcadores, o que conduz a identificar marcadores que realmente não estão associados com o fenótipo; 2) sobrestimação dos efeitos dos marcadores, uma vez que os marcadores mais prováveis a exceder o nível de significância são aqueles com erros favoráveis, conduzindo a que marcadores significativos tenham efeitos sobrestimados (FERNANDO; GARRICK, 2013); 3) diferentes genes podem interagir com outros, formando uma complexa rede de interações, situação que não pode ser levada em consideração com as análises de regressão por marcador (CORDELL, 2009); e 4) regressões simples por marcador não tem poder nem precisão suficiente para identificar interações genótipo-ambiente (CANTOR; LANGE; SINSHEIMER, 2010)

Para contornar as desvantagens anteriormente descritas, foram propostas diferentes alternativas que ajustam todos os SNPs em uma única regressão, e que podem considerar interações gênicas. Estas metodologias vão desde regressões de seleção de variáveis (“stepwise”) até metodologias mais complexas, como abordagens Bayesianas (STEPHENS; BALDING, 2009) e o uso de aprendizado de máquina (“machine learning”) (SZYMCZAK et al., 2009).

#### 1.2.3.1 Regressão “stepwise”

A regressão “stepwise”, é um método que seleciona os SNPs que estão associados com o fenótipo por meio da adição de variáveis no modelo base, testando se a nova variável é significativa. O processo é repetido para todas as variáveis preditoras (neste caso os SNPs). Porém, sua aplicação em GWAS é limitada dado o grande número de SNPs a serem testados o que leva a aumentar o tempo das análises (CANTOR; LANGE; SINSHEIMER, 2010). Ayers e Cordell (2010), em um estudo de simulação, compararam vários métodos de regressão, e encontram que “forward stepwise” teve menor poder na detecção de SNPs associados com o fenótipo. Este método é utilizado principalmente em associação com outros métodos que diminuem a quantidade de SNPs e, posteriormente, a regressão “stepwise” é implementada (MOTYER et al., 2011).



### 1.2.3.2 Regressão Bayesiana

Métodos Bayesianos têm sido aplicados com sucesso em GWAS nas espécies de interesse zootécnico (LI et al., 2011; HE; LIN, 2011; SUN et al., 2011; BODDICKER et al., 2012; PETERS et al., 2012). Os modelos utilizados são uma solução para o problema da dimensionalidade (o número de observações é menor que o número de variáveis preditivas), conseguindo predições simultâneas para milhares de SNPs ao considerar os efeitos como aleatórios. Funções de penalização são consideradas dentro do modelo para controlar o nível de “shrinkage” sobre os efeitos, de forma que os efeitos são regredidos para zero, sendo mais forte a regressão para aqueles SNPs com pouco ou nenhum efeito (AYERS; CORDELL, 2010).

Embora, estes métodos tenham sido desenvolvidos para a predição de valores genéticos (MEUWISSEN; HAYES; GODDARD, 2001), eles também podem ser usados para GWAS. Além dos efeitos aditivos, eles também podem levar em consideração efeitos não aditivos (GAO et al., 2014). O ponto de partida destes métodos é um modelo linear misto da seguinte forma.

$$Y = X\beta + Z\alpha + e$$

em que:  $\mathbf{Y}$  é um vetor  $n \times 1$  de fenótipos;  $\mathbf{X}$  é uma matriz  $n \times p$  que relaciona o vetor  $\beta$  de efeitos fixos com o fenótipo  $Y$ ;  $\mathbf{Z}$  é uma matriz  $n \times k$  de genótipos (0 para o primeiro homocigoto, 1 para o heterocigoto e 2 para o segundo homocigoto) de  $k$  SNPs;  $\alpha$  é um vetor  $k \times 1$  de coeficientes aleatórios de regressão para os SNPs (efeito de substituição alélica dos marcadores) e  $\mathbf{e}$  é um vetor de resíduos (FERNANDO; GARRICK, 2013).

Em estimação Bayesiana, distribuições *a priori* devem ser atribuídas para cada parâmetro. Para  $\beta$  é atribuída uma distribuição não informativa (“flat”);  $\mathbf{e}$  segue uma distribuição normal  $N \sim (0, I\sigma_e^2)$ , onde  $\sigma_e^2$  é a variância do resíduo e é considerada como desconhecida, com uma distribuição *a priori* Qui-quadrado escalonada invertida. A diferença entre os modelos Bayesianos está na distribuição *a priori* marginal que é assumida para o vetor de parâmetros  $\alpha$ .

- Em RR-BLUP (“Ridge Regression”) uma distribuição normal  $N \sim (0, \sigma_{\alpha_i}^2)$  é atribuída como *priori* para  $\alpha$ , e para a variância  $\sigma_{\alpha_i}^2$  é usada uma distribuição Qui-quadrada escalonada invertida.
- Em BayesA, *a priori* para  $\alpha$  é uma distribuição univariada  $t$  com média zero. Neste modelo é assumida uma variância diferente para cada marcador (MEUWISSEN; HAYES; GODDARD, 2001).

- Em BayesB, *a priori* assumida para os marcadores é uma mistura de distribuições, tendo um ponto de massa em zero com probabilidade  $\pi$  e uma distribuição univariada  $t$  com probabilidade  $1 - \pi$  e média zero. Assim BayesA é um caso especial de BayesB, quando  $\pi = 0$  (MEUWISSEN; HAYES; GODDARD, 2001).
- Em BayesC, *a priori* assumida para os efeitos dos marcadores é uma mistura de distribuições, tendo um ponto de massa em zero com probabilidade  $\pi$  e uma distribuição normal univariada  $N \sim (0, \sigma_\alpha^2)$  com probabilidade  $1 - \pi$ , a variância  $\sigma_\alpha^2$  segue uma distribuição Qui-quadrada escalonada invertida. RR-BLUP é um caso especial de BayesC quando  $\pi = 0$  (HABIER et al., 2011). Em BayesB e BayesC,  $\pi$  pode ser tratado como desconhecido e ser inferido a partir dos dados, atribuindo uma distribuições beta a priori para  $\pi$ .
- No LASSO Bayesiano, a distribuição a priori atribuída para  $\alpha$  é a dupla exponencial, que tem uma maior concentração de massa em zero quando comparada como as distribuições normal e  $t$ . Este modelo usa o parâmetro  $\lambda$  que controla o nível do “shrinkage” aplicado aos efeitos dos SNPs. O valor de  $\lambda$  pode ser definido por validação cruzada ou inferido a partir dos dados. Para este ultimo caso, uma distribuição beta ou gama é atribuída como priori ao parâmetro  $\lambda$  (PARK; CASELLA, 2008; De los CAMPOS et al., 2009).

### 1.2.3.3 Machine learning

Métodos não paramétricos, como “machine learning”, são utilizados tanto para prever o valor genético (NEVES; CARVALHEIRO; QUEIROZ, 2012) como para GWAS (GONZÁLEZ-RECIO; FORNI, 2011; WANG et al., 2013; BOTTA et al., 2014), porque permitem modelar a arquitetura da característica ao tornar possível incluir interações entre os marcadores e interações marcador-ambiente, sem precisar especificar um modelo (SZYMCZAK et al., 2009). “Machine learning” usa os computadores para aprender sobre o problema com base em amostras. Independentemente do algoritmo aplicado, a construção do modelo segue os mesmos passos básicos 1) Construção do conjunto de dados para o treinamento; 2) seleção de características e variáveis, 3) construção do modelo; e 4) validação do modelo. Estes passos são repetidos várias vezes até obter o melhor modelo (GONDRO; WERF; HAYES, 2013). Os dois métodos mais utilizados em GWAS são “boosting” e “Random Forest” (LEE et al., 2011; GONDRO; WERF; HAYES, 2013; YAO et al., 2013).

Até agora não existe um método que seja superior na estimação dos efeitos dos SNPs e na detecção de regiões associadas com o fenótipo. O melhor método

depende da arquitetura genética da característica e da estrutura da população em estudo (GONZÁLEZ-RECIO; FORNI, 2011).

### **1.3 Objetivos**

#### Objetivos Gerais

- Determinar as consequências de atribuir os escores com base em uma referência relativa, sobre a estimação de parâmetros e valores genéticos.
- Identificar as regiões do genoma que estão associadas com os escores visuais de conformação (C), precocidade (P) e musculosidade (M), mensurados à desmama em bovinos Nelore, por meio de estudos de associação genômica (GWAS).

#### Objetivos Específicos

- Avaliar o efeito da seleção e do tamanho de grupo de atribuição dos escores, sobre a estimação de parâmetros e valores genéticos em escores visuais.
- Comparar a utilização dos modelos BayesC e LASSO Bayesiano para GWAS.
- Identificar os genes localizados nas regiões associadas com os escores visuais.



## REFERÊNCIAS

- AFOLAYAN, R. A.; PITCHFORD, W. S.; DELAND, M.; MCKIERNAN, W. Breed variation and genetic parameters for growth and body development in diverse beef cattle genotypes. **Animal**, v. 1, n. 1, p. 13–20, 2007.
- AYERS, K. L.; CORDELL, H. J. SNP selection in genome-wide and candidate gene studies via penalized logistic regression. **Genetic Epidemiology**, v. 34, n. 8, p. 879–891, 2010.
- BARENDSE, W. Haplotype analysis improved evidence for candidate genes for intramuscular fat percentage from a genome wide association study of cattle. **PloS One**, v. 6, n. 12, p. e29601, 2011.
- BARICHELLO, F.; ALENCAR, M.; JÚNIOR, R. T. et al. Herdabilidade e correlações quanto a peso, perímetro escrotal e escores visuais à desmama, em bovinos Canchim. **Pesquisa Agropecuária Brasileira**, v. 45, n. 6, p. 563–570, 2010.
- BERG, R. T.; BUTTERFIELD, R. M. **New concepts of cattle growth**. [S.l.]: University of sydney, 1976.
- BIF. **BIF Guidelines for Uniform Beef Improvement Programs**. [S.l.]: Beef Improvement Federation, 1974.
- BODDICKER, N.; WAIDE, E.; ROWLAND, R.; LUNNEY, J.; GARRICK, D.; REECY, J.; DEKKERS, J. Evidence for a major QTL associated with host response to porcine reproductive and respiratory syndrome virus challenge. **Journal of Animal Science**, v. 90, n. 6, p. 1733–1746, 2012.
- BOTTA, V.; LOUPPE, G.; GEURTS, P.; WEHENKEL, L. Exploiting SNP correlations within random forest for genome-wide association studies. **PloS One**, v. 9, n. 4, p. e93379, 2014.
- BUZANSKAS, M. E.; GROSSI, D. A.; VENTURA, R. V.; SCHENKEL, F. S.; SARGOLZAEI, M.; MEIRELLES, S. L.; MOKRY, F. B.; HIGA, R. H.; MUDADU, M. A.; SILVA, M. V. B. da et al. Genome-wide association for growth traits in Canchim beef cattle. **PloS One**, v. 9, n. 4, p. e94802, 2014.
- CANTOR, R. M.; LANGE, K.; SINSHEIMER, J. S. Prioritizing GWAS results: a review of statistical methods and recommendations for their application. **The American Journal of Human Genetics**, v. 86, n. 1, p. 6–22, 2010.
- CHURCHILL, G. A.; DOERGE, R. W. Empirical threshold values for quantitative trait mapping. **Genetics**, v. 138, n. 3, p. 963–971, 1994.
- CONROY, S.; DRENNAN, M.; KENNY, D.; MCGEE, M. The relationship of live animal muscular and skeletal scores, ultrasound measurements and carcass classification scores with carcass composition and value in steers. **Animal**, v. 3, n. 11, p. 1613–1624, 2009.
- CORDELL, H. J. Detecting gene–gene interactions that underlie human diseases. **Nature Reviews Genetics**, v. 10, n. 6, p. 392–404, 2009.

CRUCHAGA, C.; KAUWE, J. S.; HARARI, O.; JIN, S. C.; CAI, Y.; KARCH, C. M.; BENITEZ, B. A.; JENG, A. T.; SKORUPA, T.; CARRELL, D. et al. GWAS of cerebrospinal fluid tau levels identifies risk variants for alzheimer's disease. **Neuron**, v. 78, n. 2, p. 256–268, 2013.

DAETWYLER, H. D.; SCHENKEL, F. S.; SARGOLZAEI, M.; ROBINSON, J. A. B. A genome scan to detect quantitative trait loci for economically important traits in Holstein cattle using two methods and a dense single nucleotide polymorphism map. **Journal of Dairy Science**, v. 91, n. 8, p. 3225–3236, 2008.

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. 1, p. 375–385, 2009.

DUDBRIDGE, F.; GUSNANTO, A. Estimation of significance thresholds for genome-wide association scans. **Genetic Epidemiology**, v. 32, n. 3, p. 227–234, 2008.

DUITAMA, L. O.; FARAH, M.; UTSONOMIYA, A. H.; ONO, R. K.; PIRES, M.; da FONSECA, R. Uso de modelos lineares mistos na avaliação genética de escores visuais: estudo de simulação. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, v. 66, n. 4, p. 1239–1246, 2014.

ELSIK, C. G.; TELLAM, R. L.; WORLEY, K. C. et al. The genome sequence of taurine cattle: a window to ruminant biology and evolution. **Science**, v. 324, n. 5926, p. 522–528, 2009.

FARIA, C. U.; MAGNABOSCO, C. U.; ALBUQUERQUE, L. G.; 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.; MAGNABOSCO, C. U.; ALBUQUERQUE, L. G.; De los Reyes, A.; BEZERRA, L. A. F.; LÔBO, R. B. Estimativas de correlações genéticas entre escores visuais e características de crescimento em bovinos da raça Nelore utilizando modelos bayesianos linear-limiar. **Ciência Animal Brasileira**, v. 9, n. 2, p. 327–340, 2008.

FARIA, C. U.; MAGNABOSCO, C. U.; ALBUQUERQUE, L. G.; BEZERRA, L. A. F. Estimativas de correlações genéticas entre escores visuais e características de carcaça medidas por ultrassonografia em bovinos Nelore utilizando modelos bayesianos linear-limiar. **Revista Brasileira de Zootecnia**, v. 38, n. 11, p. 2144–2151, 2009.

FERNANDO, R. L.; BILLINGSLEY, R.; GIANOLA, D. Effects of method of scaling on heritability estimates and sire evaluations for frame size at weaning in Angus cattle. **Journal of Animal Science**, v. 56, n. 5, p. 1047–1056, 1983.

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

FERNANDO, R. L.; NETTLETON, D.; SOUTHEY, B. R.; DEKKERS, J. C. M.; ROTHSCHILD, M. F.; SOLLER, M. Controlling the proportion of false positives in multiple dependent tests. **Genetics**, v. 166, n. 1, p. 611–619, 2004.

FORTES, M.; LEHNERT, S.; BOLORMAA, S.; REICH, C.; FORDYCE, G.; CORBET, N.; WHAN, V.; HAWKEN, R.; REVERTER, A. Finding genes for economically important traits: Brahman cattle puberty. **Animal Production Science**, v. 52, n. 3, p. 143–150, 2012.

FRIES, L. Uso de escores visuais em programas de seleção para produtividade em gado de corte. In: SEMINARIO NACIONAL - REVISÃO DE CRITERIOS DE SELEÇÃO E JULGAMENTO EM GADO DE CORTE, 1996, Uberaba - MG. **Anais ... Uberaba - MG: ABCZ**, 1996.

GAO, H.; WU, Y.; LI, J.; LI, H.; LI, J.; YANG, R. Forward lasso analysis for high-order interactions in genome-wide association study. **Briefings in Bioinformatics**, v. 15, n. 4, p. 552–561, 2014.

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

GONDRO, C.; WERF, J. Van der; HAYES, B. J. **Genome-wide Association Studies and Genomic Prediction**. [S.l.]: Springer, 2013.

GONZÁLEZ-RECIO, O.; FORNI, S. Genome-wide prediction of discrete traits using bayesian regressions and machine learning. **Geneics Selection Evolution**, v. 43, n. 7, p. 1–12, 2011.

GORDO, D. G. **Estudo genético quantitativo da qualidade da carcaça e da carne e suas associações com as características de avaliação visual na raça Nelore**. Tese (Doutorado) — Universidade estadual paulista-UNESP, 2014.

GORDO, D. M.; BALDI, F.; LÔBO, R. B.; FILHO, W. K.; SAINZ, R. D.; ALBUQUERQUE, L. 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.

HABIER, D.; FERNANDO, R. L.; KIZILKAYA, K.; GARRICK, D. J. Extension of the bayesian alphabet for genomic selection. **BMC Bioinformatics**, v. 12, n. 186, p. 1–12, 2011.

HE, Q.; LIN, D.-Y. A variable selection method for genome-wide association studies. **Bioinformatics**, v. 27, n. 1, p. 1–8, 2011.

JÚNIOR, J. J.; PITA, F. V.; FRIES, L. A.; ALBUQUERQUE, L. G. Influência de alguns fatores de ambiente sobre os escores de conformação, precocidade e musculatura à desmama em um rebanho da raça Nelore. **Revista Brasileira de Zootecnia**, v. 30, n. 6, p. 1697–1703, 2001.

KOURY FILHO, W.; ALBUQUERQUE, L. G.; ALENCAR, M.; FORNI, S.; SILVA, J.; LÔBO, R. B. Estimativas de herdabilidade e correlações para escores visuais, peso e altura ao sobreano em rebanhos da raça Nelore. **Revista Brasileira de Zootecnia**, 2009, v. 38, n. 12, p. 2362–2367, 2009.

KOURY FILHO, W.; ALBUQUERQUE, L. G.; FORNI, S.; SILVA, J.; YOKOO, M. J.; ALENCAR, M. M. de. Estimativas de parâmetros genéticos para os escores visuais e suas associações com peso corporal em bovinos de corte. **Revista Brasileira de Zootecnia**, v. 39, n. 5, p. 1015–1022, 2010.

LEE, I.; BLOM, U. M.; WANG, P. I.; SHIM, J. E.; MARCOTTE, E. M. Prioritizing candidate disease genes by network-based boosting of genome-wide association data. **Genome Research**, v. 21, n. 7, p. 1109–1121, 2011.

LI, J.; DAS, K.; FU, G.; LI, R.; WU, R. The bayesian lasso for genome-wide association studies. **Bioinformatics**, v. 27, n. 4, p. 516–523, 2011.

LONG, R. **The Ankony Scoring System: Its Uses in Herd Improvement**. Ankony Angus Corporation, 1971. Disponível em: <<http://books.google.com.br/books?id=rhkpcAAACAAJ>>.

MAIWASHE, A.; BRADFIELD, M.; THERON, H.; WYK, J. V. Genetic parameter estimates for body measurements and growth traits in South African Bonsmara cattle. **Livestock Production Science**, v. 75, n. 3, p. 293–300, 2002.

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.

MOTYER, A. J.; MCKENDRY, C.; GALBRAITH, S.; WILSON, S. R. LASSO model selection with post-processing for a genome-wide association study data set. **BMC Proceedings**, v. 5, n. Suppl 9, p. S24, 2011.

NEVES, H. H.; CARVALHEIRO, R.; QUEIROZ, S. A. A comparison of statistical methods for genomic selection in a mice population. **BMC Genetics**, v. 13, n. 100, p. 1–17, 2012.

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

PETERS, S.; KIZILKAYA, K.; GARRICK, D.; FERNANDO, R.; REECY, J.; WEABER, R.; SILVER, G.; THOMAS, M. Bayesian genome-wide association analysis of growth and yearling ultrasound measures of carcass traits in Brangus heifers. **Journal of Animal science**, v. 90, n. 10, p. 3398–3409, 2012.

PRYCE, J.; BOLORMAA, S.; CHAMBERLAIN, A.; BOWMAN, P.; SAVIN, K.; GODDARD, M.; HAYES, B. A validated genome-wide association study in 2 dairy cattle breeds for milk production and fertility traits using variable length haplotypes. **Journal of Dairy Science**, v. 93, n. 7, p. 3331–3345, 2010.

SHIOTSUKI, L.; SILVA, J.; TONHATI, H.; ALBUQUERQUE, L. G. Genetic associations of sexual precocity with growth traits and visual scores of conformation, finishing, and muscling in Nelore cattle. **Journal of Animal Science**, v. 87, n. 5, p. 1591–1597, 2009.



STEPHENS, M.; BALDING, D. J. Bayesian statistical methods for genetic association studies. **Nature Reviews Genetics**, v. 10, n. 10, p. 681–690, 2009.

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. **BMC Proceedings**, v. 5, n. Suppl 3, p. S13, 2011.

SZYMCZAK, S.; BIERNACKA, J. M.; CORDELL, H. J.; GONZÁLEZ-RECIO, O.; KÖNIG, I. R.; ZHANG, H.; SUN, Y. V. Machine learning in genome-wide association studies. **Genetic Epidemiology**, v. 33, n. S1, p. S51–S57, 2009.

UTSUNOMIYA, Y. T.; CARMO, A. S.; NEVES, H. H.; CARVALHEIRO, R.; MATOS, M. C.; ZAVAREZ, L. B.; ITO, P. K.; O'BRIEN, A. M. P.; SÖLKNER, J.; PORTO-NETO, L. R. et al. Genome-wide mapping of loci explaining variance in scrotal circumference in Nelore cattle. **PloS One**, v. 9, n. 2, p. e88561, 2014.

UTSUNOMIYA, Y. T.; CARMO, A. S. D.; CARVALHEIRO, R.; NEVES, H. H.; MATOS, M. C.; ZAVAREZ, L. B.; O'BRIEN, A. M. P.; SÖLKNER, J.; MCEWAN, J. C.; COLE, J. B. et al. Genome-wide association study for birth weight in Nelore cattle points to previously described orthologous genes affecting human and bovine height. **BMC Genetics**, BioMed Central Ltd, v. 14, n. 52, p. 1–12, 2013.

VISSCHER, P. M.; BROWN, M. A.; MCCARTHY, M. I.; YANG, J. Five years of GWAS discovery. **The American Journal of Human Genetics**, Elsevier, v. 90, n. 1, p. 7–24, 2012.

WANG, Y.; GOH, W.; WONG, L.; MONTANA, G. Random forests on hadoop for genome-wide association studies of multivariate neuroimaging phenotypes. **BMC Bioinformatics**, v. 14, n. 16, p. 1–15, 2013.

YAO, C.; SPURLOCK, D.; ARMENTANO, L.; JR, C. P.; VANDEHAAR, M.; BICKHART, D.; WEIGEL, K. Random forests approach for identifying additive and epistatic single nucleotide polymorphisms associated with residual feed intake in dairy cattle. **Journal of Dairy Science**, v. 96, n. 10, p. 6716–6729, 2013.

YOKOO, M. J.; WERNECK, J. N.; PEREIRA, M. C.; ALBUQUERQUE, L. G.; FILHO, W. K. Correlações genéticas entre escores visuais e características de carcaça medidas por ultrassom em bovinos de corte. **Pesquisa Agropecuária Brasileira**, v. 44, n. 2, p. 197–202, 2009.

YUNG, W. A. From GWAS risk foci to glioma molecular subclass. **Neuro Oncology**, v. 15, n. 5, p. 513–514, 2013.

ZHAO, H.; FERNANDO, R. L.; DEKKERS, J. C. Power and precision of alternate methods for linkage disequilibrium mapping of quantitative trait loci. **Genetics**, v. 175, n. 4, p. 1975–1986, 2007.

ZIEGLER, A.; KÖNIG, I. R.; THOMPSON, J. R. Biostatistical aspects of genome-wide association studies. **Biometrical Journal**, v. 50, n. 1, p. 8–28, 2008.

ZINK, V.; ŠTÍPKOVÁ, M.; LASSEN, J. Genetic parameters for female fertility, locomotion, body condition score, and linear type traits in Czech Holstein cattle. **Journal of Dairy Science**, v. 94, n. 10, p. 5176–5182, 2011.

## **CAPÍTULO 2 - ESTIMAÇÃO DE PARÂMETROS E VALORES GENÉTICOS EM CARACTERÍSTICAS DE ESCORES VISUAIS: ESTUDO DE SIMULAÇÃO**

**RESUMO:** Os escores visuais são medidas subjetivas de características morfológicas, amplamente usadas nos programas de melhoramento genético animal no Brasil. Uma particularidade dos escores é a forma como eles são atribuídos, dado que cada grupo avaliado usa uma referência de comparação diferente, que depende do grupo. Assim, a referência é o perfil médio da característica observado no grupo de animais avaliado. Esta particularidade pode trazer dificuldades na estimação de parâmetros e valores genéticos. Objetivou-se com este estudo determinar as consequências de atribuir os escores usando diferentes referências, sobre a estimação de parâmetros e valores genéticos. Um estudo de simulação foi conduzido, considerando 6 cenários formados pela combinação de duas populações (com seleção e sem seleção) e 3 tamanhos de grupo na atribuição dos escores (10, 40 e 100 animais). A característica categórica do escore, foi gerada a partir de uma característica contínua não observável, subjacente ao escore, com herdabilidade de 0,4. Os parâmetros e valores genéticos foram estimados usando dois modelos de limiar que diferiram só na correção para os efeitos ambientais. As estimativas de herdabilidade para as duas populações foram próximas quando o modelo corrigiu para os efeitos ambientais. No entanto, houve grande diferença quando o modelo não considerou os efeitos ambientais, sendo que nas populações com seleção houve uma queda da herdabilidade ao longo das gerações, como consequência da falta de correção para o efeito da seleção. O tamanho de grupo de avaliação teve efeito principalmente na acurácia de predição dos valores genéticos, uma vez que acurácias maiores estiveram relacionadas com maiores tamanhos de grupo de avaliação. A queda de acurácia quando grupos pequenos são utilizados está associada com a heterogeneidade de variância fenotípica entres os grupos de avaliação, o que viola as pressuposições de homogeneidade de variâncias na metodologia BLUP. Em conclusão a forma de atribuição dos escores visuais não afeta a estimação de parâmetros e valores genéticos, no entanto, é desejável que sejam formados grupos de avaliação superiores a 40 animais para obter maiores acurácias e progressos genéticos.

**Palavras-chave:** modelo de limiar, parâmetros genéticos, simulação

## 2.1 Introdução

Medidas de escores visuais em gado de corte buscam auxiliar a seleção de animais mais produtivos e adaptados ao ambiente, uma vez que os escores estão correlacionados geneticamente de forma positiva, com características de importância econômica, principalmente de crescimento (MAIWASHE et al., 2002; AFOLAYAN et al., 2007; KOURY FILHO et al., 2010; BERTIPAGLIA et al., 2012). Os principais programas de melhoramento genético incluem DEPs (Diferença esperada da progênie) para estas características, usando alguma das metodologias existentes. As metodologias se baseiam na avaliação de características morfológicas, por meio da apreciação visual, atribuindo escores para diferentes níveis de expressão do fenótipo. Estas avaliações são feitas por avaliadores treinados, minimizando a subjetividade do processo (JÚNIOR et al., 2001).

Desde a incorporação dos escores nos programas de melhoramento genético, a metodologia tem sofrido modificações, e na atualidade os escores visuais são atribuídos como sugerido por Fries (1996). O processo pode ser dividido em dois passos.

- O primeiro passo é formar grupos de animais provenientes das mesmas condições de criação, de forma que as diferenças observadas dentro de cada grupo por causas ambientais sejam minimizadas.
- Em um segundo passo, os escores são atribuídos dentro de grupo, usando como referência o fenótipo médio do grupo, cada animal é comparado com o perfil médio para atribuir o escore. Para evitar que os escores dentro de grupo fiquem concentrados em poucos níveis da escala de avaliação, a metodologia exige que dentro de grupo todos os escores sejam atribuídos.

Dado o processo usado na atribuição dos escores, dois aspectos devem ser levados em consideração no momento da estimação de parâmetros e valores genéticos. O primeiro é o fenótipo categórico (5 ou 6 níveis), formado por uma escala ordinal, que se aproxima a uma distribuição normal, tendo maior massa nos níveis intermediários da escala. Dada a proximidade com a distribuição normal, muitos estudos consideram o fenótipo como contínuo e usam o modelo linear na estimação dos parâmetros genéticos (KOURY FILHO et al., 2009; SHIOTSUKI et al., 2009; YOKOO et al., 2009; BARICHELLO et al., 2010; KOURY FILHO et al., 2010; GORDO et al., 2012; BERTIPAGLIA et al., 2012). No entanto, Fernando, Billingsley e Gianola (1983), alertaram que, quando um modelo linear é usado, a invariância do BLUP em popula-

ções sob seleção não é garantida, afetando a classificação dos animais para seleção, principalmente, quando o número de níveis da escala categórica e a herdabilidade diminuem (MEIJERING; GIANOLA, 1985; ABDEL-AZIM; BERGER, 1999). Faria et al. (2008) compararam estimativas de parâmetros genéticos usando os modelos linear e de limiar para escores visuais, não encontrando diferença significativa nos parâmetros estimados. No entanto, o modelo de limiar deve ser preferido, dadas suas propriedades estatísticas que garantem a correta estimação.

O segundo aspecto que torna os escores visuais diferentes de qualquer outra característica é a forma como eles são atribuídos, dado que não é feita com base em uma referência absoluta, como feito em características lineares de tipo em gado leiteiro (SHORT; LAWLOR, 1992; ZINK; ŠTÍPKOVÁ; LASSEN, 2011). Neste caso, a referência de avaliação é diferente para cada grupo, e depende só da composição do grupo (animais no grupo), não sendo possível comparar fenotipicamente animais pertencentes a grupos de avaliação diferentes (KOURY FILHO et al., 2009). Fatores que modifiquem a composição do grupo de avaliação, como o tamanho do grupo (número de animais), podem afetar a atribuição dos escores visuais. Duitama et al. (2014), em um estudo de simulação, reportaram que a seleção também afeta a atribuição dos escores, uma vez que os parâmetros genéticos foram subestimados.

Devido às particularidades apresentadas pelo processo de atribuição dos escores visuais, este trabalho teve por objetivo, determinar as consequências, de atribuir os escores com base em uma referência relativa sobre a estimação de parâmetros e valores genéticos, por meio de um estudo de simulação, considerando populações com e sem seleção e três tamanhos de grupo de avaliação (10, 40 e 100 animais).

## **2.2 Material e métodos**

Um estudo de simulação foi conduzido, considerando três tamanhos de grupo de avaliação (10, 40 e 100 animais) em duas populações, uma sem seleção e a outra com seleção. Para a simulação das populações foi usado o software QMSim (SARGOLZAEI; SCHENKEL, 2009), que faz o processo em dois passos. Em um primeiro passo, uma população histórica é simulada para estabilizar o equilíbrio entre as taxas de deriva genética e mutação, e em um segundo passo é construída a estrutura da população desejada. Na simulação, a população histórica foi formada por 1000 gerações. Da última geração, foram escolhidos aleatoriamente 2000 fêmeas e 50 machos, que foram a base para as populações simuladas.

O Genoma simulado foi constituído por 29 pares de cromossomos autossômicos, com comprimento variando de 40 até 145 cM, simulando o genoma bovino, com um total de 2333 cM (SNELLING et al., 2007). O número de “QTLs” controlando a característica foi de 500, distribuídos aleatoriamente ao longo do genoma. Os QTLs foram bialélicos e trialélicos, e seus efeitos foram amostrados de uma distribuição normal. Toda a variância genética aditiva foi atribuída aos efeitos dos QTLs.

Dada a natureza categórica dos escores, foi assumido que o fenótipo observável no animal é o resultado de uma característica subjacente não observável com distribuição normal padrão  $N \sim (0, 1)$ , conhecida como “liability” (FALCONER; MACKAY, 1996). Assim, a resposta categórica observável, é o resultado de um animal exceder o limite de um ponto específico (“threshold”) sobre a “liability”.

O modelo utilizado para simular a característica subjacente foi:

$$y = \mu + gc + a + e$$

em que  $y$  é a “liability”;  $\mu$  é a média populacional que foi igual a zero;  $gc$  é o efeito fixo de grupo contemporâneo, formado por sexo e ano de nascimento, o qual foi simulado a partir de uma distribuição normal com média zero e variância 0,2, equivalente a 20% da variância total da “liability”;  $a$  é o efeito genético aditivo do animal, o qual foi calculado como o somatório dos efeitos alelicos dos QTLs para cada animal; e  $e$  é o resíduo, simulado a partir de uma distribuição normal com média zero e variância 0,6. A herdabilidade da característica subjacente foi de 0,4.

Como a “liability” na população segue uma distribuição normal, é de se esperar que dentro de grupos de avaliação a distribuição normal se mantenha. Assim, a característica subjacente foi categorizada em 5 níveis dentro de cada grupo de avaliação, simulando a metodologia empregada pelos avaliadores (KOURY FILHO et al., 2009). Os “threshold” dentro de grupo, foram definidos com base na distribuição normal da “liability” no grupo, os pontos foram -1,3, -0,5, 0,5 e 1,3 desvios-padrão em relação à média da distribuição, de forma que a distribuição dos escores dentro de grupo se assemelha aos encontrados em dados reais (Figura 2.1).

Os animais das populações simuladas foram acasalados por 12 vezes, a proporção fêmea/macho foi mantida constante. A taxa de reposição por ciclo de acasalamento para fêmeas e machos foi de 15% e 20% respectivamente. Os critérios de descarte foram a idade nas populações sem seleção, e a idade e os valores genéticos estimados (VGEs) para as populações sob seleção, sendo descartados os animais mais velhos e de menores VGEs.

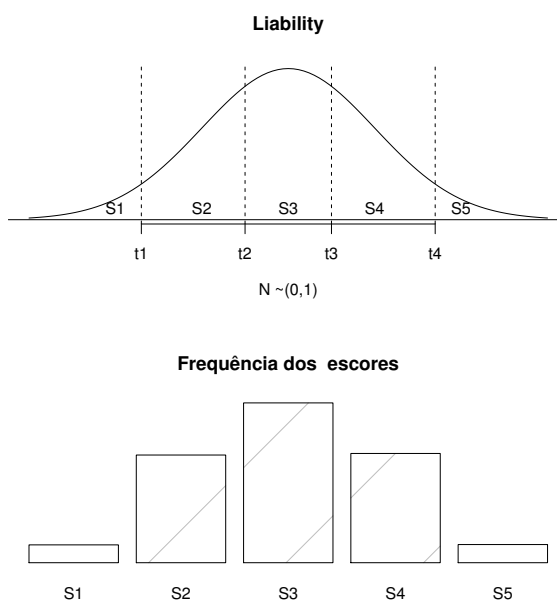


Figura 2.1 – Representação do escore visual e a característica subjacente dentro de grupo de avaliação. Em cima, a “liability” com os 4 “thresholds” que definem os 5 fenótipos. Em baixo, a frequência observada para cada categoria fenotípica.

A cada ciclo de acasalamento parâmetros e valores genéticos foram estimados usando um modelo de limiar, que modela a característica subjacente ao fenótipo observável. O modelo utilizado, em forma matricial, foi:

$$\lambda = \mathbf{F}\mathbf{t} + \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}$$

em que:  $\lambda$  é o vetor não observável da “liability”;  $\mathbf{t}$  vetor de  $m-1$  “threshold”, onde  $m$  é o número de categorias do fenótipo, neste caso 5;  $\boldsymbol{\beta}$  é o vetor de soluções para os efeitos sistemáticos de ambiente;  $\mathbf{u}$  vetor aleatório de efeitos genéticos aditivos;  $\mathbf{e}$  vetor de erros;  $\mathbf{X}$  e  $\mathbf{Z}$  são as matrizes de incidência que relacionam os vetores  $\boldsymbol{\beta}$  e  $\mathbf{u}$  com a “liability” e,  $\mathbf{F}$  a matriz que define a probabilidade de um animal estar nas distintas categorias do fenótipo observável, as probabilidades dependem dos “threshold” e da função  $\Phi$ , que é a distribuição acumulada normal padrão.

Para todos os cenários dois modelos foram utilizados, diferindo apenas na inclusão do efeito de grupo contemporâneo. O primeiro modelo considerou o grupo contemporâneo como efeito fixo e no segundo modelo o efeito não foi incluído, considerando que a atribuição dos escores visuais dentro de grupo corrige para os efeitos ambientais usados para agrupar os animais no momento da atribuição dos escores (DUTAMA et

al., 2014). Na tabela 2.1 é apresentada a combinação entre as populações simuladas e os modelos de avaliação utilizados.

Tabela 2.1 – Combinação entre as populações simuladas e os modelos utilizados na estimação de parâmetros e valores genéticos.

Com Seleção				Sem Seleção			
Nome	TG	Modelo	Replicas	Nome	TG	Modelo	Replicas
CS10_M1	10	M1	10	SS10_M1	10	M1	10
CS10_M2	10	M2	10	SS10_M2	10	M2	10
CS40_M1	40	M1	10	SS40_M1	40	M1	10
CS40_M2	40	M2	10	SS40_M2	40	M2	10
CS100_M1	100	M1	10	SS100_M1	100	M1	10
CS100_M2	100	M2	10	SS100_M2	100	M2	10

TG= tamanho de grupo de avaliação; M1 e M2= modelo sem e com grupo contemporâneo como efeito fixo, respectivamente.

Nas populações sem seleção, a cada ciclo de acasalamentos 15% e 20% das fêmeas e machos mais velhos eram substituídos por animais nascidos no ciclo anterior, sendo escolhidos de forma aleatória, mantendo assim, um tamanho constante na população. Para as populações sob seleção a mesma quantidade de fêmeas e machos foram substituídos, porém, os animais escolhidos foram os de maior valor genético estimado (VGEs).

Os componentes de variância e valores genéticos, foram estimados usando inferência Bayesiana com o algoritmo de amostragem de Gibbs, as *prioris* para os componentes de variância foram não informativas. A distribuição *a posteriori* dos componentes de variância e valores genéticos foram construídas a partir de uma cadeia de 100.000 amostras, da qual foram retiradas as primeiras 20.000 (“burn-in”). O tamanho da cadeia foi definido com base na análise de convergência dos testes Geweke (1992) e Heidelberger e Welch (1983), implementados no pacote Coda do software R (PLUMMER et al., 2006). A estimação dos componentes de variância e valores genéticos foi feita usando o software THRGIBBS1F90 (MISZTAL et al., 2002), que implementa o modelo de limiar sob estimação Bayesiana. Para simplificar a estimação dos parâmetros, a variância residual foi fixada em 1, os “threshold” t1 e t2, foram fixados em 0 e 1 e, os restantes foram estimados a partir dos dados (WANG; QUAAS; POLLAK, 1997).

Os critérios usados para determinar o efeito que tem o tamanho de grupo e a seleção, sobre a estimação de parâmetros e valores genéticos em características de escores visuais foram:



- Erro Quadrático Médio (EQM), foi usado como indicador da qualidade na estimação da herdabilidade. O EQM foi calculado como descrito por (CASELLA; BERGER, 2002)

$$\frac{1}{n} \sum_{i=1}^n (\hat{h}_i^2 - h^2)^2$$

em que:  $n$  é o número de réplicas de cada cenário simulado, neste caso 10;  $\hat{h}_i^2$  é a herdabilidade estimada para cada réplica e,  $h^2$  é o valor verdadeiro da herdabilidade.

- Correlações de Pearson entre os valores genéticos verdadeiros (VGVs) e os estimados (VGEs), como indicador da acurácia de predição (FALCONER; MACKAY, 1996; MRODE, 2005).
- Porcentagem de coincidência (só nos cenários com seleção), entre os 10 machos selecionados e os 10, 20 e 30 machos candidatos à seleção com os maiores VGVs.
- Eficiência da seleção (só nos cenários com seleção), definida como a porcentagem de ganho genético verdadeiro e estimado com referência ao máximo ganho genético possível. O ganho genético por ano foi calculado por meio da regressão dos valores genéticos verdadeiros e estimados ao longo dos anos. Foi assumido que um ciclo de acasalamento equivale a 1 ano de tempo. O máximo ganho genético é obtido quando a acurácia é igual a 1, e foi calculado como:

$$\text{máx } \frac{\Delta G}{\text{Ano}} = \frac{(i_m + i_f) * r_{\hat{a}a} * \sigma_a}{IG_m + IG_f}$$

em que:  $\frac{\Delta G}{\text{Ano}}$  é o ganho genético por ano;  $i_m$  intensidade de seleção dos machos;  $i_f$  intensidade de seleção das fêmeas;  $r_{\hat{a}a}$  acurácia;  $\sigma_a$  desvio padrão genético;  $IG_m$  intervalo de gerações dos machos e  $IG_f$  intervalo de gerações das fêmeas. Assim, o máximo ganho genético esperado nas populações sob seleção foi:

$$\text{máx } \frac{\Delta G}{\text{Ano}} = \frac{(2,66 + 1,15) * 1 * 0,64}{4 + 4,5} = 0,286 \text{ unidades da "liability"}$$

### 2.3 Resultados e discussão

Na Figura 2.2 é apresentado o comportamento médio da herdabilidade ao longo dos ciclos de acasalamentos para os cenários simulados. A inclusão do efeito fixo de grupo contemporâneo no modelo melhorou as estimativas de herdabilidade nas duas populações, principalmente para os cenários com seleção, onde a inclusão do efeito fixo corrigiu para o processo de seleção, uma vez que, considerar grupos de animais

(grupo contemporâneo) com diferentes méritos genéticos, equivale a incluir no modelo uma função linear que descreva o processo de seleção (HENDERSON, 1973). Como consequência de não corrigir para o efeito da seleção, as estimativas de herdabilidade diminuem rapidamente, como observado quando o efeito fixo de grupo contemporâneo não foi considerado no modelo para os cenários com seleção.

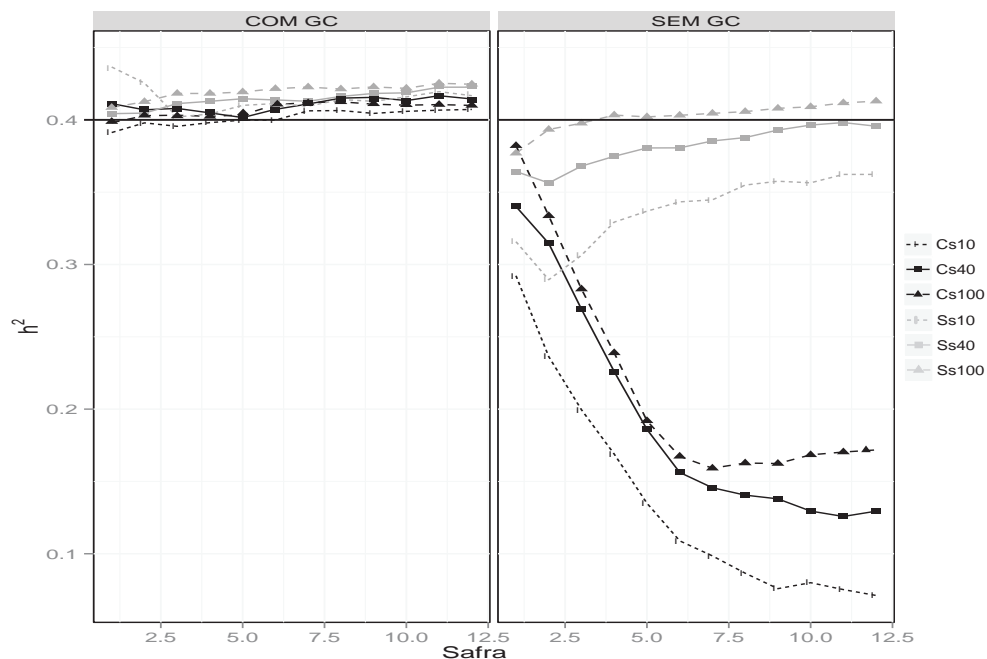


Figura 2.2 – Estimativas de herdabilidade ao longo dos acasalamentos para os cenários simulados, usando dois modelos na estimação de parâmetros genéticos (com grupo contemporâneo a esquerda e sem grupo contemporâneo a direita). Cs e Ss indicam populações com e sem seleção, respectivamente; 10, 40 e 100 representam o tamanho do grupo de avaliação.

Nos cenários sem seleção a diferença entre os modelos de avaliação foi menor. No modelo que não considerou o efeito fixo de grupo contemporâneo as estimativas de herdabilidade foram se aproximando ao valor verdadeiro à medida que a informação utilizada na estimação aumentava, especialmente quando o tamanho de grupo de avaliação foi grande. Esse comportamento ocorreu, porque, no momento de atribuir os escores visuais dentro de grupo, estes dependem da variância genética e residual contida no grupo, enquanto que a variabilidade causada pelos efeitos ambientais (grupo contemporâneo) não é levada em consideração, uma vez que a atribuição é feita dentro de grupo e os animais dentro de grupo de avaliação estão sob os mesmos efeitos ambientais. Portanto, pode-se inferir que o escore atribuído a cada animal estaria corrigido para os efeitos ambientais utilizados na formação do grupo de avaliação. Sob esse ponto de vista, animais com o mesmo escore pertencentes a diferentes gru-

pos de avaliação seriam comparáveis, uma vez que o escore reflete o desempenho dos animais se estivessem sob as mesmas condições ambientais.

O tamanho do grupo de avaliação influenciou a estimativa da herdabilidade principalmente no modelo que não considerou o efeito de grupo contemporâneo (Figura 2.2). Estas diferenças são consequência da heterogeneidade da variância fenotípica existente entre os grupos contemporâneos, de forma que quando o grupo de avaliação é menor, a heterogeneidade é maior como indicado na Figura 2.3. Uma das pressuposições do BLUP nas avaliações genéticas é a homogeneidade de variâncias entre os níveis dos efeitos fixos e aleatórios, quando isto não se verifica, os parâmetros e, em maior medida, os valores genéticos são afetados (HILL, 1984; VISSCHER; HILL, 1992). A inclusão do efeito fixo de grupo contemporâneo diminuiu as diferenças das estimativas de herdabilidade entre os diferentes tamanhos do grupo de avaliação, indicando que a inclusão deste efeito corrige o viés na estimação dos componentes de variância quando há heterogeneidade de variâncias.

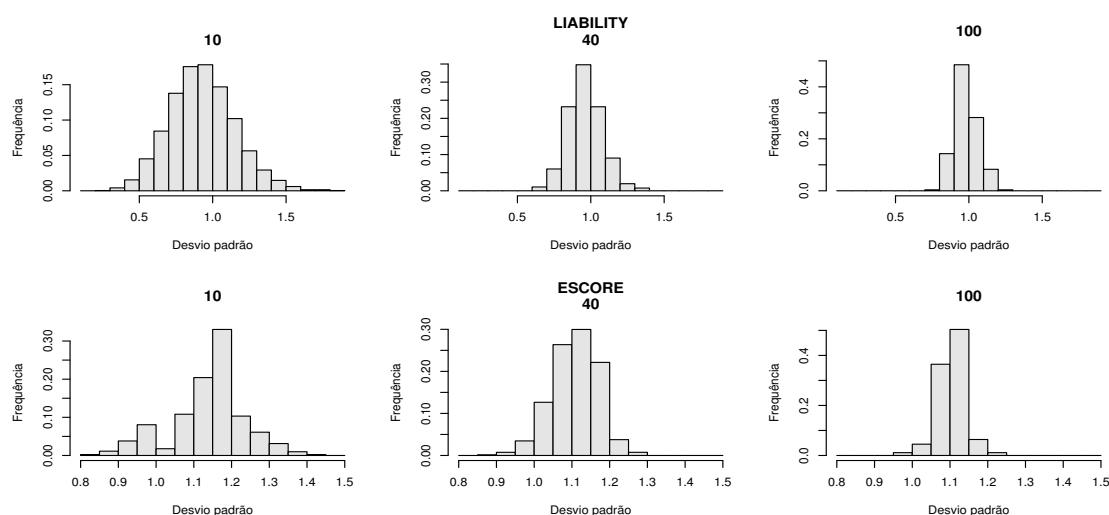


Figura 2.3 – Distribuição dos desvios padrão fenotípicos dos grupos contemporâneos para os diferentes tamanhos de grupo de avaliação (10, 40 e 100 animais) no fenótipo observável (escore) e subjacente (“liability”). Dados tomados de uma replica.

Na Tabela 2.2 são apresentadas a médias da herdabilidade e os EQM para os cenários simulados, pode-se notar a importância de incluir o efeito fixo de grupo contemporâneo no modelo quando há seleção e, ou heterogeneidade de variâncias, uma vez que as estimativas de herdabilidade ficam mais próximas do valor verdadeiro e a variabilidade das estimativas diminuem (menor EQM). Em ausência de seleção e heterogeneidade (cenários SS100\_M1 e SS100\_M2) as diferenças entre os modelos

utilizados é mínima.

Tabela 2.2 – Média das 10 replicas de cada cenário para a herdabilidade verdadeira ( $h^2$ ), estimada ( $\hat{h}^2$ ) e o erro quadrático médio (EQM).

Com Seleção				Sem Seleção			
Cenário	$h^2$	$\hat{h}^2$	EQM	Cenário	$h^2$	$\hat{h}^2$	EQM
CS10_M1	0,397	0,172	0,0075	SS10_M1	0,408	0,359	0,0005
CS10_M2	0,390	0,407	0,0004	SS10_M2	0,402	0,418	0,0008
CS40_M1	0,387	0,224	0,0071	SS40_M1	0,404	0,400	0,0002
CS40_M2	0,389	0,418	0,0003	SS40_M2	0,397	0,418	0,0003
CS100_M1	0,379	0,247	0,0029	SS100_M1	0,399	0,418	0,0003
CS100_M2	0,383	0,416	0,0004	SS100_M2	0,399	0,421	0,0002

CS e SS = populações com e sem seleção; 10, 40 e 100= tamanho de grupo de avaliação;  
M1 e M2 = modelo sem e com grupo contemporâneo como efeito fixo, respectivamente.

Na Tabela 2.3, são apresentadas as correlações entre os valores genéticos estimados (VGEs) e os verdadeiros (VGVs), para os cenários simulados. A inclusão do efeito de grupo contemporâneo no modelo aumentou a acurácia de predição dos valores genéticos nas duas populações (com e sem seleção) a menor diferença entre os modelos foi observada para os cenários SS100, concordando como os resultados obtidos na estimação de componentes de variância. O tamanho de grupo teve influência na acurácia porque, como indicado anteriormente, grupos pequenos têm maior heterogeneidade de variâncias fenotípicas, o que leva principalmente a perda de acurácia e incremento do viés nas predições de valores genéticos (VISSCHER; HILL, 1992; CARVALHEIRO et al., 2002).

Tabela 2.3 – Média das 10 replicas de cada cenário para as correlações entre os valores genéticos verdadeiros e os estimados ( $r_{aa}$ ) com o erro padrão (EP).

Com seleção			Sem Seleção		
Cenário	$r_{aa}$	EP	Cenário	$r_{aa}$	EP
CS10_M1	0,576	0,0014	SS10_M1	0,612	0,0007
CS10_M2	0,652	0,0018	SS10_M2	0,656	0,0012
CS40_M1	0,631	0,0011	SS40_M1	0,676	0,0008
CS40_M2	0,688	0,0014	SS40_M2	0,694	0,0010
CS100_M1	0,651	0,0011	SS100_M1	0,702	0,0009
CS100_M2	0,700	0,0011	SS100_M2	0,715	0,0008

CS e SS = populações com e sem seleção; 10, 40 e 100= tamanho de grupo;  
M1 e M2 = modelo sem e com grupo contemporâneo como efeito fixo.

A eficiência da seleção foi avaliada nos cenários com seleção, por meio de dois critérios, o primeiro, a porcentagem de coincidência entre os melhores machos candidatos a seleção e os selecionados (Tabela 2.4). O segundo foi a porcentagem de ganho genético obtido com referência ao máximo ganho possível (Tabela 2.5). Para o primeiro critério, de uma forma geral, a porcentagem de coincidência foi baixa e, como esperado, aumentou quando foi avaliada em um número maior de melhores machos. Quando o tamanho de grupo de avaliação aumentou, a coincidência foi maior, indicando que os valores genéticos foram estimados com maior acurácia, como apresentado na Tabela 2.3. Novamente existiram diferenças entre os modelos, sendo que o modelo que considerou o efeito fixo de grupo contemporâneo apresentou maiores porcentagens de coincidência, como consequência da correção para o efeito da seleção feito ao considerar o efeito fixo de grupo contemporâneo.

Tabela 2.4 – Média das 10 replicas dos cenários com seleção, para a porcentagem de coincidência entre os melhores machos candidatos a seleção com os selecionados em cada ciclo de acasalamentos

Cenário	% Coincidência Melhores Machos		
	10	20	30
CS10_M1	10,73	18,00	23,27
CS10_M2	16,64	27,82	34,36
CS40_M1	14,18	21,64	29,91
CS40_M2	21,64	32,00	39,82
CS100_M1	17,73	28,55	36,55
CS100_M2	21,45	33,27	43,27

CS = populações com seleção; 10, 40 e 100= tamanho de grupo; M1 e M2= modelo sem e com grupo contemporâneo como efeito fixo.

A porcentagem de ganho genético verdadeiro (Tabela 2.5) foi alta, apesar da porcentagem de coincidência dos machos selecionados ter sido baixa, indicando que os machos escolhidos foram de alto valor genético, mesmo sem ser os melhores. O ganho genético estimado ficou próximo do verdadeiro quando o modelo utilizado considerou o efeito de grupo contemporâneo, enquanto o modelo sem o efeito fixo subestimou o ganho genético estimado com referência ao verdadeiro, como consequência da menor variância genética estimada ao não corrigir para o efeito da seleção (Figura 2.2). O tamanho do grupo de avaliação influenciou a resposta à seleção, sendo maior a resposta quando maiores tamanhos de grupos de avaliação são utilizados, isto era esperado uma vez que a acurácia de predição foi maior para estes cenários.

Tabela 2.5 – Média das 10 replicas dos cenários com seleção, para a eficiência da seleção medida no ganho genético verdadeiro e estimado, expressada como porcentagem do máximo ganho genético possível

Cenário	% Ganho genético	
	$\Delta G$ Verdadeiro	$\Delta G$ Estimado
CS10_M1	51,39	3,50
CS10_M2	61,99	63,59
CS40_M1	58,16	5,38
CS40_M2	66,34	71,38
CS100_M1	65,08	6,76
CS100_M2	70,33	78,08

CS = populações com seleção; 10, 40 e 100= tamanho de grupo; M1 e M2= modelo sem e com grupo contemporâneo como efeito fixo.

Os resultados indicam que a atribuição dos escores visuais, com referência ao perfil médio de animais avaliados, corrige para os efeitos ambientais utilizados na formação do grupo, porém, o efeito fixo de grupo contemporâneo deve ser mantido nos modelos, porque este permite corrigir para o efeito da seleção, levando à estimação de valores genéticos mais acurados, e por conseguinte, maior progresso genético. Em dados reais o efeito de grupo contemporâneo também corrige para o efeito de avaliador, uma vez que os grupos não são avaliados pelo mesmo avaliador. O tamanho de grupo de avaliação influenciou principalmente a acurácia de estimação dos valores genéticos, sendo desejável a formação de grupos maiores que 40 animais para obter valores genéticos mais acurados.

## 2.4 Conclusões

As estimativas de parâmetros e os valores genéticos não são afetadas pela atribuição dos escores com referência ao grupo de avaliação, no entanto, deve-se ter cuidado quando grupos de avaliação com menos de 40 animais são utilizados, uma vez que estes diminuem a acurácia de predição dos valores genéticos.

## REFERÊNCIAS

- ABDEL-AZIM, G. A.; BERGER, P. Properties of threshold model predictions. **Journal of Animal Science**, v. 77, n. 3, p. 582–590, 1999.
- AFOLAYAN, R. A.; PITCHFORD, W. S.; DELAND, M.; MCKIERNAN, W. Breed variation and genetic parameters for growth and body development in diverse beef cattle genotypes. **Animal**, v. 1, n. 1, p. 13–20, 2007.
- BARICHELO, F.; ALENCAR, M.; JÚNIOR, R. T. et al. Herdabilidade e correlações quanto a peso, perímetro escrotal e escores visuais à desmama, em bovinos Canchim. **Pesquisa Agropecuária Brasileira**, v. 45, n. 6, p. 563–570, 2010.
- BERTIPAGLIA, T. S.; CARREÑO, L. O. D.; MACHADO, C. H. C.; ANDRIGHETTO, C.; FONSECA, R. d. Estimates of genetic parameters for visual scores and their correlation with production and reproductive traits in Brahman cattle. **Revista Brasileira de Zootecnia**, v. 41, n. 6, p. 1407–1411, 2012.
- CARVALHEIRO, R.; FRIES, L. A.; SCHENKEL, F. S.; ALBUQUERQUE, L. G. Efeitos da heterogeneidade de variância residual entre grupos de contemporâneos na avaliação genética de bovinos de corte. **Revista Brasileira de Zootecnia**, v. 31, n. 4, p. 1680–1688, 2002.
- CASELLA, G.; BERGER, R. L. **Statistical inference**. [S.l.]: Duxbury Pacific Grove, CA, 2002. v. 2.
- DUITAMA, L. O.; FARAH, M.; UTSONOMIYA, A. H.; ONO, R. K.; PIRES, M.; da FONSECA, R. Uso de modelos lineares mistos na avaliação genética de escores visuais: estudo de simulação. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, v. 66, n. 4, p. 1239–1246, 2014.
- FALCONER, D. S.; MACKAY, T. F. **Introduction to Quantitative Genetics**. 4th. ed. England: PERSON Prentice Hall, 1996. 464 p.
- FARIA, C. U.; MAGNABOSCO, C. U.; ALBUQUERQUE, L. G.; 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.
- FERNANDO, R. L.; BILLINGSLEY, R.; GIANOLA, D. Effects of method of scaling on heritability estimates and sire evaluations for frame size at weaning in Angus cattle. **Journal of Animal Science**, v. 56, n. 5, p. 1047–1056, 1983.
- FRIES, L. Uso de escores visuais em programas de seleção para produtividade em gado de corte. In: SEMINÁRIO NACIONAL - REVISÃO DE CRITÉRIOS DE SELEÇÃO E JULGAMENTO EM GADO DE CORTE, 1996, Uberaba - MG. **Anais ... Uberaba - MG: ABCZ**, 1996.
- GEWEKE, J. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In: \_\_\_\_\_. **Bayesian Statistics 4: Proceedings of the Fourth Valencia International Meeting**. 4. ed. [S.l.]: Oxford University Press, USA, 1992.

GORDO, D. M.; BALDI, F.; LÔBO, R. B.; FILHO, W. K.; SAINZ, R. D.; ALBUQUERQUE, L. 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.

HEIDELBERGER, P.; WELCH, P. D. Simulation run length control in the presence of an initial transient. **Operations Research**, v. 31, n. 6, p. 1109–1144, 1983.

HENDERSON, C. R. Sire evaluation and genetic trends. **Journal of Animal Science**, Symposium, p. 10–41, 1973.

HILL, W. On selection among groups with heterogeneous variance. **Animal Production**, v. 39, n. 03, p. 473–477, 1984.

JÚNIOR, J. J.; PITA, F. V.; FRIES, L. A.; ALBUQUERQUE, L. G. Influência de alguns fatores de ambiente sobre os escores de conformação, precocidade e musculatura à desmama em um rebanho da raça Nelore. **Revista Brasileira de Zootecnia**, v. 30, n. 6, p. 1697–1703, 2001.

KOURY FILHO, W.; ALBUQUERQUE, L. G.; ALENCAR, M.; FORNI, S.; SILVA, J.; LÔBO, R. B. Estimativas de herdabilidade e correlações para escores visuais, peso e altura ao sobreano em rebanhos da raça Nelore. **Revista Brasileira de Zootecnia**, 2009, v. 38, n. 12, p. 2362–2367, 2009.

KOURY FILHO, W.; ALBUQUERQUE, L. G.; FORNI, S.; SILVA, J.; YOKOO, M. J.; ALENCAR, M. M. de. Estimativas de parâmetros genéticos para os escores visuais e suas associações com peso corporal em bovinos de corte. **Revista Brasileira de Zootecnia**, v. 39, n. 5, p. 1015–1022, 2010.

MAIWASHE, A.; BRADFIELD, M.; THERON, H.; WYK, J. V. Genetic parameter estimates for body measurements and growth traits in South African Bonsmara cattle. **Livestock Production Science**, v. 75, n. 3, p. 293–300, 2002.

MEIJERING, A.; GIANOLA, D. Linear versus nonlinear methods of sire evaluation for categorical traits: simulation study. **Genetics Selection Evolution**, v. 17, n. 1, p. 115–132, 1985.

MISZTAL, I.; TSURUTA, S.; STRABEL, T.; AUVRAY, B.; DRUET, T.; LEE, D. et al. BLUPF90 and related programs (BGF90). In: WORLD CONGRESS ON GENETICS APPLIED TO LIVESTOCK PRODUCTION, 7th., 2002, Montpellier, France. **Proceedings ...** [S.l.]: Institut National de la Recherche Agronomique (INRA), 2002. p. 1–2.

MRODE, R. A. **Linear models for the prediction of animal breeding values**. [S.l.]: Cabi, 2005.

PLUMMER, M.; BEST, N.; COWLES, K.; VINES, K. Coda: Convergence diagnosis and output analysis for mcmc. **R News**, v. 6, n. 1, p. 7–11, 2006.

SARGOLZAEI, M.; SCHENKEL, F. S. QMSim: a large-scale genome simulator for livestock. **Bioinformatics**, v. 25, n. 5, p. 680–681, 2009.



SHIOTSUKI, L.; SILVA, J.; TONHATI, H.; ALBUQUERQUE, L. G. Genetic associations of sexual precocity with growth traits and visual scores of conformation, finishing, and muscling in Nelore cattle. **Journal of Animal Science**, v. 87, n. 5, p. 1591–1597, 2009.

SHORT, T.; LAWLOR, T. Genetic parameters of conformation traits, milk yield, and herd life in Holsteins. **Journal of Dairy Science**, v. 75, n. 7, p. 1987–1998, 1992.

SNELLING, W. M.; CHIU, R.; SCHEIN, J. E.; HOBBS, M.; ABBEY, C. A.; ADELSON, D. L.; AERTS, J.; BENNETT, G. L.; BOSDET, I. E.; BOUSSAHA, M. et al. A physical map of the bovine genome. **Genome Biology**, v. 8, n. 8, p. R165, 2007.

VISSCHER, P.; HILL, W. Heterogeneity of variance and dairy cattle breeding. **Animal Production**, v. 55, n. 03, p. 321–329, 1992.

WANG, C.; QUAAS, R.; POLLAK, E. Bayesian analysis of calving ease scores. **Genetics Selection Evolution**, v. 29, p. 117–143, 1997.

YOKOO, M. J.; WERNECK, J. N.; PEREIRA, M. C.; ALBUQUERQUE, L. G.; FILHO, W. K. Correlações genéticas entre escores visuais e características de carcaça medidas por ultrassom em bovinos de corte. **Pesquisa Agropecuária Brasileira**, v. 44, n. 2, p. 197–202, 2009.

ZINK, V.; ŠTÍPKOVÁ, M.; LASSEN, J. Genetic parameters for female fertility, locomotion, body condition score, and linear type traits in Czech Holstein cattle. **Journal of Dairy Science**, v. 94, n. 10, p. 5176–5182, 2011.



### **CAPÍTULO 3 - ESTUDO DE ASSOCIAÇÃO GENÔMICA PARA ESCORES VISUAIS MENSURADOS À DESMAMA EM BOVINOS NELORE**

**RESUMO:** Os estudos de associação genômica ampla (GWAS) são amplamente utilizados em bovinos para identificar regiões ou variantes genéticas que estão associadas com os fenótipos de interesse, e assim, desenhar estratégias que permitam aumentar a frequência dos alelos favoráveis. Os escores visuais são características importantes na bovinocultura do Brasil, porque são utilizados como critério de seleção, ajudando a escolher animais mais harmoniosos. Apesar de sua importância, ainda não há estudos de associação genômica para estas características. O objetivo do presente estudo foi identificar as regiões do genoma associadas com as características de conformação, precocidade e musculabilidade, mensuradas à desmama. Para o GWAS foram utilizados fenótipos de 2873 animais da raça Nelore com os modelos BayesC e LASSO Bayesiano. Os animais foram genotipados com o painel BovineHD da Illumina, sendo que, após o controle de qualidade um total de 309.865 SNPs foram utilizados. Nas análises, o fenótipo e os valores genéticos desregredidos foram utilizados como variável dependente, no primeiro caso foi utilizado um modelo “threshold”, e no segundo foi utilizado um modelo linear. As análises foram feitas no pacote BGLR do software R, usando cadeias de 600.000 ciclos após 200.000 ciclos de “burn-in”. O critério de associação foi a porcentagem de variância genética explicada pelos SNPs contidos em janelas de 1 Mb de comprimento. O modelo BayesC foi o que se ajustou melhor aos dados porque conseguiu explicar uma maior proporção de variância fenotípica nas duas variáveis dependentes. De acordo com os resultados não houve janelas com grandes efeitos para os escores visuais, indicando que são de natureza poligênica. No entanto, foram identificadas nos cromossomos 1, 3, 5, 7, 14, 15, 16, 19, 20 e 23 regiões que explicaram maior variância genética para os escores visuais de conformação, precocidade e musculabilidade.

**Palavras-chave:** BayesC, genes crescimento, LASSO Bayesiano

### 3.1 Introdução

Com o sequenciamento do genoma bovino (ELSIK et al., 2009), novas informações se tornaram disponíveis para auxiliar a predição de valores genéticos através da seleção genômica (GS) e para localizar regiões ou genes associados com o fenótipo de interesse, utilizando os estudos de associação genômica ampla (GWAS) (HAYES; GODDARD, 2010; DEKKERS, 2012).

Nas análises de GWAS os modelos de regressão simples são frequentemente utilizados, no entanto, este método apresenta duas limitações. A primeira é a sobrestimação da proporção de variância explicada pelo marcador, uma vez que não leva em consideração o LD existente entre os marcadores (GIANOLA; VERRIER; VERRIER, 2013) e a segunda é a alta taxa de falsos positivos se a estrutura da população não é levada em consideração (HAYES et al., 2010). Uma solução a estas dificuldades são os modelos Bayesianos de regressão simultânea dos SNPs, propostos inicialmente para GS (MEUWISSEN; HAYES; GODDARD, 2001). Entre os modelos Bayesianos mais usados em GWAS encontram-se o LASSO Bayesiano (PARK; CASELLA, 2008) e BayesC (HABIER et al., 2011). Os dois modelos diferem na distribuição *a priori* para os efeitos dos QTLs, o primeiro, assume que existem poucos genes de grande efeito e muitos com efeito pequeno ou zero, enquanto o segundo, assume que a grande maioria dos SNPs não estão associados com o fenótipo, e só uma pequena porção  $\pi$  tem algum efeito sobre a característica. O melhor modelo a ser usado depende da arquitetura genética da característica, a qual não é conhecida, principalmente quando as características têm sido pouco estudadas (GONZÁLEZ-RECIO; FORNI, 2011).

Em bovinos, estudos de associação foram conduzidos para a maioria de características de importância econômica (GODDARD; HAYES, 2009; HAYES et al., 2010; BOLORMAA et al., 2011; HAWKEN et al., 2012; UTSUNOMIYA et al., 2014). No entanto, características como os escores visuais, que são de relevância na pecuária brasileira, ainda não têm sido estudadas com as novas informações e metodologias disponíveis. Estas características são usadas como critérios de seleção pelos produtores, porque ajudam na identificação de animais harmoniosos e estão correlacionadas com características de interesse econômico, principalmente de crescimento (FARIA et al., 2008; KOURY FILHO et al., 2010; BOLIGON; ALBUQUERQUE, 2010). Um dos objetivos dos escores é servir como critério de seleção para melhorar a qualidade da carcaça, porém, os estudos de genética quantitativa mostram baixas correlações genéticas entre os escores visuais e estas características (YOKOO et al., 2009; FARIA et al., 2009; GORDO et al., 2012; GORDO, 2014).

Para características morfológicas, estudos de associação têm sido feitos em

gado leiteiro para as avaliações lineares de tipo, que avaliam vários caracteres exteriores do animal e índices que ponderam as características avaliadas (KOLBEHDARI et al., 2008; ALAM et al., 2011; COLE et al., 2011; BOLORMAA et al., 2011; WU et al., 2013). De forma geral, para estas características não foram identificadas regiões de grande efeito, indicando que são de natureza poligênica, controladas por múltiplos genes de pequeno efeito, como mostrado por Hayes et al. (2010). Conhecer a distribuição dos efeitos dos QTLs para as características é importante porque é um dos fatores que influenciam a acurácia de predição de valores genéticos genômicos em seleção genômica, sendo que a acurácia será maior se existem QTLs de grande efeito (GODDARD, 2009; HAYES et al., 2010).

Dada a importância dos escores visuais na pecuária brasileira, por serem usados como critérios de seleção, é desejável que as regiões envolvidas no controle genético tenham efeitos pleiotrópicos sobre outras características de importância econômica e que existam QTL de grande efeito, levando assim, a maiores ganhos genéticos quando usados como critério de seleção. Este trabalho objetivou identificar as regiões do genoma que estão associadas com os escores visuais de conformação (C), precocidade (P) e musculosidade (M) mensurados à desmama em bovinos Nelore e, comparar o usando dos modelos BayesC e LASSO Bayesiano em estudo de associação genômica.

### **3.2 Material e métodos**

Os dados são provenientes de animais da raça Nelore, machos e fêmeas nascidos entre 2007 e 2011, pertencentes a dois programas de melhoramento genético, DeltaGen e Paint. O fenótipo para os escores visuais foi avaliado a desmama de acordo com o seguinte procedimento: primeiramente observou-se todo o lote, que caracteriza o grupo de contemporâneos, visualizando-se o perfil médio do lote para cada característica avaliada, o que serviu de base de comparação para atribuir os escores, que variaram de 1 a 5, sendo 5 a maior expressão e 1 a menor expressão da característica. Segundo Júnior et al. (2001) cada escore objetiva: Conformação (C), avaliar o animal em termos de produção de carne se abatido naquele momento, sendo que o tamanho e o grau de musculosidade influenciam o valor do escore. Precocidade (P), avalia a capacidade do animal em acumular a quantidade de gordura mínima necessária para o abate com peso vivo não elevado, sendo observado no momento da avaliação a profundidade de costelas e a deposição de gordura na virilha e na cauda do animal. Musculosidade (M), avalia a quantidade de massa muscular, usando como referência o desenvolvimento muscular na paleta, antebraço, lombo, garupa e traseiro.

Um total de 2021 fêmeas e 1416 machos com registros para C, P e M com informação genotípica foram utilizados nas análises. Para os estudos de associação genômica o fenótipo do escore visual e os valores genéticos desregredidos (dEBV) foram usados como variável dependente. No primeiro caso, após o controle de qualidade 2873 registros foram utilizados, enquanto no segundo caso, o número de registros foi menor (Tabela 3.1), porque na desregressão dos EBVs só foram considerados animais com acurácia acima de 0,7.

Tabela 3.1 – Número de registros usados por variável dependente nas análises de associação genômica, para os escores visuais de conformação (C), precocidade (P) e musculosidade (M)

Escore	Variável Dependente	Registros			Média/ Moda
		Machos	Fêmeas	Total	
C	Fenótipo	1977	896	2873	3
	dEBV	1266	601	1867	0.2327
P	Fenótipo	1977	896	2873	3
	dEBV	1132	569	1701	0.2938
M	Fenótipo	1977	896	2873	3
	dEBV	1118	553	1671	0.3543

Para a estimação dos valores genéticos (EBVs) foi utilizado um modelo “threshold” como descrito por Mrode (2005). O modelo considerou os efeitos do animal e o resíduo como aleatórios, enquanto que os efeitos fixos incluíram o grupo contemporâneo (grupo de manejo a desmama, fazenda, ano de nascimento e sexo) e a idade a mensuração do escore, considerada como covariável. Para a estimação de parâmetros e valores genéticos foi utilizado o software THRGIBBS1F90 que implementa o modelo de limiar sob estimação Bayesiana (MISZTAL et al., 2002). Um total de 177.520 animais nascidos entre 1985 e 2012 foram incluídos na matriz de parentesco. A acurácia dos EBVs foi calculada como descrito por González-Recio et al. (2006) e desregredidos usando a metodologia descrita por Garrick et al. (2009).

### 3.2.1 Dados genotípicos

Os animais foram genotipados com o painel BovineHD BeadChip de alta densidade (Illumina, San Diego, CA, USA), segundo protocolo da Illumina. O BovineHD BeadChip contém 777.962 marcadores do tipo SNP espalhados pelo genoma, com uma distância média entre marcadores de 3,43 kb. Os critérios de inclusão usados para o controle de qualidade dos genótipos foram: SNPs localizados nos autossomos, “call

rate” por SNP superior a 0,95; “call rate” por animal superior a 0,90; “minor allele frequency” maior que 0,05; p-value para o teste de equilíbrio de Hardy-Weinberg (HWE) maior que  $1 \times 10^{-5}$  (desvios extremos do equilíbrio sugerem potenciais erros de genotipagem); SNPs altamente correlacionados ( $r^2 > 0,98$ ) foram excluídos. O processo foi realizado iterativamente até que nenhum SNP nem amostra fossem excluídas, resultando em 309.865 SNPs para as análises que consideraram o fenótipo como variável dependente, e de 308.861, 308.561 e 308.481 SNPs para C, P e M, respectivamente, quando a variável dependente foi o dEBV. O controle de qualidade e a imputação dos genótipos faltantes foram realizados com software R usando o pacote snpStats (CLAYTON, 2013).

### 3.2.2 Análises de associação

Para estimar os efeitos dos marcadores foram utilizados os modelos BayesC e LASSO Bayesiano, estes dois modelos diferem nas distribuições *a priori* assumidas para os efeitos dos marcadores. De forma geral, os modelos podem ser representados em forma matricial como:

$$y = X\beta + Z\alpha + e$$

em que  $\mathbf{y}$  é um vetor  $n \times 1$  de fenótipos ou dEBVs de  $n$  animais para os escores visuais de C, P e M;  $\mathbf{X}$  é uma matriz  $n \times p$  que relaciona o vetor  $\beta$  de  $p$  efeitos fixos com  $\mathbf{y}$ ;  $\mathbf{Z}$  é uma matriz  $n \times k$  de genótipos (0 para o primeiro homocigoto, 1 para o heterocigoto e 2 para o segundo homocigoto) de  $k$  SNPs;  $\alpha$  é um vetor  $k \times 1$  de coeficientes aleatórios de regressão para os SNPs (efeito de substituição alélica), e  $\mathbf{e}$  é um vetor de resíduos com uma distribuição Normal  $N \sim (0, I\sigma_e^2)$ , onde  $\sigma_e^2$  é a variância do resíduo, considerada como desconhecida com uma distribuição *a priori* Qui-quadrado escalonada invertida (FERNANDO; GARRICK, 2013).

A diferença entre os modelos está na distribuição *a priori* marginal assumida para  $\alpha$ , que determina o tipo de “shrinkage” nas estimativas dos efeitos dos SNPs. Para o LASSO, a distribuição assumida para os efeitos é dupla exponencial. Esta distribuição tem maior densidade em zero e caudas mais grossas que uma distribuição Normal, o que leva a um “shrinkage” efeito-dependente, porque SNPs com efeitos pequenos são regredidos para o zero com maior força que aqueles SNPs com grande efeito (PARK; CASELLA, 2008; GIANOLA, 2013). O grau do “shrinkage” é controlado pelo hiperparâmetro  $\lambda$ , o qual foi inferido a partir dos dados, usando uma distribuição Gamma *a priori*. Para BayesC a distribuição dos efeitos *a priori* é uma mistura de um ponto de massa em zero e uma distribuição Normal (HABIER et al., 2011). A proporção de SNPs com efeito diferente do zero é controlada pelo hiperparâmetro  $\pi$ , o qual foi fixado em 0,01 ao invés de ser inferido a partir dos dados, uma vez que Berg, Fritz

e Boichard (2013) encontraram problemas de convergência quando o  $\pi$  foi inferido a partir dos dados.

Quando o fenótipo dos escores foi considerado como variável dependente, um modelo “threshold” com função de ligação *probit* foi utilizado. O modelo assume que há uma variável aleatória subjacente ao fenótipo observável, conhecida como “liability”, que segue uma distribuição Normal padrão. Para facilitar a estimação, a variância residual foi fixada em 1. Os efeitos fixos considerados no modelo foram: o grupo contemporâneo (formado pelo grupo de manejo à desmama, fazenda, ano de nascimento e sexo) e à idade a desmama considerada como covariável. Quando os dEBV foram a variável dependente, um modelo linear foi utilizado e só o efeito da média foi considerado no modelo como efeito fixo.

As análises foram realizadas com o pacote BGLR do software R (de los CAMPOS; PEREZ, 2013), que implementa “Gibbs sampler” para amostrar as distribuições *a posteriori* dos parâmetros. Foram amostradas cadeias de 800.000 ciclos, dos quais os primeiros 200.000 foram descartados (“Burn-in”), restando 600.000 amostras para a inferência dos parâmetros.

### 3.2.3 Critério de associação

Com o uso de um chip de alta densidade o efeito de um QTL pode ser distribuído ao longo de vários SNPs que estejam em LD com o QTL, resultando em efeitos de SNPs individuais não significativos (BODDICKER et al., 2012; ONTERU et al., 2012), por tal motivo os SNPs foram agrupados em janelas de 1 Mb de comprimento, sobrepostas a cada 100 Kb, totalizando 25.250 janelas. A quantidade de SNPs por janela variou de 1 até 336, com média de  $122 \pm 25,2$  SNPs. O critério usado para identificar janelas associadas com os escores foi a porcentagem de variância genética explicada por cada janela, a qual foi calculada como segue.

$$\% \sigma_j^2 = \frac{\sigma_j^2}{\sigma_{\text{SNP}}^2} \times \frac{h^2}{h_{\text{SNP}}^2}$$

em que:  $\% \sigma_j^2$  é a porcentagem de variância genética explicada pela janela  $j$ ;  $\sigma_j^2$  é a variância genética da janela  $j$ ;  $\sigma_{\text{SNP}}^2$  é a variância genética explicada pelos SNPs;  $h^2$  é a herdabilidade da característica e  $h_{\text{SNP}}^2$  é a proporção de variância fenotípica explicada pelos marcadores (herdabilidade dos marcadores).  $\sigma_j^2$  e  $\sigma_{\text{SNP}}^2$  foram calculadas como a variância dos valores genéticos genômicos (GEBV) de cada janela ou de todo o genoma em todos os animais, respectivamente. O GEBV para o animal  $i$  na janela  $j$  foi calculado como:

$$\text{GEBV}_{ij} = \sum_{k=1}^K X_{ik} \hat{\alpha}_k$$



em que:  $k$  é o número de SNPs contidos na janela  $j$ ;  $x_{ik}$  é o genótipo do animal  $i$  para o SNP  $k$  e  $\hat{\alpha}_k$  é o efeito de substituição alélica para o SNP  $k$ . Para janelas de 1 SNP, este método é equivalente a  $2p_k(1 - p_k)\hat{\alpha}_k^2$  (FALCONER; MACKAY, 1996). O modelo que explicou a maior proporção de variância fenotípica foi escolhido para identificar as regiões do genoma associadas com o fenótipo, sendo que para as janelas que explicaram valores superiores a 0,25% da variância genética aditiva foram identificados os genes e proposto um gene candidato, como responsável pela variância explicada pela janela.

### 3.2.4 Identificação dos genes

Para a identificação dos genes contidos nas janelas foi utilizada a bases de dados “National Center for Biotechnology Information” NCBI <http://www.ncbi.nlm.nih.gov/snp/> carregada com o genoma bovino versão UMD 3.1. A classificação dos genes quanto à função biológica foi realizada por meio da base de dados “The Database for Annotation, Visualization and Integrated Discovery” DAVID (DENNIS JR et al., 2003) disponível em <http://david.abcc.ncifcrf.gov/>.

## 3.3 Resultados e discussão

De forma geral, a proporção de variância fenotípica explicada pelos SNPs foi menor que a herdabilidade da característica (Tabela 3.2), diferença conhecida como herdabilidade perdida (MANOLIO et al., 2009). Esta diferença pode ser explicada por vários fatores, entre eles a falta de LD entre os marcadores e os QTLs. Outra causa é o tamanho da amostra, uma vez que para capturar QTLs de pequenos efeitos é necessário aumentar o número de animais genotipados, como mostrado por Wood et al. (2014) que capturaram maior variância fenotípica na medida que aumentou o tamanho da amostra.

Entre os modelos utilizados o BayesC foi o que maior variância fenotípica explicou, indicando que conseguiu descrever melhor a distribuição dos efeitos dos marcadores sobre a variável dependente. Entre as variáveis dependentes, o fenótipo foi superior aos dEBV, apesar dos valores genéticos estimados (EBV) serem considerados por alguns autores como a melhor variável dependente para análises de GWAS (BECKER et al., 2013) uma vez que estes são a melhor estimativa do verdadeiro valor genético. A menor variância fenotípica capturada pelos modelos que usaram os dEBV pode ser consequência da baixa acurácia na estimação, uma vez que a maioria dos animais utilizados tinham pouca ou nenhuma descendência, o que levou a que 75% dos EBVs utilizados tiveram uma acurácia entre 0,7 e 0,8.

Tabela 3.2 – Proporção de variância fenotípica explicada pelos modelos BayesC e LASSO Bayesiano, e número de janelas que explicam 10% da variância genética para os escores visuais de conformação (C), precocidade (P) e musculosidade (M)

Escore	$h^2$	Modelo	Variável depend.	$\sigma_{\text{SNP}}^2$	$\sigma_e^2$	$h_{\text{SNP}}^2$	10% $\sigma_a^2$
C	0,44	BayesC	dEBV	1,031	0,619	0,275 [0,625]	139
			Fenótipo	0,503	1,00	0,335	87
		LASSO	dEBV	0,970	0,711	0,254 [0,577]	175
			Fenótipo	0,412	1,00	0,292	101
P	0,43	BayesC	dEBV	0,992	0,735	0,247 [0,574]	172
			Fenótipo	0,390	1,00	0,281	82
		LASSO	dEBV	0,824	0,921	0,203 [0,472]	198
			Fenótipo	0,322	1,00	0,244	99
M	0,42	BayesC	dEBV	0,902	0,753	0,229 [0,545]	159
			Fenótipo	0,416	1,00	0,294	86
		LASSO	dEBV	0,850	0,834	0,212 [0,504]	205
			Fenótipo	0,355	1,00	0,262	104

dEBV= valores genéticos estimados desregredidos;  $h^2$ = herdabilidade estimada usando modelo “threshold”;  $\sigma_{\text{SNP}}^2$ = variância explicada pelos SNPs;  $\sigma_e^2$ = variância residual;  $h_{\text{SNP}}^2$  = proporção de variância fenotípica explicada pelos SNPs, valores entre colchetes [ ] indicam a proporção com referência à variância genética aditiva; 10%  $\sigma_a^2$  = número de janelas que capturam 10% da variância genética.

O grande número de janelas necessárias para explicar 10% da variância genética (Tabela 3.2), indica que não existem QTLs de grande efeito, assim, os escores visuais avaliados em bovinos Nelore são de natureza poligênica, influenciados por muitos genes de pequeno efeito. Como relatado para características morfológicas avaliados em gado leiteiro, onde não têm sido identificadas regiões de grande efeito e na maioria dos casos as regiões identificadas não são comuns entre os estudos (KOLBEHDARI et al., 2008; ALAM et al., 2011; COLE et al., 2011; WU et al., 2013).

Para identificar as regiões associadas com os escores visuais foi escolhido o modelo BayesC com o fenótipo como variável dependente, uma vez que, este foi o que explicou a maior proporção de variância fenotípica nos três escores visuais. Nas Tabelas 3.3, 3.4, 3.5 e Figura 3.1 são apresentadas as janelas que explicaram porcentagens superiores a 0,25% da variância genética aditiva, para conformação, precocidade e musculosidade. Dado o grande número de janelas necessárias para explicar 10% da variância genética, o número de janelas que superaram o limiar de 0,25% foram poucas, motivo pelo qual, janelas que estiveram muito próximas do limiar também foram incluídas. Para estas janelas foram identificados os genes e proposto genes

candidatos responsáveis pela variância genética explicada pela janela.

Um total de 191 genes foram localizados nas 14 janelas associadas com o fenótipo dos escores visuais, 150 foram genes que codificaram proteínas, 16 ncRNA (RNA não codificante), 4 miRNA (microRNA), 6 tRNA (RNA transportador) e 15 pseudogenes. Na identificação dos genes candidatos foram considerados só genes que codificam proteínas. Assim, foram identificados 18 genes candidatos, classificados em 4 grupos de acordo com a função que desempenham no organismo: metabolismo basal e celular (TBC1D5, LPAR2, TMEM9, NDUFA13, GSTA3 e FBXL17), regulação e transcrição de outros genes (CREG1, POU2F1, MAPK11, MAPK12, ESRRG e STK3), metabolismo de lípidos (SORL1 e ASCF2) e genes relacionados com o crescimento e o músculo esquelético (C1QTNF3, PPP1R39, WFIKKN2 e IGFBP6).

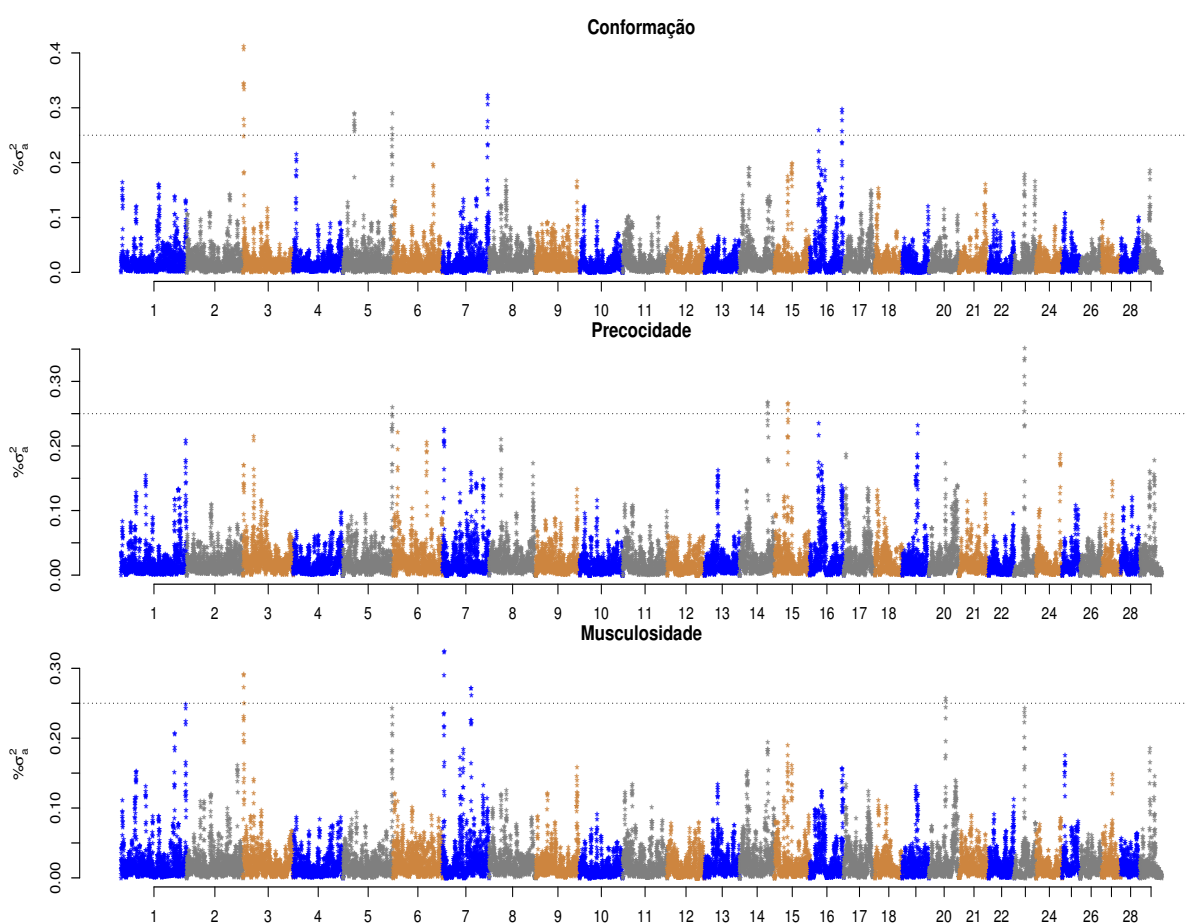


Figura 3.1 – Porcentagem da variância genética explicada pelas janelas do modelo BayesC, com o fenótipo como variável dependente para os escores visuais de conformação, precocidade e musculosidade. A linha pontilhada marca o limiar de 0,25% de variância genética explicada pelas janelas

Tabela 3.3 – Identificação dos genes e porcentagem de variância genética explicada pelas janelas associadas com o escore visual de conformação, usando o modelo BayesC com o fenótipo como variável dependente

BTA	Mb	% $\sigma_a^2$	Genes	Candidato	Descrição
3	0,4-1,4	0,41	10 Proteínas 1 tRNA 3 Pseudogene	CREG1 POU2F1	Cellular repressor of E1A-stimulated; POU class 2 homeobox 1
5	26,7-27,7	0,29	30 Proteínas	IGFBP6	Insulin-like growth factor protein 6
5	118,7-119,7	0,29	3 Proteínas 4 ncRNA	MAPK11 MAPK12	Mitogen-activated protein kinase 11 e 12
7	108,8-109,8	0,32	2 Proteínas 1 tRNA 1 ncRNA	FBXL17	F-Box And Leucine-Rich Repeat Protein 17
16	20,7-21,7	0,26	2 Proteínas 1 tRNA	ESRRG	Estrogen-related receptor gamma
16	80,7-81,7	0,30	7 Proteínas	TMEM9	Transmembrane protein 9
Total		1,87			

BTA= cromossomo bovino, Mb= posição da janela em megabases; % $\sigma_a^2$ = % de variancia genética explicada por cada janela.

A presença de genes relacionados ao metabolismo basal e celular, localizados em janelas significativas é explicado pelo fato desses genes atuarem em vários tecidos, afetando o desenvolvimento celular e corporal como um todo, e assim, contribuindo para o desempenho geral. Esses genes têm um efeito pleiotrópico importante já que não são tecido-específicos e atuam nas mais diversas células do corpo e em diversas vias metabólicas que incluem: sinalização entre células, síntese e transporte de proteínas, proliferação e sobrevivência de células, transporte e formação da membrana celular e seus receptores, etc. Neste grupo de genes, encontra-se, por exemplo, a proteína do gene *NDUFA13* (conhecido também como *GRIM-19*) que é um componente funcional do complexo mitocondrial I, que está envolvido em processos de apoptose e produção de energia celular (FEARNLEY et al., 2001), modificações neste gene poderiam levar a uma menor disponibilidade de energia para os processos celulares, diminuindo o crescimento dos tecidos. Outro exemplo, é o gene *FBXL17* que intervém em processos de reciclagem de proteínas ao fazer parte do proteossoma 26S, atuando desta forma em diversos processos celulares. Sadkowski et al. (2006) encontraram diferentes níveis de expressão para este gene em músculo esquelético

de bovinos entre 8 e 12 meses de idade, indicando que está associado com o desenvolvimento muscular nestas idades.

Tabela 3.4 – Identificação dos genes e porcentagem de variância genética explicada pelas janelas associadas com o escore visual de precocidade, usando o modelo BayesC com o fenótipo como variável dependente

BTA	Mb	% $\sigma_a^2$	Genes	Candidato	Descrição
5	118,7-119,7	0,26	3 Proteínas 4 tRNA	MAPK11 MAPK12	Mitogen-activated protein kinase 11 e 12
14	67,5-68,5	0,27	11 Proteínas 1 tRNA 2 ncRNA	STK3	Serine/threonine kinase 3
15	31,7-32,7	0,27	5 Proteínas miRNA	SORL1	Sortilin-related receptor, L(DLR class)
16	20,7-21,7	0,24	2 Proteínas 1 tRNA	ESRRG	Estrogen-related receptor gamma
19	36,0-37,0	0,23	23 Proteínas 3 ncRNA 4 Pseudogene	WFIKKN2 ACSF2	Follistatin/kazal, immunoglobulin, kunitz and netrin domain; Hypothetical protein LOC768237
23	24,1-25,1	0,35	16 Proteínas 1 tRNA 2 ncRNA 2miRNA 3 Pseudogene	GSTA3	Glutathione S-transferase, alpha 3
Total		1,62			

BTA= cromossomo bovino, Mb= posição da janela em megabases; % $\sigma_a^2$ = % de variância genética explicada por cada janela.

Os genes que regulam a transcrição podem ser divididos em dois subgrupos, o primeiro formado pelos genes STK3, MAPK11 e MAPK12 que pertencem à família das quinases (subgrupo “serine/threonine-protein”), e o segundo que inclui os genes POU2F1, CREG1 e ESRRG, que têm como função principal modular a expressão genética (“transcription factor”). As quinases são enzimas que catalisam a fosforilação de proteínas, através da transferência de um grupo fosforila de ATP, mudando a configuração das proteínas o que ocasiona sua ativação ou desativação. As quinases são a maior família de proteínas em eucariotos e estão envolvidas em múltiplos processos celulares, incluindo mecanismos de sinalização celular e ativação de fatores de

transcrição (MANNING et al., 2002). Lee et al. (2011) encontraram que o gene STK3 (conhecido em humanos como MST2) atua como um repressor dos genes envolvidos na deposição de gordura em bovinos Hanwoo, uma vez que altos níveis de expressão deste gene estiveram associados com baixo nível de gordura intramuscular, mostrando que este gene inibe a proliferação de adipócitos. Os genes MAPK11 e MAPK12 fazem parte da via metabólica p38 MAPK que é um mecanismo preservado de resposta celular a uma ampla variedade de sinais extracelulares, sendo proposto como regulador da diferenciação, proliferação e desenvolvimento celular (SEGER; KREBS, 1995). O fato do p38 MAPK não ser tecido-específico explicaria porque esta região esteve associada com o fenótipo para conformação, precocidade e musculabilidade.

Tabela 3.5 – Identificação dos genes e porcentagem de variância genética explicada pelas janelas associadas com o escore visual de musculabilidade, usando o modelo BayesC com o fenótipo como variável dependente

BTA	Mb	% $\sigma_a^2$	Genes	candidato	Descrição
1	155,9-156,9	0,25	1 Proteínas	TBC1D5	TBC1 domain family, member 5
3	0,5-1,5	0,29	10 Proteínas 1 tRNA 3 Pseudogene	CREG1 POU2F1	Cellular repressor of E1A-stimulated; POU class 2 homeobox 1
5	118,7-119,7	0,24	3 Proteínas 4 ncRNA	MAPK11 MAPK12	Mitogen-activated protein kinase 11 e 12
7	3,0-4,0	0,32	23 Proteínas 5 Pseudogene 2 ncRNA	LPAR2 NDUFA13	Lysophosphatidic acid receptor 2; NADH dehydrogenase 1 alpha, 13
7	59,0-60,0	0,27	9 Proteínas 2 ncRNA	PPP1R39	SH3 domain ring finger 2
20	39,4-40,4	0,26	8 Proteínas 1 tRNA	C1QTNF3	C1q factor related protein 3
23	24,1-25,1	0,24	16 Proteínas 1 tRNA 2 ncRNA 2 miRNA 3 Pseudogene	GSTA3	Glutathione S-transferase alpha 3
Total		1,87			

BTA= cromossomo bovino, Mb= posição da janela em megabases; % $\sigma_a^2$ = % de variância genética explicada por cada janela.

No segundo subgrupo, os genes POU2F1 e CREG1 são fatores de transcrição que regulam a atividade de múltiplos genes que atuam em diversas vias metabólicas, entre eles, genes envolvidos nos processos de crescimento, proliferação e diferenciação celular. Pérez-Montarelo et al. (2014) identificaram POU2F1 como um dos fatores de transcrição que regula os genes associados com características de crescimento e deposição de gordura em suínos Ibericos x Landrace, enquanto que, Moreno-Sánchez et al. (2010) encontram diferença no nível de expressão gênica para CREG1 nos músculos *Psoas major* e *Flexor digitorum* em bovinos, indicando que os dois genes agem no músculo esquelético e tecido adiposo. O gene ESRRG é membro da família de receptores relacionados ao estrógeno (ESRR) que agem como ativadores de transcrição para o gene PERM1 envolvido no metabolismo energético do músculo esquelético e cardíaco (CHO et al., 2013), o que leva a que este gene possa estar associado como a deposição de tecido adiposo e muscular.

Os dois genes envolvidos no metabolismo de lípidos estiveram associados com o escore de precocidade, que avalia a capacidade do animal para depositar gordura. O gene SORL1 pertence à família de receptores de lipoproteínas de baixa densidade (LDLR) envolvidos no metabolismo do colesterol, enquanto que, ACSF2 está associado com o metabolismo de lípidos e a diferenciação de adipócitos. Estes dois genes têm sido associados com a deposição de gordura intramuscular em suínos e bovinos (WANG et al., 2013; NEUSTAETER et al., 2014).

Dos 4 genes relacionados com crescimento, 3 estiveram associados com precocidade e musculabilidade, sendo que a região do gene da miostatina (GDF8) não mostrou associação como os escores visuais. O gene da miostatina é importante no desenvolvimento muscular e é um dos responsáveis pela presença de musculatura dupla em raças como Belgian Blue e Limousin. O gene PPP1R39 (conhecido também como SH3RF2) ao igual que o gene da miostatina regula de forma negativa o crescimento do tecido muscular, uma vez que, baixos níveis de expressão estão associados com hipertrofia muscular. Rubin et al. (2010), estudando linhas de crescimento rápido e lento em frango de corte, encontram que para o gene SH3RF2, tinha sido fixada uma deleção na linha de crescimento rápido, sendo identificada como a causa do maior crescimento. Em bovinos, Boitard e Rocha (2013) encontraram que este gene esteve sob seleção na raça Blonde d'Aquitane, sendo um dos responsáveis pela presença de musculatura dupla nos animais.

O gene WFIKKN2 é uma proteína inibitória dos genes GDF8 (miostatina) e GDF11 implicados no desenvolvimento muscular, o que leva a que a sobreexpressão deste gene produza hipertrofia muscular, uma vez que o gene da miostatina regula de forma negativa o crescimento muscular, como observado em ratos e em ovinos

(MONESTIER et al., 2012; WANG et al., 2014). Os genes IGFBP6 e C1QTNF3 regulam de forma positiva o crescimento muscular em bovinos, sendo que altos níveis de expressão estão associados com crescimento celular no músculo esquelético. O gene C1QTNF3 também está envolvido na deposição de gordura subcutânea e intramuscular (ZHANG et al., 2011; SHENG et al., 2014).

As janelas localizadas nos cromossomos 3 (0,4-1,4 Mb), 5 (118,7-119,7), 16 (20,7-21,7) e 23 (24,1-25,1) estiveram associadas com pelo menos dois escores (Tabelas 3.3, 3.4 e 3.5). Isto era esperado uma vez que as correlações genéticas entre os escores foram altas 0,80 até 0,92 (dados não apresentados), indicando que os mesmos genes estão controlando os escores visuais. Os resultados deste estudo auxiliarão no processo de seleção dos escores visuais, uma vez que o aumento das frequências dos alelos favoráveis para os genes identificados, trará maiores ganhos genéticos.

### **3.4 Conclusões**

De acordo com os resultados pode-se dizer que os escores visuais são de natureza poligênica, uma vez que não foram encontradas regiões explicando grandes porcentagens de variância genética. No entanto, foram identificadas regiões associadas com os escores visuais nos cromossomos 1, 3, 5, 7, 14, 15, 16, 19, 20 e 23, contendo genes que fazem parte de processos biológicos e funções moleculares importantes no desenvolvimento corporal.



## REFERÊNCIAS

- ALAM, M.; LEE, Y.; PARK, B.; KIM, J.; LEE, S.; SHIN, H.; KIM, K.; KIM, N.; KIM, J. A whole genome association study to detect single nucleotide polymorphisms for body conformation traits in a Hanwoo population. **Asian-Australasian Journal of Animal Science**, v. 24, n. 3, p. 322–329, 2011.
- BECKER, D.; WIMMERS, K.; LUTHER, H.; HOFER, A.; LEEB, T. A genome-wide association study to detect QTL for commercially important traits in Swiss Large White boars. **PloS One**, v. 8, n. 2, p. e55951, 2013.
- BERG, I. V. D.; FRITZ, S.; BOICHARD, D. QTL fine mapping with BayesC ( $\pi$ ): a simulation study. **Genetics Selection Evolution**, v. 45, n. 19, p. 1–11, 2013.
- BODDICKER, N.; WAIDE, E.; ROWLAND, R.; LUNNEY, J.; GARRICK, D.; REECY, J.; DEKKERS, J. Evidence for a major QTL associated with host response to porcine reproductive and respiratory syndrome virus challenge. **Journal of Animal Science**, v. 90, n. 6, p. 1733–1746, 2012.
- BOITARD, S.; ROCHA, D. Detection of signatures of selective sweeps in the Blonde d'Aquitaine cattle breed. **Animal Genetics**, v. 44, n. 5, p. 579–583, 2013.
- BOLIGON, A. A.; ALBUQUERQUE, L. G. Correlações genéticas entre escores visuais e características reprodutivas em bovinos Nelore usando inferência bayesiana. **Pesquisa Agropecuária Brasileira**, v. 45, n. 12, p. 1412–1418, 2010.
- BOLORMAA, S.; HAYES, B.; SAVIN, K.; HAWKEN, R.; BARENDSE, W.; ARTHUR, P.; HERD, R.; GODDARD, M. Genome-wide association studies for feedlot and growth traits in cattle. **Journal of Animal Science**, v. 89, n. 6, p. 1684–1697, 2011.
- CHO, Y.; HAZEN, B. C.; RUSSELL, A. P.; KRALLI, A. Peroxisome proliferator-activated receptor  $\gamma$  coactivator 1 (*pgc-1*)-and estrogen-related receptor (*err*)-induced regulator in muscle 1 (*perm1*) is a tissue-specific regulator of oxidative capacity in skeletal muscle cells. **Journal of Biological Chemistry**, v. 288, n. 35, p. 25207–25218, 2013.
- CLAYTON, D. **snpStats: SnpMatrix and XSnpmatrix classes and methods**. [S.I.], 2013. R package version 1.14.0.
- COLE, J. B.; WIGGANS, G. R.; MA, L.; SONSTEGARD, T. S.; LAWLOR, T. J.; CROOKER, B. A.; TASSELL, C. P. V.; YANG, J.; WANG, S.; MATUKUMALLI, L. K. et al. Genome-wide association analysis of thirty one production, health, reproduction and body conformation traits in contemporary US Holstein cows. **BMC Genomics**, v. 12, n. 408, p. 1–17, 2011.
- de los CAMPOS, G.; PEREZ, R. P. **BGLR: Bayesian Generalized Linear Regression**. [S.I.], 2013. R package version 1.0.2. Disponível em: <<http://CRAN.R-project.org/package=BGLR>>.
- DEKKERS, J. C. Application of genomics tools to animal breeding. **Current Genomics**, v. 13, n. 3, p. 207–212, 2012.

DENNIS JR, G.; SHERMAN, B. T.; HOSACK, D. A.; YANG, J.; GAO, W.; LANE, H. C.; LEMPICKI, R. A. et al. David: database for annotation, visualization, and integrated discovery. **Genome Biology**, v. 4, n. 9, p. R60, 2003.

ELSIK, C. G.; TELLAM, R. L.; WORLEY, K. C. et al. The genome sequence of taurine cattle: a window to ruminant biology and evolution. **Science**, v. 324, n. 5926, p. 522–528, 2009.

FALCONER, D. S.; MACKAY, T. F. **Introduction to Quantitative Genetics**. 4th. ed. England: PERSON Prentice Hall, 1996. 464 p.

FARIA, C. U.; MAGNABOSCO, C. U.; ALBUQUERQUE, L. G.; De los Reyes, A.; BEZERRA, L. A. F.; LÔBO, R. B. Estimativas de correlações genéticas entre escores visuais e características de crescimento em bovinos da raça Nelore utilizando modelos bayesianos linear-limiar. **Ciência Animal Brasileira**, v. 9, n. 2, p. 327–340, 2008.

FARIA, C. U.; MAGNABOSCO, C. U.; ALBUQUERQUE, L. G.; BEZERRA, L. A. F. Estimativas de correlações genéticas entre escores visuais e características de carcaça medidas por ultrassonografia em bovinos Nelore utilizando modelos bayesianos linear-limiar. **Revista Brasileira de Zootecnia**, v. 38, n. 11, p. 2144–2151, 2009.

FEARNLEY, I. M.; CARROLL, J.; SHANNON, R. J.; RUNSWICK, M. J.; WALKER, J. E.; HIRST, J. GRIM-19, a cell death regulatory gene product, is a subunit of bovine mitochondrial NADH: ubiquinone oxidoreductase (complex i). **Journal of Biological Chemistry**, v. 276, n. 42, p. 38345–38348, 2001.

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

GARRICK, D. J.; TAYLOR, J. F.; FERNANDO, R. L. et al. Deregressing estimated breeding values and weighting information for genomic regression analyses. **Genetics Selection Evolution**, v. 41, n. 55, p. 1–8, 2009.

GIANOLA, D. Priors in whole-genome regression: the bayesian alphabet returns. **Genetics**, v. 194, n. 3, p. 573–596, 2013.

GIANOLA, D.; VERRIER, E.; VERRIER, E. Contribution of an additive locus to genetic variance when inheritance is multi-factorial with implications on interpretation of GWAS. **Theoretical and Applied Genetics**, v. 126, n. 6, p. 1457–1472, 2013.

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

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

GONZÁLEZ-RECIO, O.; CHANG, Y. M.; GIANOLA, D.; WEIGEL, K. A. Comparison of models using different censoring scenarios for days open in spanish holstein cows. **Animal Science**, Cambridge Univ Press, v. 82, n. 2, p. 233–239, 2006.

GONZÁLEZ-RECIO, O.; FORNI, S. Genome-wide prediction of discrete traits using bayesian regressions and machine learning. **Geneics Selection Evolution**, v. 43, n. 7, p. 1–12, 2011.

GORDO, D. G. **Estudo genético quantitativo da qualidade da carcaça e da carne e suas associações com as características de avaliação visual na raça Nelore**. Tese (Doutorado) — Universidade estadual paulista-UNESP, 2014.

GORDO, D. M.; BALDI, F.; LÔBO, R. B.; FILHO, W. K.; SAINZ, R. D.; ALBUQUERQUE, L. 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.

HABIER, D.; FERNANDO, R. L.; KIZILKAYA, K.; GARRICK, D. J. Extension of the bayesian alphabet for genomic selection. **BMC Bioinformatics**, v. 12, n. 186, p. 1–12, 2011.

HAWKEN, R.; ZHANG, Y.; FORTES, M.; COLLIS, E.; BARRIS, W.; CORBET, N.; WILLIAMS, P.; FORDYCE, G.; HOLROYD, R.; WALKLEY, J. et al. Genome-wide association studies of female reproduction in tropically adapted beef cattle. **Journal of Animal Science**, v. 90, n. 5, p. 1398–1410, 2012.

HAYES, B.; GODDARD, M. Genome-wide association and genomic selection in animal breeding. **Genome**, v. 53, n. 11, p. 876–883, 2010.

HAYES, B. J.; PRYCE, J.; CHAMBERLAIN, A. J.; BOWMAN, P. J.; GODDARD, M. E. Genetic architecture of complex traits and accuracy of genomic prediction: coat colour, milk-fat percentage, and type in Holstein cattle as contrasting model traits. **PLoS Genetics**, v. 6, n. 9, p. e1001139, 2010.

JÚNIOR, J. J.; PITA, F. V.; FRIES, L. A.; ALBUQUERQUE, L. G. Influência de alguns fatores de ambiente sobre os escores de conformação, precocidade e musculatura à desmama em um rebanho da raça Nelore. **Revista Brasileira de Zootecnia**, v. 30, n. 6, p. 1697–1703, 2001.

KOLBEHDARI, D.; WANG, Z.; GRANT, J.; MURDOCH, B.; PRASAD, A.; XIU, Z.; MARQUES, E.; STOTHARD, P.; MOORE, S. A whole-genome scan to map quantitative trait loci for conformation and functional traits in Canadian Holstein bulls. **Journal of Dairy Science**, v. 91, n. 7, p. 2844–2856, 2008.

KOURY FILHO, W.; ALBUQUERQUE, L. G.; FORNI, S.; SILVA, J.; YOKOO, M. J.; ALENCAR, M. M. de. Estimativas de parâmetros genéticos para os escores visuais e suas associações com peso corporal em bovinos de corte. **Revista Brasileira de Zootecnia**, v. 39, n. 5, p. 1015–1022, 2010.

LEE, S. H.; WERF, J. van der; KIM, N. K.; LEE, S. H.; GONDRO, C.; PARK, E. W.; OH, S. J.; GIBSON, J.; THOMPSON, J. QTL and gene expression analyses identify genes affecting carcass weight and marbling on BTA14 in Hanwoo (korean cattle). **Mammalian Genome**, v. 22, n. 9-10, p. 589–601, 2011.

MANNING, G.; WHYTE, D. B.; MARTINEZ, R.; HUNTER, T.; SUDARSANAM, S. The protein kinase complement of the human genome. **Science**, v. 298, n. 5600, p. 1912–1934, 2002.

MANOLIO, T. A.; COLLINS, F. S.; COX, N. J.; GOLDSTEIN, D. B.; HINDORFF, L. A.; HUNTER, D. J.; MCCARTHY, M. I.; RAMOS, E. M.; CARDON, L. R.; CHAKRAVARTI, A. et al. Finding the missing heritability of complex diseases. **Nature**, v. 461, n. 7265, p. 747–753, 2009.

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

MISZTAL, I.; TSURUTA, S.; STRABEL, T.; AUVRAY, B.; DRUET, T.; LEE, D. et al. BLUPF90 and related programs (BGF90). In: WORLD CONGRESS ON GENETICS APPLIED TO LIVESTOCK PRODUCTION, 7th., 2002, Montpellier, France. **Proceedings ...** [S.I.]: Institut National de la Recherche Agronomique (INRA), 2002. p. 1–2.

MONESTIER, O.; BRUN, C.; COCQUEMPOT, O.; PETIT, D.; BLANQUET, V. GASP/WFIKK proteins: evolutionary aspects of their functions. **PloS One**, v. 7, n. 8, p. e43710, 2012.

MORENO-SÁNCHEZ, N.; RUEDA, J.; CARABAÑO, M. J.; REVERTER, A.; MCWILLIAM, S.; GONZÁLEZ, C.; DÍAZ, C. Skeletal muscle specific genes networks in cattle. **Functional & Integrative Genomics**, v. 10, n. 4, p. 609–618, 2010.

MRODE, R. A. **Linear models for the prediction of animal breeding values**. [S.I.]: Cabi, 2005.

NEUSTAETER, A.; GROSSI, D.; JAFARIKIA, M.; SARGOLZAEI, M.; SCHENKEL, F. Genome-wide association study for loin marbling score in Canadian Duroc pigs. In: WORLD CONGRESS OF GENETICS APPLIED TO LIVESTOCK PRODUCTION, 10th., 2014, Vancouver, BC. **Proceedings...** Vancouver, BC: World Congress of Genetics Applied to Livestock Production, 2014. p. 1–3.

ONTERU, S.; FAN, B.; DU, Z.-Q.; GARRICK, D.; STALDER, K.; ROTHSCHILD, M. A whole-genome association study for pig reproductive traits. **Animal Genetics**, v. 43, n. 1, p. 18–26, 2012.

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

PÉREZ-MONTARELO, D.; MADSEN, O.; ALVES, E.; RODRÍGUEZ, M. C.; FOLCH, J. M.; NOGUERA, J. L.; GROENEN, M. A.; FERNÁNDEZ, A. I. Identification of genes regulating growth and fatness traits in pig through hypothalamic transcriptome analysis. **Physiological genomics**, American Physiological Society, v. 46, n. 6, p. 195–206, 2014.

RUBIN, C.-J.; ZODY, M. C.; ERIKSSON, J.; MEADOWS, J. R.; SHERWOOD, E.; WEBSTER, M. T.; JIANG, L.; INGMAN, M.; SHARPE, T.; KA, S. et al. Whole-genome resequencing reveals loci under selection during chicken domestication. **Nature**, v. 464, n. 7288, p. 587–591, 2010.

SADKOWSKI, T.; JANK, M.; OPRZADEK, J.; MOTYL, T. Age-dependent changes in bovine skeletal muscle transcriptomic profile. **Journal of Physiology and Pharmacology**, v. 57, n. 7, p. 95–110, 2006.

SEGER, R.; KREBS, E. G. The mapk signaling cascade. **The FASEB Journal**, v. 9, n. 9, p. 726–735, 1995.

SHENG, X.; NI, H.; LIU, Y.; LI, J.; ZHANG, L.; GUO, Y. Rna-seq analysis of bovine intramuscular, subcutaneous and perirenal adipose tissues. **Molecular Biology Reports**, v. 41, n. 3, p. 1631–1637, 2014.

UTSUNOMIYA, Y. T.; CARMO, A. S.; NEVES, H. H.; CARVALHEIRO, R.; MATOS, M. C.; ZAVAREZ, L. B.; ITO, P. K.; O'BRIEN, A. M. P.; SÖLKNER, J.; PORTO-NETO, L. R. et al. Genome-wide mapping of loci explaining variance in scrotal circumference in Nelore cattle. **PLoS One**, v. 9, n. 2, p. e88561, 2014.

WANG, H.; ZHENG, Y.; WANG, G.; LI, H. Identification of microRNA and bioinformatics target gene analysis in beef cattle intramuscular fat and subcutaneous fat. **Molecular BioSystems**, v. 9, n. 8, p. 2154–2162, 2013.

WANG, J.; ZHOU, H.; FANG, Q.; LUO, Y.; HICKFORD, J. G. Variation in the ovine *wfikkn2* gene. **Gene**, v. 543, n. 1, p. 53–57, 2014.

WOOD, A. R.; ESKO, T.; YANG, J.; VEDANTAM, S.; PERS, T. H.; GUSTAFSSON, S.; CHU, A. Y.; ESTRADA, K.; LUAN, J.; KUTALIK, Z. et al. Defining the role of common variation in the genomic and biological architecture of adult human height. **Nature Genetics**, v. 46, n. 11, p. 1173–1186, 2014.

WU, X.; FANG, M.; LIU, L.; WANG, S.; LIU, J.; DING, X.; ZHANG, S.; ZHANG, Q.; ZHANG, Y.; QIAO, L. et al. Genome wide association studies for body conformation traits in the chinese Holstein cattle population. **BMC Genomics**, v. 14, n. 897, p. 1–10, 2013.

YOKOO, M. J.; WERNECK, J. N.; PEREIRA, M. C.; ALBUQUERQUE, L. G.; FILHO, W. K. Correlações genéticas entre escores visuais e características de carcaça medidas por ultrassom em bovinos de corte. **Pesquisa Agropecuária Brasileira**, v. 44, n. 2, p. 197–202, 2009.

ZHANG, Y.-Y.; ZAN, L.-S.; WANG, H.-B.; QING, L.; WU, K.-X.; QUAN, S.-A.; LI, C.-Q.; ZHONG, X.; WANG, C.-J. Differentially expressed genes in skeletal muscle tissues from castrated qinchuan cattle males compared with those from intact males. **Livestock Science**, v. 135, n. 1, p. 76–83, 2011.