

**Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”**

**Seleção e análise de associação genômica em dados simulados e
da qualidade da carne de ovinos da raça Santa Inês**

Simone Fernanda Nedel Pértile

Tese apresentada para obtenção do título de
Doutora em Ciências. Área de concentração:
Ciência Animal e Pastagens

**Piracicaba
2015**

Simone Fernanda Nedel Pértile
Zootecnista

**Seleção e análise de associação genômica em dados simulados e da qualidade
da carne de ovinos da raça Santa Inês**

versão revisada de acordo com a resolução CoPGr 6018 de 2011

Orientador:
Prof. Dr. **GERSON BARRETO MOURÃO**

Tese apresentada para obtenção do título de
Doutora em Ciências. Área de concentração:
Ciência Animal e Pastagens

Piracicaba
2015

**Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA - DIBD/ESALQ/USP**

Pértile, Simone Fernanda Nedel

Seleção e análise de associação genômica em dados simulados e da qualidade da carne de ovinos da raça Santa Inês / Simone Fernanda Nedel Pértile. - - versão revisada de acordo com a resolução CoPGr 6018 de 2011. - - Piracicaba, 2015.

91 p. : il.

Tese (Doutorado) - - Escola Superior de Agricultura "Luiz de Queiroz".

1. Coeficiente de herdabilidade 2. Habilidade preditiva 3. ssGBLUP 4. Validação cruzada I. Título

CDD 636.3082
P469s

“Permitida a cópia total ou parcial deste documento, desde que citada a fonte – O autor”

DEDICATÓRIA

Aos meus pais Gentil e Lícia, por todo amor e incentivo, por sempre me apoiarem em minhas decisões.

Às minhas irmãs Renata e Raquel, pelo apoio, amizade e carinho.

Ao meu noivo Gilson, por todo amor, cuidado, dedicação e apoio.

Dedico.

AGRADECIMENTOS

À Deus, por iluminar diariamente o meu caminho.

À Escola Superior “Luiz de Queiroz”, Universidade de São Paulo, em especial ao Departamento de Zootecnia, pela oportunidade concedida para a realização deste estudo.

À FAPESP, pela concessão da bolsa de estudos e de outros auxílios financeiros na realização deste estudo (Processo FAPESP nº2012/21006-4).

Ao Professor Gerson Barreto Mourão, pela orientação, dedicação, ensinamentos, compreensão e confiança, indispensáveis na realização deste estudo e para o meu crescimento profissional.

Aos professores do Departamento de Zootecnia, ESALQ/USP, em especial aos Professores Luiz Lehmann Coutinho e Carla Bittar, pelos ensinamentos, pelas contribuições para meu aperfeiçoamento profissional e pelo apoio.

Aos alunos do Grupo de Estudos em Estatística e Melhoramento Animal (GEMA) da ESALQ/USP: Aline, Ana, Caroline, Fabiane, Gilson, Grégori, João, Johanna, Juliana, Laiza, Mary Anna, Mayara e Tiago, pela companhia, apoio, auxílios na realização deste trabalho, momentos de descontração e experiências durante o período do doutorado.

Ao Professor Fabyano Fonseca e Silva, pelo o apoio na realização deste trabalho, pelos ensinamentos e pelo auxílio no meu crescimento profissional.

Ao Professor Eduardo Delgado e aos alunos de pós-graduação do laboratório de anatomia e fisiologia animal (Daiane, Giuliana, Patrícia, Marcelo e Giancarlo), pelos ensinamentos, apoio nas coletas de dados e amizade.

Aos Professores da Universidade Federal da Bahia (UFBA), em especial ao Professor Luís Fernando e ao Professor Gleidson, e aos alunos de pós-graduação e graduação desta Universidade, pelos ensinamentos, convívio e auxílios nas coletas de dados.

Aos colegas de Pós-Graduação em Ciência Animal e Pastagens.

À Sheila Tavares Nascimento, pela amizade, por se fazer sempre presente mesmo à distância, pelas palavras de incentivo e apoio nas horas mais difíceis.

Às companheiras de república Glaucia Moreira, Isadora Mendes, Isabela Nascimento Kathelyn Guimarães e Rafaela Pereira, pelo apoio, amizade, companhia e momentos de descontração.

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

Muito Obrigada!

EPIGRAFE

“A satisfação reside no esforço, não no resultado obtido. O esforço total é a plena vitória.”

Mahatma Gandhi

“Não importa onde você parou, em que momento da vida você cansou, o que importa é que sempre é possível e necessário recomeçar.”

Carlos Drummond de Andrade

“Suba o primeiro degrau com fé. Você não tem que ver toda a escada. Você só precisa dar o primeiro passo.”

Martin Luther King Jr.

SUMÁRIO

RESUMO.....	11
ABSTRACT	13
1 INTRODUÇÃO	15
1.1 Hipóteses	16
1.2 Objetivos	16
Referências	17
2 REVISÃO BIBLIOGRÁFICA	19
2.1 Seleção e associação genômica ampla	19
2.2 Controle de qualidade dos dados.....	20
2.2.1 Controle de qualidade por amostra	21
2.2.2 Controle de qualidade dos marcadores genéticos	23
2.3 Desequilíbrio de ligação	24
2.4 Métodos de predição dos efeitos dos marcadores.....	26
2.4.1 GBLUP e ssGBLUP	27
2.4.2 Métodos bayesianos.....	29
2.5 Análises dos efeitos dos marcadores.....	32
Referências	34
3 SELEÇÃO E ASSOCIAÇÃO GENÔMICA AMPLA CONSIDERANDO SITUAÇÕES REAIS DE PROGRAMAS DE MELHORAMENTO GENÉTICO ANIMAL	41
Resumo.....	41
Abstract.....	41
3.1 Introdução	42
3.2 Material e métodos.....	43
3.2.1 Avaliação da importância das informações de pedigree, coeficiente de herdabilidade e efeito poligênico	43
3.2.1.1 Dados simulados.....	44
3.2.1.2 Predição dos valores genéticos genômicos e dos efeitos de marcadores genéticos por meio dos métodos GBLUP e ssGBLUP.....	46
3.2.1.3 Predição dos valores genéticos genômicos via métodos bayesianos	48
3.2.2 Avaliação da importância de diferentes números de QTL e pressões de seleção para seleção genômica	49

3.3 Resultados e discussão	50
3.3.1 Avaliação da importância das informações de pedigree, coeficiente de herdabilidade e efeito poligênico na seleção genômica via métodos ssGBLUP e GBLUP.....	50
3.3.2 Avaliação da importância das informações de pedigree, coeficiente de herdabilidade e efeito poligênico nas análises de associação genômica ampla via métodos ssGBLUP e GBLUP.....	55
3.3.3 Seleção genômica ampla via método GBLUP e abordagem bayesiana para fenótipos com diferentes coeficiente de herdabilidade e efeito poligênico.....	61
3.3.4 Avaliação da importância de diferentes números de QTL e pressões de seleção.....	64
3.4 Conclusões	65
Referências.....	66
4 ASSOCIAÇÃO GENÔMICA AMPLA PARA COMPOSIÇÃO DE GORDURA INTRAMUSCULAR E CARACTERÍSTICAS DE QUALIDADE DA CARNE E CARÇAÇA EM OVINOS DA RAÇA SANTA INÊS.....	69
Resumo	69
Abstract.....	69
4.1 Introdução.....	70
4.2 Material e métodos	71
4.2.1 Conjunto de dados.....	71
4.2.2 Extração de DNA e genotipagem.....	73
4.2.3 Controle de qualidade e análise de associação genômica ampla	73
4.3 Resultados e discussão	75
4.4 Conclusões	83
Referências.....	84
ANEXOS.....	87

RESUMO

Seleção e análise de associação genômica em dados simulados e da qualidade da carne de ovinos da raça Santa Inês

Informações de milhares de marcadores genéticos têm sido incluídas nos programas de melhoramento genético, permitindo a seleção dos animais considerando estas informações e a identificações de regiões genômicas associadas às características de interesse econômico. Devido ao alto custo associado a esta tecnologia e às coletas de dados, os dados simulados apresentam grande importância para que novas metodologias sejam estudadas. O objetivo deste trabalho foi avaliar a eficiência do método ssGBLUP utilizando pesos para os marcadores genéticos, informações de genótipo e fenótipos, com ou sem as informações de pedigree, para seleção e associação genômica ampla, considerando diferentes coeficientes de herdabilidade, presença de efeito poligênico, diferentes números de QTL (*quantitative trait loci*) e pressões de seleção. Adicionalmente, dados de qualidade da carne de ovinos da raça Santa Inês foram comparados com a os padrões descritos para esta raça. A população estudada foi obtida por simulação de dados, e foi composta por 8.150 animais, sendo 5.850 animais genotipados. Os dados simulados foram analisados utilizando o método ssGBLUP com matrizes de relacionamento com ou sem informações de pedigree, utilizando pesos para os marcadores genéticos obtidos em cada iteração. As características de qualidade da carne estudadas foram: área de olho de lombo, espessura de gordura subcutânea, cor, pH ao abate e após 24 horas de resfriamento das carcaças, perdas por cocção e força de cisalhamento. Quanto maior o coeficiente de herdabilidade, melhores foram os resultados de seleção e associação genômica. Para a identificação de regiões associadas a características de interesse, não houve influência do tipo de matriz de relacionamento utilizada. Para as características com e sem efeito poligênico, quando considerado o mesmo coeficiente de herdabilidade, não houve diferenças para seleção genômica, mas a identificação de QTL foi melhor nas características sem efeito poligênico. Quanto maior a pressão de seleção, mais acuradas foram as predições dos valores genéticos genômicos. Os dados de qualidade da carne obtidos de ovinos da raça Santa Inês estão dentro dos padrões descritos para esta raça e foram identificadas diversas regiões genômicas associadas às características estudadas.

Palavras-chaves: Bayesiana; Coeficiente de herdabilidade; ssGBLUP; Validação cruzada

ABSTRACT

Genomic selection and association analysis in simulated data and meat quality of Santa Inês sheep breed

Thousands of genetic markers data have been included in animal breeding programs to allow the selection of animals considering this information and to identify genomic regions associated to traits of economic interest. Simulated data have great importance to the study of new methodologies due to the high cost associated with this technology and data collection. The objectives of this study were to evaluate the efficiency of the ssGBLUP method using genotype and phenotype information, with or without pedigree information, and attributing weights for genetic markers, for selection and genome-wide association considering different coefficients of heritability, the presence of polygenic effect, different numbers of quantitative trait loci and selection pressures. Additionally, meat quality data of Santa Ines sheep breed were compared with the standards for the breed. The population of simulated data was composed by 8.150 individuals and 5.850 genotyped animals. The simulated data was analysed by the ssGBLUP method and by two relationship matrix, with or without pedigree information, and weights for genetic markers were obtained in every iteration. The traits of meat quality evaluated were: rib eye area, fat thickness, color, pH at slaughter and 24 hours after the carcass cooling, cooking losses and shear force. The results of selection and genomic association were better for the traits with the highest heritability coefficients. For traits with the greater selection pressure, more accurate predictions of the genomic breeding values were obtained. There was no difference between the relationship matrix studied to identify the regions associated with traits of interest. For the traits with and without polygenic effect, considering the same heritability coefficient, they did not show differences in genomic selection, but the identification of the QTL was better for traits without polygenic effect. The meat quality data obtained from Santa Ines sheep breed are in accordance with the standards for this breed and different genomic regions associated to the studied characteristics were identified.

Keywords: Bayesian; Heritability coefficient; ssGBLUP; Cross-validation

1 INTRODUÇÃO

A ovinocultura é uma atividade agropecuária em expansão no Brasil, pelo aumento no número de animais produzidos, melhorias na produtividade e aumento no consumo de carne ovina no país. Esta expansão tem ocorrido na maioria dos estados brasileiros, dentre os quais o estado da Bahia apresenta destaque, o qual possuía um rebanho em 2013 de 2.926.601 cabeças de ovinos, tendo um rebanho inferior apenas ao do estado do Rio Grande do Sul, o qual possuía 4.250.932 neste ano, enquanto que o Brasil possuía 17.290.519 cabeças, segundo o senso pecuário do IBGE (2013).

Paralelo a este cenário, ainda há demanda do mercado consumidor para que haja crescimento na produção de carne ovina no país. Além disso, os consumidores buscam produtos com melhor qualidade, especialmente nutricional, devido a recente procura dos consumidores por alimentos que contribuam positivamente na qualidade da alimentação e, conseqüentemente, à saúde dos consumidores.

Para atingir tais objetivos, no melhoramento genético clássico, a identificação dos genótipos superiores é realizada a partir de informações de pedigree e genótipo. Recentemente, Meuwissen et al. (2001) propuseram um método denominado seleção genômica, o qual trata-se da seleção assistida por marcadores genéticos do tipo SNP (*single nucleotide polymorphism*), considerando uma escala genômica ampla (MEUWISSEN, 2007); neste caso a seleção é simultânea para dezenas ou centenas de milhares marcadores SNPs, os quais podem cobrir de modo denso e amplo todo o genoma de uma referida espécie. Assim, a partir da estimação dos efeitos dos SNPs, pode-se prever os valores genéticos dos indivíduos para características complexas.

Além da seleção genômica, as informações genotipagem obtidas a partir de chips de marcadores genéticos de alta densidade podem ser utilizadas para estudos de associação destes SNPs com características de interesse (GWAS), tornando possível desvendar parte da arquitetura genética que atua no controle destas características de herança complexa (DAETWYLER et al., 2010), pela identificação de regiões associadas com estas características.

Visando melhorias na aplicação dos métodos estatísticos de seleção e associação genômica ampla, a simulação de dados pode ser uma excelente alternativa, principalmente pelo alto custo relacionado às coletas de dados e à

tecnologia de genotipagem com chips de alta densidade de marcadores, e por permitir a obtenção dos valores genéticos reais dos indivíduos da população estudada (HICKEY; GORJANC, 2012).

1.1 Hipóteses

- Nas análises de seleção e associação genômica ampla, o uso das informações de pedigree, por meio da matriz de relacionamento H e em substituição da matriz G, ocasionam em maiores acurácias na predição dos valores genéticos genômicos e melhorias na identificação de QTL (*quantitative trait loci*), quando utilizado o método ssGBLUP;
- Quanto maior o coeficiente de herdabilidade da característica, maior será a acurácia da predição dos valores genéticos genômicos, melhorando também a identificação de regiões associadas ao fenótipo;
- Os valores genéticos genômicos preditos utilizando abordagem bayesiana possuem maiores acurácias do que os preditos pelo método ssGBLUP;
- A presença de efeito poligênico, como componente da variância genética aditiva de um fenótipo, quando utilizado o método ssGBLUP, não influencia na predição dos valores genéticos genômicos e nas análises de GWAS;
- O número de regiões do genoma associadas a um fenótipo influencia na predição dos valores genéticos genômicos utilizado no método ssGBLUP, assim como a pressão de seleção na população estudada;
- Os dados de qualidade da carne obtidos para ovinos da raça Santa Inês estão dentro dos padrões descritos na literatura para machos inteiros desta raça;
- Há múltiplas regiões do genoma associadas com as características de carcaça, qualidade da carne e composição de gordura intramuscular em ovinos.

1.2 Objetivos

O objetivo deste trabalho foi avaliar a eficiência do método GBLUP utilizando informação de genótipo e fenótipos, com ou sem as informações de pedigree, para seleção e associação genômica ampla, considerando diferentes coeficientes de

herdabilidade, presença de efeito poligênico, diferentes números de marcadores SNPs e pressões de seleção. Adicionalmente, foram estudados os benefícios do método ssGBLUP com processo iterativo, utilizando pesos para dos marcadores genéticos, e foram comparadas as acurácias obtidas pelo método GBLUP com métodos bayesianos. Em relação aos dados de ovinos da raça Santa Inês, os objetivos deste estudo foram comparar os dados de qualidade obtidos com os padrões descritos para esta raça e identificar regiões genômicas associadas a estas características, via método GBLUP.

Referências

DAETWYLER, H.D.; PONG-WONG, R.; VILLANUEVA, B.; WOOLLIAMS, J.A. The Impact of Genetic Architecture on Genome-Wide Evaluation Methods. **Genetics**, Austin, v. 185, p. 1021–1031, 2010.

GODDARD, M. Genomic selection: prediction of accuracy and maximization of long term response. **Genetics**, Austin, v. 136, p. 245–257, 2009.

HICKEY, J.M.; GORJANC, G. Simulated data for genomic selection and genome-wide association studies using a combination of coalescent and gene drop methods. **Genes, Genomes and Genetics**, Toronto, v. 2, p. 425-427, 2012.

IBGE. Instituto brasileiro de geografia e estatística: Produção da pecuária municipal em 2013. Disponível em: ftp://ftp.ibge.gov.br/Producao_Pecuaria/Producao_da_Pecuaria_Municipal/2013/tabelas_pdf/tab19.pdf . Acesso em: 11 maio 2015.

MEUWISSEN, T.H.E. Genomic selection: marker assisted selection on genome-wide scale. **Journal of Animal Breeding and Genetics**, Berlim, v. 124, p. 321-322, 2007.

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

2 REVISÃO BIBLIOGRÁFICA

2.1 Seleção e associação genômica ampla

A eficiência de um programa de melhoramento genético depende da identificação precisa dos genótipos superiores, visando os objetivos do programa de seleção. Para isto, no melhoramento genético clássico, são preditos os valores genéticos dos indivíduos, utilizando registros de fenótipo e de pedigree, ou seja, pelos dados fenotípicos e de pedigree pretende-se fazer inferências nos genótipos dos animais.

As estimativas de valores genéticos dos indivíduos são obtidas frequentemente a partir dos fenótipos analisados sob modelos estatísticos e genéticos, sendo menos eficiente em características de difícil mensuração, como de qualidade da carne, que exigem o abate do animal. Portanto, para melhorar essas características seria vantajoso e útil identificar os genes para realizar a seleção de animais portadores de alelos desejáveis (GODDARD; HAYES, 2007).

Neste contexto, o uso de informações obtidas a partir de técnicas de genética molecular no melhoramento genético animal tem como principal vantagem a utilização direta da informação do material genético na seleção, permitindo melhorias na eficiência de seleção e diminuição no tempo de avaliação, em comparação com a tradicional seleção baseada essencialmente em dados fenotípicos. O uso de informações genéticas moleculares e fenotípicas para as avaliações caracteriza a seleção assistida por marcadores genéticos. As informações genéticas utilizadas na seleção assistida por marcadores eram oriundas de estudos de alguns genes individuais ou de poucos *loci* associados às características quantitativas (DEKKERS; HOSPITAL, 2002; DEKKERS, 2004).

Nos últimos anos, informações de milhares de marcadores genéticos, os quais cobrem de forma densa o genoma, têm sido utilizadas a partir de um método proposto por Meuwissen et al. (2001), pelo qual pressupõe-se que os genes de interesse provavelmente apresentarão desequilíbrio de ligação com parte importante dos marcadores genéticos. A partir deste método, espera-se obter melhorias na seleção dos animais (GWS) e na identificação de genes de interesse (GWAS).

A seleção genômica (MEUWISSEN et al., 2001) tem como objetivo aumentar a eficiência da seleção genética que tem sido praticada com base apenas em dados

fenotípicos avaliados a campo. Tal método surge como um complemento ao melhoramento genético clássico, contribuindo para que os produtores selecionem do seu rebanho, o mais breve possível, animais com potenciais produtivos que são desejáveis, podendo reduzir gastos com a manutenção de animais improdutivos, proporcionando maior progresso genético, a partir de maior velocidade e precisão dos processos seletivos.

No mesmo contexto, os estudos de associação genômica ampla (GWAS) possuem vantagens quando comparados com a seleção assistida por marcadores, devido ao fato de que a maioria dos genes candidatos ou QTL (*Quantitative trait loci*) capturavam apenas uma proporção pequena da variação genética total.

Em relação ao tipo de marcadores genéticos utilizados nos estudos de GWS e GWAS, a variância genética pode ser mais bem explicada pelos marcadores genéticos SNPs (*single nucleotide polymorphism*) em especial por sua abundância e abrangência genômica. Assumindo que esses marcadores são densos o suficiente para estarem em desequilíbrio de ligação com os QTL, os marcadores podem ser utilizados para selecionar o alelo favorável em cada QTL, sem a necessidade de realmente identificar o QTL ou o polimorfismo funcional. Além disso, a partir dos dados desses marcadores, o valor genético de um indivíduo pode ser obtido com alta acurácia, utilizando registros de marcadores genéticos e fenótipos em uma população (GODDARD, 2009).

Os marcadores genéticos do tipo SNPs são a forma mais abundante de variação do DNA genômico e são preferidos em relação a outros marcadores genéticos devido à sua baixa taxa de mutação e facilidade de genotipagem (RESENDE et al., 2008).

2.2 Controle de qualidade dos dados

O sucesso dos estudos de GWS e GWAS está relacionado com o cuidado adequado dos dados após a genotipagem, ou seja, controle de qualidade dos dados, predição dos efeitos dos SNPs e estabelecimentos de níveis adequados de seleção dos marcadores com efeitos significativos ou não. Após a obtenção dos dados de genotipagem, o primeiro passo é a realização do controle de qualidade destes,

utilizando critérios como call rate, MAF (*minor allele frequency*), aderência ao Equilíbrio de Hardy-Weinberg, ocorrência de indivíduos repetidos, correção de parentesco, entre outros. A importância do controle de qualidade dos dados de genotipagem está no aumento na confiabilidade dos efeitos dos marcadores SNPs e valores genéticos genômicos preditos, pois a acurácia é reduzida quando este controle não é realizado ou é ineficiente (WIGGANS et al., 2009). Adicionalmente, segundo os mesmos autores, marcadores SNPs que não contribuem para a acurácia das avaliações podem ser eliminados, reduzindo assim esforços computacionais. Mesmo quando a taxa de erro ou viés é baixa, manter marcadores que não atendam aos critérios de qualidade pode ser prejudicial, pelo número de marcadores estudados (ANDERSON et al., 2010).

2.2.1 Controle de qualidade por amostra

O controle de qualidade das amostras deve ser realizado pela: (i) identificação de indivíduos com informação sexual discordante, o qual pode ser verificado pela homozigose do cromossomo X; (ii) identificação de indivíduos com genótipos faltando (call rate) ou desvios de heterozigosidade; (iii) identificação de indivíduos duplicados; (iv) a identificação de indivíduos de ascendência divergentes, (v) estratificação da população (ANDERSON et al., 2010; ZANELLA, 2011).

A eficiência da genotipagem ou *call rate* é uma medida de qualidade da amostra de DNA, concentração de DNA, falhas no processo de genotipagem, contaminação por altos níveis de proteína e RNA. Neste critério de controle de qualidade, amostras que apresentarem *call rate* abaixo de um limiar, estabelecido de acordo com o tamanho da população, devem ser retiradas das análises (SETTLES et al., 2009).

A heterozigosidade é definida como uma a relação entre o número de heterozigotos esperados e observados, sendo considerada uma média dos *loci* estudados de um indivíduo e a média da população. Assim, uma amostra que apresenta baixo número de marcadores SNPs em heterozigose em relação à média da população pode ser indicativo de endogamia, e caso contrário, um alto número de heterozigotos apresentado por uma amostra pode indicar contaminação. Dessa forma, a heterozigosidade média pode ser calcula pela fórmula $(N-O) / N$, em que N é

o número total de genótipos observados; e O é o número de genótipos homozigotos observados para dado indivíduo, sendo que os cromossomos sexuais não são considerados neste cálculo. Como os desvios de heterozigosidade são analisados em relação à média da população, amostras com heterozigosidade da ordem de 3 desvios-padrão acima ou abaixo da média devem ser excluídos da análise (ANDERSON et al., 2010).

A correção de parentesco é realizada após a comparação entre os genótipos do pai e da progênie. Esta correção é feita a partir dos coeficientes de parentesco calculados por métodos tradicionais, propostos por Malécot (1969) e Wright (1922). O primeiro a propor um coeficiente de parentesco eficiente a partir da genealogia foi Wright, o qual foi generalizado por outros autores, tais como Malécot (1969), e definido como a probabilidade de que dois alelos de um mesmo gene presentes em um mesmo indivíduo sejam idênticos por descendência, ou seja, quando um alelo de um indivíduo é tomado ao acaso, qual a probabilidade deste alelo ser idêntico por descendência ao alelo de outro indivíduo (FALCONER; MACKAY, 1996). Os valores de coeficiente de parentesco obtidos por estes métodos poderão ser comparados com os obtidos a partir das semelhanças genômicas do material genético dos indivíduos genotipados (WIGGANS et al., 2009), permitindo a correção de incoerências.

Um dos grandes problemas nos estudos de associação genética é a estratificação da população, em que as diferenças genótípicas entre animais são geradas devido a diferentes origens de população e não devido a qualquer efeito sobre o fenótipo (CARDON; PALMER, 2003). Assim, a estratificação da população ocorre quando indivíduos pertencentes à população em estudo diferem sistematicamente em relação à ascendência genética e ao fenótipo. Assim, as informações de ascendência são indispensáveis para estudos de associação genômica ampla, para o controle da estratificação da população e assim diminuir a ocorrência de falsos – positivos (TURNER et al., 2011). Uma alternativa é a adoção de modelo com inclusão do vetor de efeitos poligênicos, o qual contempla a matriz de parentesco e permite correção para estrutura de população (MACLEOD et al., 2010).

Em estudos utilizando milhares de marcadores, a estratificação da população pode ser avaliada pela proporção média de alelos compartilhados entre dois indivíduos, em relação ao total de marcadores genéticos avaliados, desconsiderando

marcadores genéticos presentes nos cromossomos sexuais (ANDERSON et al., 2010; PURCELL et al., 2007). Além disso, o gráfico de escalonamento multidimensional (*multidimensional scaling*, MDS) pode ser utilizado como teste para estudar a estratificação da população, considerando marcadores que não estejam em desequilíbrio de ligação ($r^2 < 0,02$), no qual a similaridade dos alelos entre os *loci* independentes é mensurada dentre todos os animais (ZANELLA, 2011). Outro método que pode ser utilizado é gráfico quantil-quantil (Q-Q plot) (SETTLES et al. 2009; ZANELLA et al. 2011), e por este gráfico é possível verificar a existência de subestrutura na população, pela relação entre os valores qui-quadrado esperado contra os valores qui-quadrado observado (ZANELLA, 2011).

2.2.2 Controle de qualidade dos marcadores genéticos

O controle de qualidade dos SNPs é realizado utilizando as frequências alélicas e genotípicas, para os critérios: proporção de genótipos perdidos (*call rate*), desvios do equilíbrio de Hardy-Weinberg e frequência do menor alelo (ANDERSON et al., 2010).

A eficiência de genotipagem ou *call rate* é um critério de qualidade indicativo da quantidade de genótipos perdidos por SNP. A remoção de marcadores de qualidade inferior é muito importante para os estudos GWAS, porque eles podem apresentar-se como os falsos positivos e reduzir a capacidade de identificar associações verdadeiras (ANDERSON et al., 2010). Apesar do valor indicativo da eficiência de genotipagem variar de acordo estudo, especialmente em relação ao número de indivíduos avaliados, Tunner et al. (2011) sugeriram que seja considerado um *call rate* de 98-99%.

Em relação ao critério de Equilíbrio de Hardy Weinberg, as frequências alélicas e genotípicas se mantêm constantes de geração para geração, em uma população onde não há seleção, mutação, migração e quando o tamanho populacional é grande (FALCONER; MACKAY, 1996). Ainda segundo estes autores, a relação matemática entre as frequências alélicas e genotípicas, quando considerados marcadores genéticos ou indivíduos diplóides, é dada por $p^2 + 2pq + q^2 = 1$, em que p é a frequência de um dos alelos e q é a frequência do outro alelo. Assim, a partir das frequências alélicas são calculadas as frequências genotípicas

esperadas, as quais são comparadas com as frequências genotípicas observadas na população. Quando verificados desvios do Equilíbrio de Hardy-Weinberg, alguma pressuposição (seleção, por exemplo) pode ter sido violada, ou seja, os alelos não estão segregando na população como esperado, pela ocorrência de um maior número de indivíduos homocigotos ou heterocigotos do que pela lei do Equilíbrio de Hardy-Weinberg. Desvios do Equilíbrio de Hardy-Weinberg também podem estar relacionados a erros de genotipagem. Além disso, segundo Wiggans et al. (2009), desvios do Equilíbrio de Hardy-Weinberg podem indicar problemas em determinar acuradamente o genótipo do SNP.

A avaliação do critério Equilíbrio de Hardy-Weinberg é geralmente feita pela escolha de um limiar, do qual o valor varia de estudo pra estudo (ANDERSON et al., 2010; TUNNER et al., 2011). Wiggans et al. (2009) consideraram que se a diferença entre as frequências genotípicas observadas e esperadas fosse maior 0,15, o SNP deveria ser excluído.

O critério de seleção de SNPs conhecido como MAF (*minor allele frequency*) ou frequência do menor alelo, utiliza as frequências alélicas para verificar a ocorrência de alelos com baixa frequência ou monomórficos, os quais não são informativos, e assim haveria dificuldade em selecionar na população alelos que são encontrados em poucos indivíduos. Considerando um alelo de um marcador SNP com MAF menor que 2%, em uma população de 100 indivíduos, apenas dois indivíduos terão o alelo. Assim, a escolha do limiar para o critério MAF pode variar de estudo pra estudo.

2.3 Desequilíbrio de ligação

O desequilíbrio de ligação ocorre quando os alelos de dois genes não segregam independentemente, ou seja, genes em que os alelos são associados aleatoriamente estão em equilíbrio de ligação, caso contrário, quando os alelos não segregam independentemente estes genes estão em desequilíbrio de ligação (HARTL; CLARK, 2007). A partir deste conceito, marcadores genéticos podem ser utilizados para explicar a parte da variação genética de um fenótipo, quando estes estão em desequilíbrio de ligação com o QTL responsável pela variação.

A variação genética de características poligênicas pode ser explicada por muitos *loci*, e assim, para explicar esta variação podemos utilizar milhares de marcadores que cobrem de forma densa todo o genoma. Assumindo que esses marcadores são densos o suficiente para estarem em desequilíbrio de ligação com os QTL, os marcadores podem ser utilizados para selecionar o alelo favorável em cada QTL, sem a necessidade de realmente identificar o QTL ou o polimorfismo funcional (GODDARD, 2009). Considerando os marcadores que são utilizados em estudos de GWAS, os SNPs apresentam destaque por sua abundância e abrangência genômica, sendo úteis na explicação da variância genética das características de herança complexa.

O conceito de desequilíbrio de ligação está associado à fração de recombinação, em que esta é utilizada para determinar se dois genes estão ligados. A recombinação pode ocorrer durante a meiose, processo a partir do qual células diplóides dividem-se em quatro células haplóides, as quais são os gametas. A meiose é composta por dois ciclos de divisão celular, sendo que no primeiro ciclo, mais precisamente na primeira fase deste ciclo (Prófase I) pode ocorrer a troca de fragmentos entre os cromossomos homólogos. A fração de recombinação é maior quanto maior a distância entre os genes, entre marcadores genéticos, e entre estes e o QTL (GRIFFITHS et al., 2000). Assim, quanto mais próximos os *loci*, maior é a probabilidade de estes genes segregarem juntos.

A estatística de desequilíbrio de ligação proposta por Hill (1981) considera que o desequilíbrio de ligação ocorre quando a frequências dos gametas em associação e repulsão diferem. Assim, considerando dois marcadores com alelos A e a, e B e b, tem-se:

$$D = Prob(AB).Prob(ab) - Prob(Ab).Prob(aB)$$

Este método considera dois *loci*, sendo muito dependente das frequências individuais dos alelos, não sendo adequada para a comparação de múltiplos *loci*. Para este segundo caso, uma estatística mais adequada é o r^2 proposto por Hill e Robertson (1968), na qual os valores variam de zero, quando os pares de *loci* não tem desequilíbrio de ligação entre eles, a 1, que representa pares de loci com desequilíbrio de ligação completo. Assim, o desequilíbrio de ligação é dado por:

$$r^2 = \frac{D^2}{f(A)f(a)f(B)f(b)},$$

em que $D=f(AB)-f(A)f(B)$, e $f(AB)$, $f(A)$, $f(a)$, $f(B)$, $f(b)$, são as frequências observadas do haplótipo AB e dos alelos A, a, B, b, respectivamente (SARGOLZAEI et al., 2008).

Quando considerada a relação entre um marcador genético e um QTL não observado, a estatística r^2 mede a proporção da variação causada por alelos do QTL que é explicada pelos marcadores (PRITCHARD; PRZEWORSKI, 2001). Assim, esta é uma medida de correlação entre os alelos, a qual nos indica haplótipos herdados de um único ancestral ou haplótipos idênticos por descendência (REICH et al. 2001).

Dentre os fatores que podem afetar o desequilíbrio de ligação estão a seleção, a taxa de recombinação e mutação (MATUKUMALLI et al., 2009). Assim, estes fatores, especialmente seleção e taxa de recombinação, tornam necessárias que os estudos dos efeitos dos marcadores sejam refeitos após gerações de seleção, pois se espera que os marcadores que estejam em desequilíbrio de ligação com os QTL possam chegar ao equilíbrio, pela taxa de recombinação, ou possam ser fixados na população, pela seleção.

2.4 Métodos de predição dos efeitos dos marcadores

Na busca do método ideal para a predição dos efeitos dos marcadores e dos valores genéticos genômicos, a escolha deste deve considerar que este acomode a arquitetura genética da característica, realizar a regularização do processo de estimação, considerando a presença de multicolinearidade e o grande número de marcadores (estimadores do tipo *shrinkage*), e realizar a seleção de covariáveis (marcadores) que afetam a característica em análise (RESENDE et al., 2012). A escolha do método quanto a arquitetura genética da característica deve ser realizada pelas pressuposições de cada método, por exemplo, se este considera a presença de genes de efeito maior.

Diversos métodos para estimação dos efeitos dos marcadores são descritos na literatura, possibilitando escolher o melhor método para a estimação dos efeitos dos marcadores para a característica em estudo. Estes métodos envolvem desde a

incorporação de informações moleculares na abordagem tradicional para a obtenção dos valores GBLUP, no qual a matriz de parentesco probabilística é substituída por uma matriz de parentesco real ou genômica (SCHAEFFER, 2006), métodos bayesianos, como o BayesA e o Bayes B (MEUWISSEN, et al., 2001; GIANOLA et al., 2009), o BayesC, o BayesD, e mais recentemente, BayesC π e BayesD π (HABIER et al., 2011), o LASSO Bayesiano (DE LOS CAMPOS et al., 2009; LI et al., 2011; PARK; CASELLA, 2008; SILVA et al., 2011), entre outros.

Quanto à escolha do método de estimação dos efeitos dos marcadores, alguns problemas podem ocorrer, principalmente pelo número de marcadores superar o número de animais genotipados, isto é, o número de parâmetros a serem estimados (efeitos dos marcadores) é muito maior que o número de observações. Se, por exemplo, for utilizado o método dos mínimos quadrados, haverá falta de graus de liberdade para ajustar simultaneamente todos os efeitos, pois neste caso, os graus de liberdade do resíduo são dados por $n - p$, em que n é número de indivíduos e p o número de marcadores, e p é muito maior do que n . Além disso, se a estimação dos efeitos dos marcadores for realizada de forma não simultânea, como por uma análise de regressão simples para cada marcador, pode ocorrer superestimação dos efeitos dos marcadores, reduzindo a acurácia do método e tornando possível apenas a identificação de QTLs de efeito maior (RESENDE et al., 2008).

De acordo com Gianola, Perez-Enciso e Toro (2003), contornar o problema de dimensionalidade, causado pelo número de marcadores ser maior que o número de indivíduos genotipados, demanda a utilização de métodos estatísticos que considerem a seleção de covariáveis (problema de multicolinearidade) e a regularização do processo de estimação (problema de dimensionalidade).

2.4.1 GBLUP e ssGBLUP

A obtenção das estimativas dos efeitos de marcadores pelo método GBLUP é caracterizada pela utilização das equações de Henderson (HENDERSON, 1984) para obtenção dos valores BLUP, que neste caso é dado para o vetor dos efeitos dos marcadores, sendo que a matriz de parentesco obtida pelos coeficientes de

Wright substituída pela matriz de parentesco genômica. Neste método é assumido que todos os *loci* ou genes explicam a mesma variância, considerando dessa forma, que a variância por *loci* é dada por uma razão da variância genética total e o número de *loci* (MEUWISSEN et al., 2001). Pressupõe-se que os efeitos dos marcadores seguem uma distribuição normal com média zero e variância comum a todos marcadores.

Quando são utilizadas apenas informações de pedigree e fenótipo para obter os valores genéticos, o modelo linear misto e as soluções, conforme Henderson (1984) e Mrode (2005), são:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{a} + \mathbf{e},$$

em que:

\mathbf{y} é o vetor das observações, sendo $\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{V})$, $\mathbf{V}=\mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R}$, $\mathbf{R}=\text{var}(\mathbf{e})$;

$\boldsymbol{\beta}$ é o vetor de efeitos fixos;

\mathbf{X} é a matriz de incidência, associando os elementos de $\boldsymbol{\beta}$ a \mathbf{y} ;

\mathbf{a} é o vetor de efeito aleatório, composto pelos valores genéticos dos indivíduos;

\mathbf{Z} é a matriz de incidência, associando respectivamente os elementos de \mathbf{a} a \mathbf{y} ;

\mathbf{e} é o vetor de resíduos aleatórios, NID $(0, \mathbf{I}\sigma^2)$.

Assim, no método G-BLUP ou BLUP genômico, a matriz de parentesco \mathbf{A} é substituída pela matriz de relacionamento genômica ou realizada, e neste caso, o vetor de efeitos aleatórios será o vetor de valores genéticos genômicos preditos, e os efeitos dos marcadores podem ser obtidos em um segundo passo, considerando a matriz \mathbf{Z} e o vetor de valores genéticos genômicos.

A matriz de parentesco genômica pode ser construída a partir das informações dos marcadores genéticos (matriz de relacionamento \mathbf{G} , método GBLUP), ou combinando as informações de pedigree e marcadores genéticos (matriz de relacionamento \mathbf{H} , método ssGBLUP), conforme descrito por Aguiar et al. (2010) e Legarra et al., 2009.

Primeiramente, o método ssGBLUP foi descrito para estudos de seleção genômica. Recentemente, Wang, H. et al. (2012) descreveram uma adaptação do método ssGBLUP para estudos de GWAS, sendo utilizado um processo iterativo com pesos para os marcadores genéticos, e a partir disso são obtidos os valores genéticos genômicos e os efeitos dos marcadores genéticos. Como o método

utilizado pelos autores foi o ssGBLUP, a matriz de relacionamento utilizada foi a H , conforme descrito por Aguilar et al. (2010), em que a inversa da matriz H dada por:

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

em que A é o numerador da matriz de relacionamento para todos os animais, A_{22} é o numerador da matriz de relacionamento para os animais genotipados, e G é a matriz de relacionamento genômico. Para obtenção dos valores genéticos genômicos e efeitos dos marcadores do tipo SNP, os pesos para os SNP foram obtidos no mesmo software em que as análises de associação genômica foram realizadas (família BLUPF90), e a cada iteração é permitido utilizar os pesos obtidos na iteração anterior para que os valores genéticos genômicos e/ou os efeitos dos marcadores genéticos sejam recalculados. O processo ocasiona em ganhos na detecção de QTL, por evidenciar regiões com a presença de QTL de maior efeito ou de maior número de QTL de pequeno efeito (WANG, H. et al., 2012).

2.4.2 Métodos bayesianos

A abordagem bayesiana também pode ser utilizada na predição dos valores genéticos dos animais e efeitos dos marcadores, sendo que as estimativas são obtidas considerando que a variância pode ser diferente entre *loci*, em que a variância explicada pelo *i-ésimo locus* é obtida a partir de uma distribuição (MEUWISSEN et al., 2001).

A abordagem bayesiana envolve métodos estatísticos que atribuem probabilidades ou distribuições a eventos ou parâmetros, nos permitindo incorporar informações em relação a quantidade de interesse desconhecida, por meio de $p(\theta)$, sendo θ o parâmetro desconhecido. A atualização da incerteza é feita utilizando o teorema de Bayes, pela combinação da *priori* ($p(\theta)$) e verossimilhança ($L(x|\theta)$), as quais nos levam à distribuição a *posteriori* de θ . Tal combinação é dada por:

$$p(\theta|x) \propto L(x|\theta)p(\theta)$$

Desta forma, utilizando a abordagem bayesiana temos um resumo da amostra estudada, o qual nos é dado pela verossimilhança, e podemos incorporar informações pela distribuição *a priori* (BOLSTAD, 2007).

O modelo estatístico na abordagem bayesiana (MEUWISSEN et al., 2001) é:

$$\mathbf{y} = \mu \mathbf{1}_n + \mathbf{Z}\mathbf{a} + \mathbf{e},$$

em que: \mathbf{y} é o vetor das observações; μ é a média geral; $\mathbf{1}_n$ é um vetor com valores um, de tamanho n ; \mathbf{u} é o vetor aleatório dos efeitos dos SNPs; \mathbf{Z} é a matriz de genótipos (com valores 0, 1 e 2, sendo 0 e 2 para os genótipos homocigotos e 1 para heterocigoto); \mathbf{e} é o vetor de efeitos residuais, $\mathbf{e} \sim N(0, \mathbf{I}\sigma_e^2)$.

O método BayesA foi descrito por Meuwissen et al. (2001), e considera que a variância pode ser diferente para os segmentos do genoma. Para este método, a informação *a priori* é que a variância dos marcadores com efeito maior do que zero tem distribuição qui-quadrado invertida e escalonada, e os efeitos dos marcadores podem ser ajustados a uma distribuição normal com média zero.

O método BayesB (MEUWISSEN et al., 2001), assim como o Bayes A, também considera que a variância genética explicada por cada *locus* pode ser diferente. A principal diferença entre os métodos BayesA e BayesB é que neste segundo método são aceitos marcadores genéticos com efeito igual a zero, sendo que a variância dos marcadores com efeito igual a zero é dada por π , e quando a variância explicada pelos marcadores com efeito maior do que zero tem distribuição qui-quadrado invertida e escalonada, com probabilidade $1 - \pi$, ou seja, π é a probabilidade de um SNP k ter efeito igual a zero (HABIER et al., 2011), e esta proporção é tratada como conhecida.

A probabilidade de um marcador apresentar efeito zero (π) afeta o encurtamento dos marcadores, e assim os métodos BayesA e BayesB parecem apresentar desvantagens em relação a outros, como o BayesC π e o BayesD π , pelo fato de que nos dois primeiros métodos a proporção de marcadores genéticos com efeito maior que zero ser tratada como conhecida (π), ao contrário dos métodos BayesC π e o BayesD π , nos quais a probabilidade π é tratada como desconhecida e deve ser inferida a partir dos dados. Comparando os métodos BayesC π e BayesD π , o primeiro método apresenta vantagens computacionais, mas assume que todos os

marcadores genéticos tem a mesma variância, enquanto que o método BayesD π assume diferentes variâncias (HABIER et al., 2011).

No método BayesC π (HABIER et al., 2011) as variâncias dos marcadores são consideradas homogêneas. A *priori* considerada para μ será uma constante. Considerando que \mathbf{A} é o numerador da matriz de relacionamento, a priori de $u|\mathbf{A}, \sigma^2_u$ será considerada normal, com média zero e variância $\mathbf{A}\sigma^2_u$ e σ^2_u é a variância genética aditiva, sendo π a probabilidade do marcador k ter efeito zero:

$$u_k|\pi, \sigma^2_{u_k} = \begin{cases} 0 \text{ com probabilidade} & \pi, \\ \sim N(0, \sigma^2_{u_k}) \text{ com probabilidade} & (1 - \pi). \end{cases}$$

Assim, π tem distribuição a priori uniforme (0,1), e assim o valor de π estará entre 0 e 1, com probabilidade igual para todas as possibilidades de valores dentro deste intervalo (RESENDE et al., 2012).

No método BayesD π as informações consideradas a *priori* para a variância dos efeitos dos marcadores são as mesmas do método BayesB, em que a variância dos marcadores com efeito maior do que zero tem distribuição qui-quadrado invertida e escalonada, com probabilidade $1 - \pi$, e π quando o efeito do marcador é igual a zero (HABIER et al., 2011).

As vantagens do uso dos métodos BayesC π e BayesD π estão no fato destes métodos propiciarem informação sobre a arquitetura genética das características (HABIER et al., 2011; WANG, C. et al., 2012), e assim são indicados para estudos de GWAS. O método BayesB considera π como conhecido, enquanto que nos métodos BayesC π e BayesD π este é tratado como desconhecido, sendo sua modelagem muito interessante para os estudos de associação, pelo fato de que a maioria das marcas não está em desequilíbrio de ligação com os QTLs e assim é necessária a seleção das marcas que estão em desequilíbrio de ligação com os QTLs responsáveis pela variação genética de um fenótipo (RESENDE et al., 2012).

Conforme apresentado por Meuwissen et al. (2001), a regressão Bayesiana pode ser utilizada nas situações em que se têm mais marcadores (covariáveis) do que observações, atribuindo distribuição a priori aos coeficientes de regressão. Essas distribuições impõem regularização no ajuste do modelo, sob a forma de encurtamento dos coeficientes de regressão (*shrinkage*). Entretanto, os autores também demonstraram que a forma dessa regularização deve ser diferenciada, pois

se espera que quando se tem SNPs de todo o genoma de um indivíduo muitos marcadores estarão em regiões que não influenciam o valor mensurado da característica, enquanto poucos estarão em desequilíbrio de ligação com alelos que influenciam a característica. Uma abordagem interessante para a situação descrita é a utilização do método de regressão LASSO (Least Absolute Shrinkage and Selection Operator, TIBSHIRANI, 1996), que combina seleção de variáveis e regularização via encurtamento dos coeficientes de regressão. A implementação Bayesiana da regressão LASSO (PARK; CASELLA, 2008) foi adaptada para seleção genômica por De Los Campos et al. (2009). Nesta adaptação, informações de relacionamento e outras covariáveis que não sofrem o efeito da regularização são consideradas no modelo.

2.5 Análises dos efeitos dos marcadores

Nos estudos de GWAS, deseja-se estudar a associação de marcadores genéticos com os QTLs responsáveis pela expressão fenotípica das características, e a evidência estatística para uma associação entre um alelo e um fenótipo pode pertencer a três situações. Em primeiro lugar, o alelo em si pode ser funcional e afetar diretamente a expressão do fenótipo. Em segundo lugar, o alelo pode estar correlacionado com, ou estar em desequilíbrio de ligação com, um alelo causador localizados nas proximidades. Terceiro, a associação pode ser atribuída ao acaso ou artefato, por exemplo, confusão ou viés de seleção (CARDON; PALMER, 2003).

Dessa forma, quando os efeitos dos marcadores genéticos são obtidos a partir de uma análise de regressão simples para cada marcador, torna-se necessário selecionar os marcadores que tem efeito real no fenótipo, o que pode ser um problema quando se trata de milhares de marcadores testados. Estes problemas estão relacionados com a realização de múltiplos testes de hipótese, e algumas considerações devem ser tomadas quanto ao nível de significância.

Há dois tipos de erros associados aos testes de hipótese: tipo I e tipo II. O erro tipo I ocorre quando se rejeita a hipótese H_0 quando esta é verdadeira, e o erro tipo II ocorre quando não rejeitamos H_0 quando está hipótese é falsa (TEDESCHI, 2006). Quando realizamos múltiplos testes de hipóteses, a probabilidade de cometermos pelo menos um erro do tipo I aumenta devido ao grande número de

hipóteses testadas, gerando assim falsos positivos (RESENDE et al., 2008). Uma alternativa seria realizar uma correção de Bonferroni, na qual tem-se um nível geral de significância α^* e um α individual para cada teste, e assim o nível de significância para cada teste é dado por $\alpha = \alpha^*/n$. Essa abordagem é conservativa e diminui o poder dos testes, permitindo a ocorrência de falsos positivos. Outro problema desta correção é que os testes são considerados independentes, porém, não podemos considerar que os marcadores em um mesmo cromossomo e que estão em desequilíbrio de ligação são independentes. Uma alternativa para a correção de Bonferroni descrita na literatura é o teste de permutação, proposto por Churchill e Doerge (1994) para contornar a questão de múltiplos testes nos experimentos de mapeamento de QTL.

Um critério mais apropriado para esse caso é a taxa de falsos positivos (FDR; *false discovery rate*) definida como a proporção esperada de falsos positivos dentre todos os testes significativos. Dentre os marcadores com efeitos significativos podemos ter: (i) os falsos positivos ou número de erros tipo I, nomeados pela letra “F”; (ii) os marcadores que realmente tem efeito, e são representados pela letra “V”; (iii) e o total de marcadores, dado pela soma de F e V, e representados pela letra “S” (BENJAMINI; HOCHBERG, 1995). Dos marcadores pertencentes ao grupo S, o número dos que realmente são significativos (V) é dado por: $(1 - FDR)$, e assim FDR pode ser definida como a proporção esperada de falsas descobertas dentre as hipóteses H_0 rejeitadas.

Adicionalmente ao conceito de FDR, tem-se o q-valor. Em um único teste de hipótese, é comum utilizar o p-valor para verificar se aceita-se ou rejeita-se H_0 . No entanto, o p-valor é uma medida de significância em termos da taxa de falsos positivos, enquanto que o q-valor é a medida FDR base de significância que pode ser calculada simultaneamente para múltiplos testes de hipóteses. O significado do q-valor é que se testes com q-valor menor ou igual a 0,05 são considerados significativos, o resultado disso é uma FDR de 5% entre os testes significativos, ou seja, 5% dos testes considerados significativos estão sob hipótese nula em média (STOREY; TIBSHIRANI, 2003).

Além das análises de regressão simples, os efeitos dos marcadores genéticos podem ser obtidos por regressão múltipla, evitando as correções para múltiplos testes. Neste caso, além do método ssGBLUP com pesos para os marcadores, no qual são evidenciadas regiões associadas aos fenótipos de interesse (WANG, H. et

al., 2012), as análises de associação podem ser realizadas por regiões do genoma, considerando o tamanho dos blocos de haplótipos. Cesar et al. (2014) utilizaram regiões genômicas, chamadas de janelas genômicas, no tamanho de 1Mpb para realizar análises de associação genômica em bovinos da raça Nelore.

Referências

- AGUILAR, I.; MISZTAL, I.; JOHNSON, D.L.; LEGARRA, A.; TSURUTA, S.; LAWLOR, T.J. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. **Journal of Dairy Science**, Champaign, v. 93, p.743–752, 2010.
- ANDERSON, C.A.; PETTERSSON F.H.; CLARKE, G.M.; CARDON, L.R.; MORRIS, P.; ZONDERVAN, K.T. Data quality control in genetic case-control association studies. **Nature Protocols**, London, v. 5, p.1564–1573, 2010.
- BENJAMINI, Y.; HOCHBERG, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. **Journal of the Royal Statistical Society**, Malden, v. 85, p. 289-300, 1995.
- BOLSTAD, W.M. **Introduction to Bayesian statistics**. 2nd ed. Hoboken: John Wiley, 2007. 463p.
- BOURDON, R.M. **Understanding animal breeding**. 2nd ed. Upper Saddle River: Prentice-Hall, 2000. 538p.
- CARDON, L.R.; PALMER, L.J. Population stratification and spurious allelic association. **Lancet**, London, v. 361, p. 598–604, 2003.
- CESAR, A.S.M.; REGITANO, L.C.A.; MOURÃO, G.B.; TULLIO, R.R.; LANNA, D.P.D.; NASSU, R.T.; MUDADU, M.M.; NASCIMENTO, M.L.; CHAVES, A.S.; ALENCAR, M.M.; SONSTEGARD, T.S.; GARRICK, D.J.; REECY, J.M.; COUTINHO, L.L. Genome-wide association study for intramuscular fat deposition and composition in Nelore cattle. **BMC Genetics**, London, v.15, p.39, 2014.
- CHURCHILL, G.A.; DOERGE, R.W. Empirical threshold values for quantitative trait mapping. **Genetics**, Austin, v. 138, p. 963-971, 1994.

COLE, J.B.; VANRADEN, P.M.; O'CONNELL, J.R.; VAN TASSELL, C.P.; SONSTEGARD, T.S.; SCHNABEL, R.D.; TAYLOR, J.F.; WIGGANS, G.R. Distribution and location of genetic effects for dairy traits. **Journal of Dairy Science**, Champaign, v. 92, p. 2931-2946, 2009.

DAETWYLER, H.D.; PONG-WONG, R.; VILLANUEVA, B.; WOOLLIAMS, J.A. The Impact of Genetic Architecture on Genome-Wide Evaluation Methods. **Genetics**, Austin, v. 185, p. 1021–1031, 2010.

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

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

DEKKERS, J.C.; HOSPITAL, F. The use of molecular genetics in the improvement of agricultural populations. **Nature Genetic**, New York, v. 3, p. 22-32, 2002.

FALCONER, D.S.; MACKAY, T.F.C. **Introduction to quantitative genetics**. 4th ed. Harlow: Longmans Green, 1996. 464p.

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

GIANOLA, D.; PEREZ-ENCISO, M.; TORO, M.A. On marker-assisted prediction of genetic value: beyond the ridge. **Genetics**, Austin, v. 163, p. 347-365, 2003.

GODDARD, M. Genomic selection: prediction of accuracy and maximization of long term response. **Genetics**, Austin, v. 136, p. 245–257, 2009.

GODDARD, M.E.; HAYES, B.J. Genomic selection. **Journal of Animal Breeding and Genetics**, Berlin, v. 124, p. 323-330, 2007.

GRIFFITHS, A.J.F.; MILLER, J.H.; SUZUKI, D.T.; LEWONTIN, R.C.; GELBART, W.M. **Introdução à genética**. 7.ed. New York: W. H. Freeman, 2000. 794p.

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

HARTL, D.L.; CLARK, A.G. **Principles of population genetics**. 4th ed. Sunderland: Sinauer, 2007. 652p.

HENDERSON, C.R. **Applications of Linear Models in Animal Breeding**. 3rd ed. Guelph: University of Guelph, 1984. 423p.

HICKEY, J.M.; GORJANC, G. Simulated data for genomic selection and genome-wide association studies using a combination of coalescent and gene drop methods. **Genes, Genomes and Genetics**, Toronto, v. 2, p. 425-427, 2012.

HILL W.G. Estimation of effective population-size from data on linkage disequilibrium. **Genetical Research**, Cambridge, v. 38, p. 209–216, 1981.

HILL, W.G.; ROBERTSON, A. Linkage disequilibrium in finite populations. **Theoretical and Applied Genetics**, New York, v. 38, p. 226-231, 1968.

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

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

MACLEOD, I.M.; HAYES, B.J.; SAVIN, K.; CHAMBERLAIN, A.J.; MCPARTLAN, H.; GODDARD, M.E. Power of dense bovine single nucleotide polymorphisms (SNPs) for genome scans to detect and position quantitative trait loci (QTL). **Journal of Animal Breeding and Genetics**, Berlin, v. 127, p. 133-142, 2010.

MALÉCOT, G. **The Mathematics of Heredity**. San Francisco: W.H. Freeman, 1969. 88p.

MATUKUMALLI, L.K.; LAWLEY, C.T. ; SCHNABEL, R.D.; TAYLOR, J.F.; ALLAN, M.F.; HEATON, M.P.; O'CONNELL, J.; MOORE, S.S.; SMITH, T.P.L.; SONSTEGARD, T.S.; VAN TASSELL, C.P. Development and characterization of a high density SNP genotyping assay for cattle. **PLoS One**, San Francisco, v. 4, e5350, 2009.

MEUWISSEN, T.H.E.; SOLBERG, T.R.; SHEPHERD, R.; WOOLIAMs, J.A. A fast algorithm for BayesB type of prediction of genome-wide estimates of genetic value. **Genetics Selection Evolution**, London, v.41, p.2, 2009.

MEUWISSEN, T.H.E. Genomic selection: marker assisted selection on genome-wide scale. **Journal of Animal Breeding and Genetics**, Berlin, v. 124, p. 321-322, 2007.

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

MRODE, R.A. **Linear models for prediction of animal breeding values**. 2nd ed. Wallingford: CABI, 2005. 344p.

PARK, T.; CASELLA, G. The Bayesian LASSO. **Journal of the American Statistical Association**, Alexandria, v. 103, p. 681-686, 2008.

PRITCHARD, J.K.; PRZEWORSKI, M. Linkage disequilibrium in humans: models and data. **American Journal of Human Genetics**, Baltimore, v. 69, p. 1-14, 2001.

PURCELL, S., NEALE, B.; TODD-BROWN, K.; THOMAS, L.; FERREIRA, M.A.; BENDER, D.; MALLER, J.; SKLAR, P.; DE BAKKER, P.I.; DALY, M.J.; SHAM, P.C. PLINK: a tool set for whole-genome association and population-based linkage analyses. **American Journal of Human Genetics**, Baltimore, v. 81, p. 559-575, 2007.

REICH, D.E.; CARGILL, M.; BOLK, S.; IRELAND, J.; SABETI, P. C.; RICHTER, D.J.; LAVERY, T.; KOUYOUJIAN, R; FARHADIAN, S.F.; WARD, R; LANDER, E.S. Linkage disequilibrium in the human genome. **Nature**, New York, n. 411, p. 199-204, 2001.

RESENDE, M.D.V.; SILVA, F.F.; LOPES, P.S.; AZEVEDO, C.F. **Seleção Genômica Ampla (GWS) via Modelos Mistos (REML/BLUP), Inferência Bayesiana (MCMC), Regressão Aleatória Multivariada (RRM) e Estatística Espacial**. Viçosa: Universidade Federal de Viçosa, 2012. 291p.

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

SARGOLZAEI, M.; SCHENKEL, F. S. QMSim: a largescale genome simulator for livestock. **Bioinformatics**, Oxford, v. 25, p. 680-681, 2009.

SARGOLZAEI, M.; SCHENKEL, F.S.; JANSEN, G. B.; SCHAEFFER, L.R. Extent of linkage disequilibrium in Holstein cattle in North America. **Journal of Dairy Science**, Champaign, v. 91, p. 2106-2117, 2008.

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

SETTLES, M.; ZANELLA, R.; MCKAY, S. D.; SCHNABEL, R.D.; TAYLOR, J.F.; WHITLOCK, R.; SCHUKKEN, Y.; VAN KESSEL, J.S.; SMITH, J.M.; NEIBERGS; H.A whole genome association analysis identifies loci associated with Mycobacterium avium subsp. paratuberculosis infection status in US holstein cattle. **Animal Genetics**, New York, v. 40, p. 655-662, 2009.

SILVA, F.F.; VARONA, L.; RESENDE, M.D.V.; BUENO FILHO, J.S.S.; ROSA, G. J.M.; VIANA, J.M.S. A note on accuracy of Bayesian LASSO regression in GWS. **Livestock Science**, Amsterdam, v. 141, p. 310-314, 2011.

STOREY, J.D.; TIBSHIRANI, R. Statistical significance for genome-wide studies, **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 100, p. 9440-9445, 2003.

TEDESCHI, L.O. Assessment of the adequacy of mathematical models. **Agricultural Systems**, Essex, v. 89, p. 225-247, 2006.

TIBSHIRANI, R. Regression shrinkage and selection via the Lasso. **Journal of the Royal Statistics Society Series B**, London, v. 58, p. 267-288, 1996.

TURNER, S.; ARMSTRONG, L.L.; BRADFORD, Y.; CARLSON, C.S.; CRAWFORD, D.C.; CRENSHAW, A.T.; ANDRADE, M.; DOHENY, K.F.; HAINES, J.L.; HAYES, G.; JARVIK, G.; JIANG, L.; KULLO, I.J.; LI, R.; LING, H.; MANOLIO, T.A.; MATSUMOTO, M.; MCCARTY, C.A.; MCDAVID, A.N.; MIREL, D.B.; PASCHALL, J.E.; PUGH, E.W.; RASMUSSEN, L.V.; WILKE, R.A.; ZUVICH, R.L.; RITCHIE, M.D. Quality Control Procedures for Genome Wide Association Studies. **Current Protocols Human Genetics**, New York, v. 68, p. 1-18, 2011.

WANG, C.; MA, P.; ZHANG, Z.; DING, X.; LIU, J.; FU, W.; WENG, Z.; ZHANG, Q. Comparison of five methods for genomic breeding value estimation for the common dataset of the 15th QTL-MAS Workshop. **BMC Proceedings**, London, v. 6, p. 1-5, 2012.

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

WEEDON, M.N.; LANGO, H.; LINDGREN, C.M.; WALLACE, C.; EVANS, D.M.; MANGINO, M.; FREATHY, R.M.; PERRY, J.R.; STEVENS, S.; HALL, A.S.; SAMANI, N.J.; SHIELDS, B.; PROKOPENKO, I.; FARRALL, M.; DOMINICZAK, A.; JOHNSON, T.; BERGMANN, S.; BECKMANN, J.S.; VOLLENWEIDER, P.; WATERWORTH,

D.M.; MOOSER, V.; PALMER, C.N.; MORRIS, A.D.; OUWEHAND, W.H.; ZHAO, J.H.; LI, S.; LOOS, R.J.; BARROSO, I.; DELOUKAS, P.; SANDHU, M.S.; WHEELER, E.; SORANZO, N.; INOUE, M.; WAREHAM, N.J.; CAULFIELD, M.; MUNROE, P.B.; HATTERSLEY, A.T.; MCCARTHY, M.I.; FRAYLING, T.M. Genome-wide association analysis identifies 20 loci that influence adult height. **Nature Genetics**, New York, v. 54, p. 575-583, 2008.

WIGGANS, G.R.; SONSTEGARD, T.S.; VANRADEN, P.M.; MATUKUMALLI, L.K.; SCHNABEL, R.D.; TAYLOR, J.F.; SCHENKEL, F.S.; VAN TASSELL, C.P. Selection of single-nucleotide polymorphisms and quality of genotypes used in genomic evaluation of dairy cattle in the United States and Canada. **Journal of Dairy Science**, Champaign, v. 92, p. 3431-3436, 2009.

WRIGHT, S. Coefficients of inbreeding and relationship. **The American Naturalist**, Chicago, v. 56, p. 330-338, 1922.

ZANELLA, R. **Identification of chromosomal regions associated with infectious diseases in cattle**. 2011. 254p. Dissertação (Doctor of Philosophy in Animal Science) - Washington State University, Washington, 2011.

ZANELLA, R.; SETTLES, M.L.; MCKAY, S.D.; SCHNABEL, R.; TAYLOR, J.; WHITLOCK, R.H.; SCHUKKEN, Y.; VAN KESSEL, J.S.; SMITH, J.M.; NEIBERGS, H.L. Identification of loci associated with tolerance to Johne's disease in Holstein cattle. **Animal Genetics**, Oxford, v. 42, p. 28-38, 2011.

3 SELEÇÃO E ASSOCIAÇÃO GENÔMICA AMPLA CONSIDERANDO SITUAÇÕES REAIS DE PROGRAMAS DE MELHORAMENTO GENÉTICO ANIMAL

Resumo

Informações de milhares de marcadores genéticos têm sido incluídas nos programas de melhoramento genético, permitindo a seleção dos animais considerando estas informações e a identificação de regiões genômicas associadas às características de interesse econômico. O objetivo deste trabalho foi avaliar a eficiência do método ssGBLUP utilizando pesos para os marcadores genéticos, informações de genótipo e fenótipos, com ou sem as informações de pedigree, para seleção e associação genômica ampla, considerando diferentes coeficientes de herdabilidade. A população estudada foi obtida por simulação de dados, e foi composta por 8.150 animais, sendo 5.850 animais genotipados. Os dados simulados foram analisados utilizando o método ssGBLUP com matrizes de relacionamento com ou sem informações de pedigree, utilizando pesos para os marcadores genéticos obtidos em cada iteração. Quanto maior o coeficiente de herdabilidade, melhores foram os resultados de seleção e associação genômica. Para a identificação de regiões associadas a características de interesse, não houve influência do tipo de matriz de relacionamento utilizada. Para as características com e sem efeito poligênico, quando considerado o mesmo coeficiente de herdabilidade, não houve diferenças para seleção genômica, mas a identificação de QTL foi melhor nas características sem efeito poligênico. Foram obtidas maiores acurácias quando utilizados os métodos bayesianos para predição dos efeitos dos marcadores genéticos e dos valores genéticos, em comparação com o método GBLUP. Quanto maior a pressão de seleção, mais acuradas foram as predições dos valores genéticos genômicos. O método ssGBLUP foi eficiente tanto para a seleção quanto para a identificação de regiões associadas às características estudadas.

Palavras-chaves: Bayesiana; Coeficiente de herdabilidade; ssGBLUP; Validação cruzada

Abstract

Data from thousands genetic markers have been included in animal breeding programs to allow the selection of animals considering this information and to identify genomic regions associated to traits of economic interest. The objectives of this study were to evaluate the efficiency of the ssGBLUP method using genotype and phenotype information, with or without pedigree information, and to attribute weights for genetic markers, for selection and for genome-wide association considering different coefficients of heritability. The population of simulated data was composed by 8.150 individuals and 5.850 genotyped animals. The simulated data was analyzed by the ssGBLUP method and by two relationship matrixes, with or without pedigree information, and weights were obtained for genetic markers in every iteration. The results of selection and genomic association were better for the traits with the highest heritability coefficients. There was no difference between the relationship matrix studied to identify the regions associated with traits of interest. The greater accuracies were obtained from Bayesian methods, in comparison with the GBLUP method. For traits with the greater selection pressure, more accurate predictions of

the genomic breeding values were obtained. The ssGBLUP method was efficient for both selection and to the identification of regions associated with the studied traits.

Keywords: Bayesian; Cross-validation; Heritability coefficient; ssGBLUP

3.1 Introdução

A seleção genômica ampla foi proposta por Meuwissen et al. (2001) visando estimar o mérito genético de indivíduos sob seleção por meio da utilização de marcadores de alta densidade (dezenas ou centenas de milhares marcadores) cujos efeitos são estimados simultaneamente em um mesmo modelo preditivo (MEUWISSEN, 2007). A partir da estimação de tais efeitos, torna-se possível prever os valores genéticos dos indivíduos para características de interesse econômico de forma mais acurada, e também identificar regiões do genoma que contribuem de forma mais efetiva para a expressão da característica em questão. Mesmo apresentando as mencionadas vantagens, a inclusão desta tecnologia em programas de melhoramento genético animal ainda apresenta limitações, principalmente pelos altos custos envolvidos no processo de genotipagem, o que limitam o número de animais genotipados.

No melhoramento genético tradicional, os valores genéticos dos animais são preditos utilizando dados fenotípicos e de pedigree, via metodologia dos modelos mistos. Neste caso, a matriz de parentesco (ou relacionamento, denominada A) é obtida por meio do valor esperado da proporção de *loci* idênticos por descendência (IBD). Tal valor geralmente é denominado de coeficiente de parentesco. Porém, ao utilizar informações de marcadores moleculares de alta densidade, tal matriz pode ser substituída pela matriz de parentesco genômica (ou realizada), a qual é construída considerando apenas as informações destes marcadores, ou seja, apenas é válida para animais genotipados. Nesta matriz, denominada G, os coeficientes são dados pela proporção de locus idênticos por estado (IBS), os quais geralmente capturam mais informações que o coeficiente tradicional. Existe ainda a possibilidade de se utilizar uma matriz que combina os coeficientes IBD e IBS, denominada H, a qual permite utilizar simultaneamente informações de parentesco de indivíduos genotipados e não-genotipados, o que reflete a atual situação atual da

maioria dos programas de melhoramento animal, nos quais nem todos os indivíduos foram genotipados.

A utilização desta matriz H é postulada em um método denominado “single step”, ou ssGBLUP (AGUILAR et al., 2010; LEGARRA et al. 2009), o qual tem-se mostrado altamente eficiente quanto a predição de valores genéticos quando comparado com métodos mais complexos de seleção genômica, tais como as regressões Bayesianas (WANG, H. et al., 2012). Porém, ainda são escassos estudos considerando também tais comparações para as estimativas dos efeitos de marcadores (além dos valores genéticos), principalmente em situações envolvendo de diferentes graus de ausência de animais genotipados e valores de herdabilidades para as características consideradas.

Diante da escassez de estudos desta natureza, a simulação de dados apresenta-se como uma importante alternativa para explorar situações em que se deseja testar novas metodologias (tal como ssGBLUP), ou quando o conjunto de dados de interesse apresenta características particulares, como diferentes estruturas de populações ou quebra de pressuposições geralmente assumidas pelos modelos em questão (HICKEY; GORJANC, 2012). Adicionalmente, a simulação de dados passa a ser especialmente útil para estudar fenótipos cujas observações muitas vezes são impossibilitadas de serem coletadas durante a vida produtiva dos animais, tal como as características de qualidade de carne, as quais geralmente são provenientes de abatedouros e apresentam limitações de informações básicas tais como grupo racial, idade, e genealogia (pedigree).

Assim, o objetivo deste trabalho foi avaliar a eficiência do método ssGBLUP para seleção e associação genômica utilizando informações fenotípicas e genotípicas, com ou sem as informações de pedigree, e diferentes coeficientes de herdabilidade. Para tanto, foi proposto um sistema de simulação de dados que permite mimetizar estes cenários na atual situação dos programas de melhoramento genético animal.

3.2 Material e métodos

3.2.1 Avaliação da importância das informações de pedigree, coeficiente de herdabilidade e efeito poligênico

3.2.1.1 Dados simulados

Os dados foram simulados utilizando o software QMSim (SARGOLZAEI; SCHENKEL, 2009), utilizando parâmetros semelhantes aos considerados por Brito et al. (2011) e Neves et al. (2014). A população histórica foi constituída inicialmente por 1.000 indivíduos por geração, número que foi mantido até a geração 1.000, seguido de uma redução gradual no tamanho da população até a geração 2.020, a qual foi composta por 200 indivíduos. Além disso, esta população foi gerada considerando acasalamentos aleatórios, assim as progênies foram geradas a partir da união aleatória de gametas, pois os gametas do macho e da fêmea que geraram cada progênie foram obtidos a partir de pools de gametas de machos e fêmeas. Além disso, as frequências alélicas na primeira geração foram fixadas em 0,5, e também foi considerado que não houve migração, as gerações foram discretas, a abordagem utilizada para simular os dados foi forward-time, e o número de machos e fêmeas foi igual em cada uma das gerações.

No próximo passo da simulação, os animais da última geração da população histórica foram considerados como fundadores (100 machos e 100 fêmeas) da população de expansão, a qual foi composta por oito gerações, considerando cinco progênies por fêmea por geração, crescimento exponencial do número de fêmeas, união aleatória de gametas e ausência de seleção.

A população recente foi simulada em dois passos. Para o primeiro passo, foram selecionados de forma aleatória 640 machos e 32.000 fêmeas da última geração da população de expansão. Os acasalamentos foram realizados de forma aleatória, sendo que os animais que permaneceram no rebanho a cada geração foram os com maiores valores genéticos, considerando uma taxa de substituição de 60% nos machos e 20% nas fêmeas. Os animais com menores valores genéticos em cada geração foram descartados. O número de progênies considerado foi um descendente por fêmea por geração. Para o segundo passo da população recente, o grupo de fundadores da população de estudo foi formado a partir da seleção aleatória de 50 machos e 1350 fêmeas na última geração da primeira população recente, ou seja, da população formada no passo anterior. Estes indivíduos formaram um rebanho constituído por cinco gerações de seleção para os indivíduos com valores genéticos altos, e como na população anterior, os acasalamentos foram

realizados de forma aleatória, considerando uma taxa de substituição de 20% nos machos e 5% nas fêmeas, e descarte dos animais com menores valores genéticos.

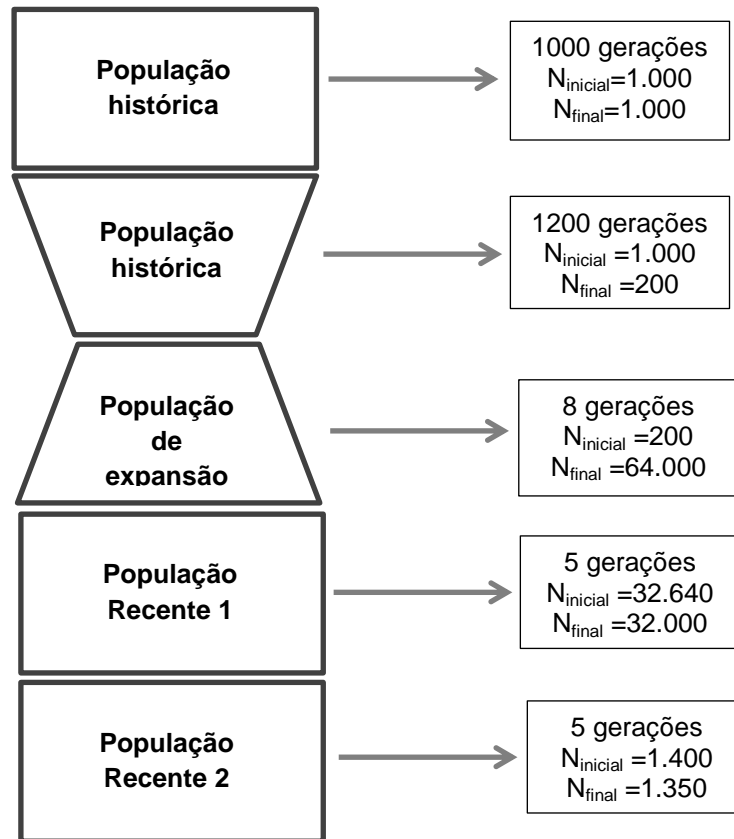


Figura 1 – Fluxograma da população de dados simulados, com número de gerações e número de indivíduos inicial ($N_{inicial}$) e final (N_{final}) para cada população

O genoma foi simulado com cinco cromossomos de 100 cM cada, nos quais foram distribuídos de forma equidistante 7.500 marcadores bialélicos, sendo aproximadamente 1.500 marcadores segregantes por cromossomo. Foi considerada a presença de 30 QTL distribuídos aleatoriamente ao longo dos cinco cromossomos e os efeitos dos QTL foram simulados considerando a distribuição gama com parâmetro shape 0,4. Os QTL e marcadores genéticos foram obtidos considerando MAF de 0,01. As taxas de mutação consideradas foram: 1×10^{-5} para *locus* de marcadores genéticos e de QTL, ambos por geração. Tanto os marcadores genéticos, como os QTL, foram considerados como bi-alélicos, e assim as mutações geradas alteraram o alelo 1 para o 2 ou o alelo 2 para o 1, desconsiderando a possibilidade de gerar novos alelos.

Foram simulados seis diferentes fenótipos, considerando diferentes coeficientes de herdabilidade e a presença ou ausência do efeito poligênico, conforme apresentado na Tabela 1.

Tabela 1 - Descrição dos diferentes fenótipos simulados

Característica ¹	h_{QTL}^2	h^2
1	0,40	0,40
2	0,20	0,40
3	0,25	0,25
4	0,13	0,25
5	0,10	0,10
6	0,05	0,10

¹Característica 1: herdabilidade 0,40 e sem efeito poligênico; Característica 2: herdabilidade 0,40 e com efeito poligênico; Característica 3: herdabilidade 0,25 e sem efeito poligênico; Característica 4: herdabilidade 0,25 e com efeito poligênico; Característica 5: herdabilidade 0,10 e sem efeito poligênico; Característica 6: herdabilidade 0,10 e com efeito poligênico; h_{QTL}^2 : coeficiente de herdabilidade oriundo dos efeitos aditivos dos QTLs; h^2 : coeficiente de herdabilidade oriundo dos efeitos genético aditivo poligênico e de QTL.

Para todas as características, a variância fenotípica considerada foi 1,0. Para cada característica, as simulações foram repetidas 10 vezes, e os resultados foram obtidos a partir das médias e desvios-padrão das repetições para cada cenário.

3.2.1.2 Predição dos valores genéticos genômicos e dos efeitos de marcadores genéticos por meio dos métodos GBLUP e ssGBLUP

Para a predição dos valores genéticos genômicos, o método empregado foi o ssGBLUP, sendo que os dados foram analisados no software BLUPF90.

Cenário 1: as análises foram realizadas considerando o pedigree, e neste caso, a matriz de parentesco utilizada foi a H, conforme descrito por Aguilar et al. (2010), na qual são combinados dados de genótipos e pedigree. A matriz H é dada por:

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

em que A é o numerador da matriz de relacionamento para todos os animais, A_{22} é o numerador da matriz de relacionamento para os animais genotipados, e G é a matriz de relacionamento genômico.

Cenário 2: As análises de associação foram realizadas apenas com dados de genótipos e fenótipos, e assim a matriz de relacionamento utilizada foi a genômica (G), a qual é construída apenas com as informações dos marcadores genéticos.

Para os dois cenários, o modelo utilizado foi:

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Z}\mathbf{u} + \mathbf{e},$$

em que: \mathbf{y} é o vetor de fenótipos simulados, $\mathbf{1}$ é o vetor de uns, μ é a média geral dos fenótipos, \mathbf{u} é o vetor de efeitos individuais dos animais, \mathbf{Z} é a matriz de incidência que relaciona os indivíduos aos fenótipos e \mathbf{e} é o vetor de resíduos.

A partir dos métodos GBLUP e ssGBLUP, Wang, H. et al (2012) propuseram utilizar os valores genéticos preditos por estes métodos para obtenção dos efeitos dos marcadores, a partir da equação:

$$\hat{\mathbf{a}} = (\mathbf{Z}'\mathbf{Z})^{-1}(\mathbf{Z}'\hat{\mathbf{u}}),$$

em que: $\hat{\mathbf{a}}$ é o vetor de efeitos preditos dos marcadores genéticos, \mathbf{Z} é a matriz de incidência que relaciona os indivíduos aos fenótipos e \mathbf{u} é o vetor de valores genéticos dos animais. Além disso, a predição dos valores genéticos genômicos e dos efeitos dos marcadores genéticos por ser realizada com diferentes iterações, utilizando pesos para as efeitos dos marcadores gerados em cada iteração, processo que evidencia regiões genômicas associadas aos fenótipos de interesse. Assim, foram utilizadas três iterações para a predição dos valores genéticos genômicos e dos efeitos dos marcadores genéticos.

A validação cruzada foi realizada a partir da divisão da população de estudo em dois grupos: as gerações 1, 2, 3 e 4 formaram o grupo de treinamento, enquanto que a geração 5, para a qual os fenótipos foram zerados, foi considerada como grupo de validação, possibilitando o estudo da habilidade de predição nas gerações futuras.

A eficiência de predição dos valores genéticos genômicos foi avaliada pelos valores obtidos para as correlações entre os valores genéticos verdadeiros e os preditos (acurácias), para as diferentes iterações, tanto para o grupo de validação quanto para o rebanho. Para as análises de associação genômica ampla foram construídos gráficos de soluções dos efeitos de SNPs e QTLs, para cada cenário,

considerando apenas a iteração de maior correlação entre os valores genéticos preditos e verdadeiros. Para a análise de convergência, foram calculadas correlações entre os valores genéticos genômicos obtidos, para toda a população, em diferentes iterações.

3.2.1.3 Predição dos valores genéticos genômicos via métodos bayesianos

A predição dos valores genéticos utilizando a abordagem bayesiana foi realizada por meio dos métodos BayesA, BayesB, BayesC π e LASSO bayesiano.

O modelo considerado foi:

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Z}\mathbf{a} + \mathbf{e},$$

em que: \mathbf{y} é o vetor de fenótipos simulados, $\mathbf{1}$ é o vetor de uns, μ é a média geral dos fenótipos, \mathbf{a} é o vetor de efeitos dos marcadores genéticos, \mathbf{Z} é a matriz de incidência que relaciona os indivíduos aos fenótipos e \mathbf{e} é o vetor de resíduos.

O vetor de efeitos dos marcadores genéticos foi predito para os animais do grupo de treinamento, e os valores genéticos genômicos para o grupo de validação foram obtidos por:

$$\hat{\mathbf{u}} = \mathbf{Z}\hat{\mathbf{a}},$$

em que \mathbf{u} é o vetor de valores genéticos genômicos, \mathbf{Z} é a matriz de incidência dos marcadores SNPs e \mathbf{a} é o vetor de efeitos dos SNPs, obtido na população de treinamento. Assim, as acurácias para o grupo de validação foram obtidas pelas correlações entre os valores genéticos genômicos preditos e os valores genéticos verdadeiros.

Para os métodos BayesA e BayesB, foi utilizado o *software* GWP (MEUWISSEN et al., 2009), considerando diferentes valores de π para o método BayesB (0,7, 0,5 e 0,3). Para os métodos BayesC π e BLASSO, foi utilizado o pacote “BGLR”, disponível para o *software* R, para os quais o número de iterações considerado foi 60.000, com burn-in de 40.000 e thinning igual a 6.

3.2.2 Avaliação da importância de diferentes números de QTL e pressões de seleção para seleção genômica

Para a avaliação da influência do número de regiões, ao longo do genoma, associadas a um fenótipo, os dados também foram simulados utilizando o software QMSim (SARGOLZAEI; SCHENKEL, 2009), com parâmetros semelhantes aos considerados por Brito et al. (2011) e Neves et al. (2014), e assim a população de estudo foi simulada utilizando parâmetros idênticos aos considerados para a avaliação da importância das informações de pedigree, coeficiente de herdabilidade e efeito poligênico. Assim como para os parâmetros populacionais, o genoma foi simulado de forma semelhante, com diferença no número de QTL. Para cada fenótipo, foram simulados, em média, 6, 30 e 100 QTL por cromossomo, com coeficiente de herdabilidade 0,40, sendo este oriundo apenas dos efeitos genéticos aditivos dos QTL.

Para estudar a influência de diferentes pressões de seleção, a população de estudo foi simulada utilizando parâmetros para a população histórica e de expansão idênticos aos considerados para a avaliação da importância das informações de pedigree, coeficiente de herdabilidade e efeito poligênico, assim como para o genoma. Para os três cenários, foi considerado um coeficiente de herdabilidade de 0,40, oriundo apenas dos efeitos genéticos aditivos dos QTL. As taxas de substituição de indivíduos na segunda população recente foram diferentes para os três cenários:

- Cenário 1: 60% nos machos e 20% nas fêmeas, em ambas as populações recentes, com descarte dos animais com menores valores genéticos;
- Cenário 2: taxa de substituição de 60% nos machos e 20% nas fêmeas na primeira população recente e 20% dos machos e 5% das fêmeas na segunda população recente, com descarte dos animais com menores valores genéticos;
- Cenário 3: 20% dos machos e 5% das fêmeas em ambas as populações recentes, com descarte dos animais pela idade.

Para cada cenário, as simulações foram repetidas 10 vezes, e os resultados foram obtidos a partir das médias e desvios-padrão das repetições.

O método de predição dos valores genéticos genômicos foi o ssGBLUP, sendo os valores genéticos recalculadas para cada iteração realizada com base dos pesos dos marcadores genéticos, conforme descrito por Wang, H. et al (2012).

O modelo utilizado foi:

$$y = \mathbf{1}\mu + \mathbf{Z}u + e,$$

em que: y é o vetor de fenótipos simulados, $\mathbf{1}$ é o vetor de uns, μ é a média geral dos fenótipos, u é o vetor de efeitos individuais dos animais, sendo que $u \sim N(\mathbf{0}, H\sigma_u^2)$, Z é a matriz de incidência que relaciona os indivíduos aos fenótipos e e é o vetor de resíduos.

A validação cruzada foi realizada a partir da divisão da população de estudo em dois grupos: as gerações 1, 2, 3 e 4 formaram o grupo de treinamento, enquanto que a geração 5, para a qual os fenótipos foram zerados, foi considerada como grupo de validação, possibilitando o estudo da habilidade de predição nas gerações futuras.

A eficiência de predição dos valores genéticos genômicos foi avaliada pelos valores obtidos para as correlações (acurácias) entre os valores genéticos verdadeiros e os preditos, para as diferentes iterações, tanto para o grupo de validação quanto para o rebanho.

3.3 Resultados e discussão

3.3.1 Avaliação da importância das informações de pedigree, coeficiente de herdabilidade e efeito poligênico na seleção genômica via métodos ssGBLUP e GBLUP

Para a análise de convergência (Tabela 2) foram calculadas as correlações entre os valores genéticos genômicos obtidos em diferentes iterações. A cada iteração, foram utilizados os pesos obtidos para os SNPs na iteração anterior, iniciando com peso 1 para todos os marcadores genéticos. Com os pesos dos marcadores genéticos sendo computados a cada iteração, marcadores genéticos com menor efeito terão menores efeitos a cada iteração, assim os marcadores

genéticos de maior efeito terão seus efeitos maximizados. Assim, a maior acurácia será obtida com a combinação dos pesos que minimize a estimação dos erros (WANG, H. et al., 2012).

Na análise da convergência, todos os fenótipos estudados tiveram resultados semelhantes, variando de 0,96 a 0,99 entre a primeira e segunda iteração, e 0,98 e 1,00 na segunda e terceira iteração. Assim, as semelhanças entre os valores genéticos são indicativos de que não há necessidade de que novas iterações sejam feitas. Para os fenótipos com maior coeficiente de herdabilidade foram obtidas as maiores correlações, indicativo de convergência mais rápida para fenótipos com maior herdabilidade. Os valores obtidos para os desvios-padrões, de 0,001 até 0,008 são indicativos de pequena variação entre as repetições dentro de cada fenótipo e cenário.

Tabela 2 - Correlações (desvios-padrão) entre valores genéticos preditos obtidos em diferentes iterações

Característica ¹	h^2	h_{QTL}^2	Matriz de relacionamento	it1-it2	it2-it3
1	0,40	0,40	H	0,99 (0,002)	1,00 (0,001)
	0,40	0,40	G	0,98 (0,004)	0,99 (0,001)
2	0,40	0,20	H	0,99 (0,002)	1,00 (0,001)
	0,40	0,20	G	0,98 (0,004)	0,99 (0,002)
3	0,25	0,25	H	0,98 (0,003)	1,00 (0,001)
	0,25	0,25	G	0,98 (0,005)	0,99 (0,001)
4	0,25	0,13	H	0,98 (0,002)	0,99 (0,002)
	0,25	0,13	G	0,98 (0,004)	0,99 (0,002)
5	0,10	0,10	H	0,97 (0,008)	0,99 (0,006)
	0,10	0,10	G	0,96 (0,006)	0,98 (0,003)
6	0,10	0,05	H	0,97 (0,003)	0,99 (0,006)
	0,10	0,05	G	0,96 (0,003)	0,98 (0,003)

¹it1 a 3=iterações 1 a 3; Característica 1: herdabilidade 0,40 e sem efeito poligênico; Característica 2: herdabilidade 0,40 e com efeito poligênico; Característica 3: herdabilidade 0,25 e sem efeito poligênico; Característica 4: herdabilidade 0,25 e com efeito poligênico; Característica 5: herdabilidade 0,10 e sem efeito poligênico; Característica 6: herdabilidade 0,10 e com efeito poligênico.

As acurácias dos valores genéticos genômicos preditos foram obtidas como correlações entre os valores genéticos verdadeiros e os valores genéticos genômicos preditos no grupo de validação (Tabela 3) e no rebanho estudado (Tabela 4).

Os valores obtidos para as acurácias no grupo de validação foram maiores para os fenótipos com maior coeficiente de herdabilidade, comparando os resultados dentro de cada cenário. Porém, para os fenótipos 1, 2, 3 e 4, ou seja, quando o coeficiente de herdabilidade foi igual a 0,40 e 0,25, as acurácias foram próximas.

Quando comparados os fenótipos com e sem efeito poligênico no grupo de validação, foram obtidos maiores acurácias para os fenótipos com herdabilidade

0,40 e 0,25 quando não foi simulado efeito poligênico, e somente para as características com baixa herdabilidade (fenótipos 5 e 6) o fenótipo com efeito poligênico teve maior acurácia do que quando havia apenas o efeito aditivo de QTL. No rebanho, as diferenças de acurácias entre fenótipos com ou sem efeito poligênico foram pequenas, sendo indicativo que não há uma diferenciação da presença do efeito poligênico utilizando o método ssGBLUP.

As acurácias obtidas foram maiores quando utilizada a matriz H em relação a matriz G, tanto no rebanho quanto no grupo de validação. Assim, o uso de uma matriz de relacionamento que combine informações de pedigree e genômicas parece ser mais eficiente na predição dos valores genéticos genômicos do que apenas as informações genômicas, para a população estudada. Nos resultados obtidos para os desvios-padrões, houve uma maior variação quando foi utilizada a matriz G em substituição na matriz H. Além disso, houve uma menor variação entre as repetições no rebanho do que no grupo de validação.

Para o fenótipo com herdabilidade igual a 0,40 sem efeito poligênico, foram obtidas altas correlações entre valores genéticos verdadeiros e preditos de 0,87, 0,85, 0,85, utilizando as informações de pedigree, respectivamente para cada iteração.

Wang, H. et al. (2012) estudaram uma característica com coeficiente de herdabilidade de 0,50, oriundo apenas dos efeitos aditivos dos QTL, em uma população contendo cinco gerações, pedigree com 15.800 animais e 1.500 indivíduos genotipados. Nesse caso, para o método ssGBLUP com processo iterativo, utilizando a matriz H para obtenção dos valores genéticos genômicos, foram obtidas as correlações (desvios-padrão) de 0,87 (0,01), 0,89 (0,01), 0,88 (0,01), para a primeira, segunda e terceira iteração. Os valores mais altos obtidos neste estudo podem ser devidos ao maior coeficiente de herdabilidade utilizado, e as diferenças entre iterações podem estar relacionadas com as diferentes estruturas das populações. Segundo os mesmos autores, as vantagens do uso das iterações para recalcular os valores genéticos genômicos estão relacionadas ao número e distribuição dos efeitos dos QTLs, sendo que quando há um pequeno número de QTLs parece não haver melhoria na acurácia na predição dos valores genéticos genômicos.

Tabela 3 - Correlações entre valores genéticos preditos e verdadeiros em diferentes iterações considerando o grupo de validação

Característica ¹	h^2	h_{QTL}^2	Matriz de relacionamento	i1	i2	i3
1	0,40	0,40	H	0,68 (0,03)	0,67 (0,03)	0,67 (0,03)
	0,40	0,40	G	0,53 (0,09)	0,53 (0,09)	0,53 (0,09)
2	0,40	0,20	H	0,63 (0,04)	0,62 (0,04)	0,62 (0,04)
	0,40	0,20	G	0,48 (0,09)	0,48 (0,08)	0,47 (0,08)
3	0,25	0,25	H	0,64 (0,02)	0,63 (0,03)	0,63 (0,03)
	0,25	0,25	G	0,45 (0,05)	0,45 (0,05)	0,45 (0,06)
4	0,25	0,13	H	0,62 (0,02)	0,60 (0,02)	0,60 (0,02)
	0,25	0,13	G	0,40 (0,05)	0,40 (0,05)	0,39 (0,05)
5	0,10	0,10	H	0,54 (0,06)	0,50 (0,06)	0,49 (0,06)
	0,10	0,10	G	0,31 (0,09)	0,30 (0,08)	0,29 (0,08)
6	0,10	0,05	H	0,56 (0,05)	0,52 (0,05)	0,52 (0,05)
	0,10	0,05	G	0,34 (0,07)	0,33 (0,06)	0,33 (0,06)

¹it1 a 3=iterações 1 a 3; Característica 1: herdabilidade 0,40 e sem efeito poligênico; Característica 2: herdabilidade 0,40 e com efeito poligênico; Característica 3: herdabilidade 0,25 e sem efeito poligênico; Característica 4: herdabilidade 0,25 e com efeito poligênico; Característica 5: herdabilidade 0,10 e sem efeito poligênico; Característica 6: herdabilidade 0,10 e com efeito poligênico.

Tabela 4 - Correlações entre valores genéticos preditos e verdadeiros em diferentes iterações considerando o rebanho

Característica ¹	h^2	h_{QTL}^2	Matriz de relacionamento	i1	i2	i3
1	0,40	0,40	H	0,87 (0,01)	0,85 (0,02)	0,85 (0,02)
	0,40	0,40	G	0,68 (0,07)	0,67 (0,07)	0,67 (0,07)
2	0,40	0,20	H	0,86 (0,01)	0,85 (0,01)	0,85 (0,01)
	0,40	0,20	G	0,69 (0,05)	0,67 (0,05)	0,66 (0,06)
3	0,25	0,25	H	0,81 (0,01)	0,79 (0,02)	0,79 (0,02)
	0,25	0,25	G	0,61 (0,04)	0,60 (0,04)	0,59 (0,04)
4	0,25	0,13	H	0,83 (0,01)	0,80 (0,01)	0,80 (0,01)
	0,25	0,13	G	0,58 (0,03)	0,56 (0,04)	0,55 (0,03)
5	0,10	0,10	H	0,72 (0,04)	0,66 (0,04)	0,65 (0,04)
	0,10	0,10	G	0,50 (0,05)	0,47 (0,04)	0,46 (0,04)
6	0,10	0,05	H	0,75 (0,02)	0,69 (0,02)	0,67 (0,02)
	0,10	0,05	G	0,50 (0,05)	0,47 (0,05)	0,45 (0,05)

¹ it1 a 3=iterações 1 a 3; Característica 1: herdabilidade 0,40 e sem efeito poligênico; Característica 2: herdabilidade 0,40 e com efeito poligênico; Característica 3: herdabilidade 0,25 e sem efeito poligênico; Característica 4: herdabilidade 0,25 e com efeito poligênico; Característica 5: herdabilidade 0,10 e sem efeito poligênico; Característica 6: herdabilidade 0,10 e com efeito poligênico.

3.3.2 Avaliação da importância das informações de pedigree, coeficiente de herdabilidade e efeito poligênico nas análises de associação genômica ampla via métodos ssGBLUP e GBLUP

Os gráficos das soluções dos efeitos dos marcadores genéticos foram obtidos na primeira iteração para todos os fenótipos (Figuras 2 a 7), pela maior acurácia na predição dos valores genéticos genômicos obtidos nesta iteração para todos os fenótipos e cenários. Os gráficos para cada fenótipo foram semelhantes para os dois cenários, sendo indicativo da mesma capacidade de identificação QTL independente da ausência ou presença das informações de pedigree na análise de GWAS, e assim estas informações não foram determinantes para a detecção de QTL. Além disso, o uso do método GBLUP foi eficiente no estudo da arquitetura genética dos fenótipos estudados, mostrando muitos QTL com pequeno efeito e poucos com grande efeito.

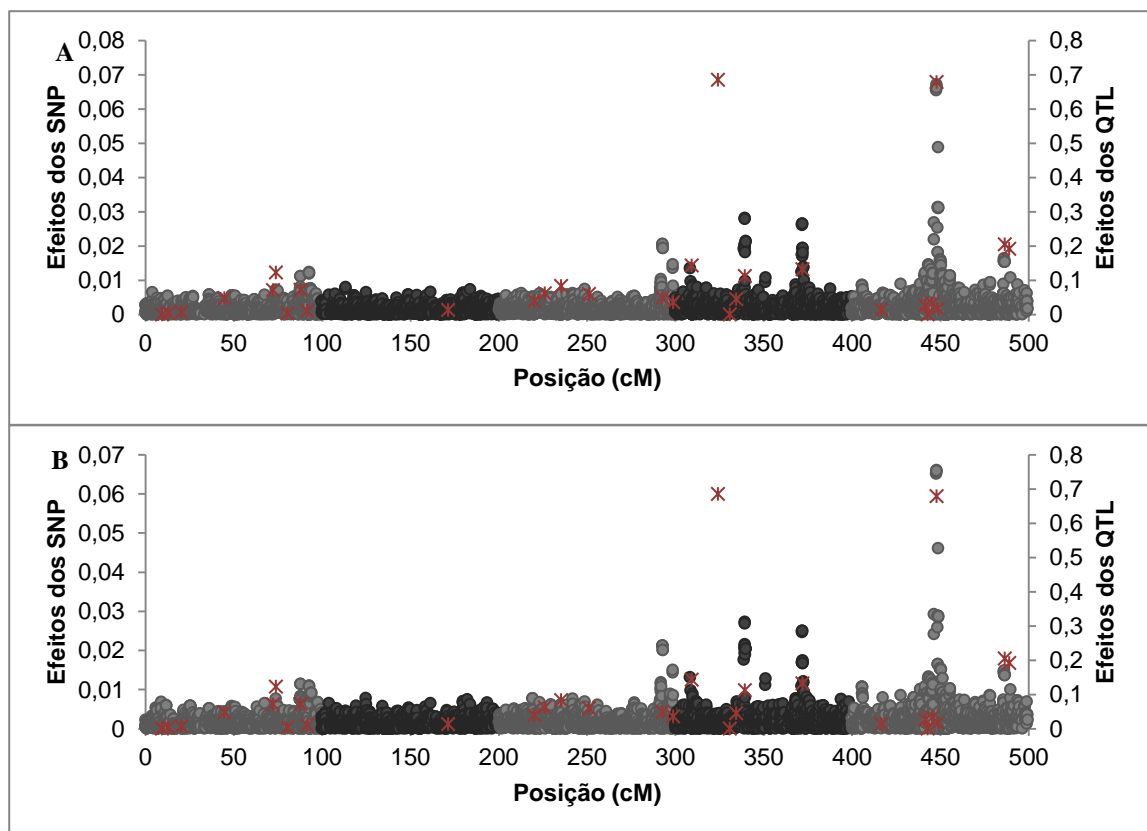


Figura 2 – Soluções dos SNP (● para cromossomos 1, 3, e 5; e ● para cromossomos 2 e 4) e efeitos dos QTL (*) para o fenótipo com herdabilidade 0,40 e sem efeito poligênico, utilizando informações de pedigree (Figura A) ou apenas genótipo e fenótipo (Figura B)

Para a característica com coeficiente de herdabilidade igual a 0,40 e sem efeito poligênico (Figura 2), por meio dos gráficos de soluções para os efeitos dos marcadores genéticos, foi identificada a presença de um QTL no cromossomo cinco, o qual foi confirmado pelos dados simulados, além de vários outros QTL de pequeno efeito presentes na mesma região. Além disso, foram encontrados indícios da presença de QTL no final do cromossomo três e início do cromossomo quatro, na parte central do cromossomo quatro, e na parte final do cromossomo cinco, onde há presença de QTL simulados em posições próximas.

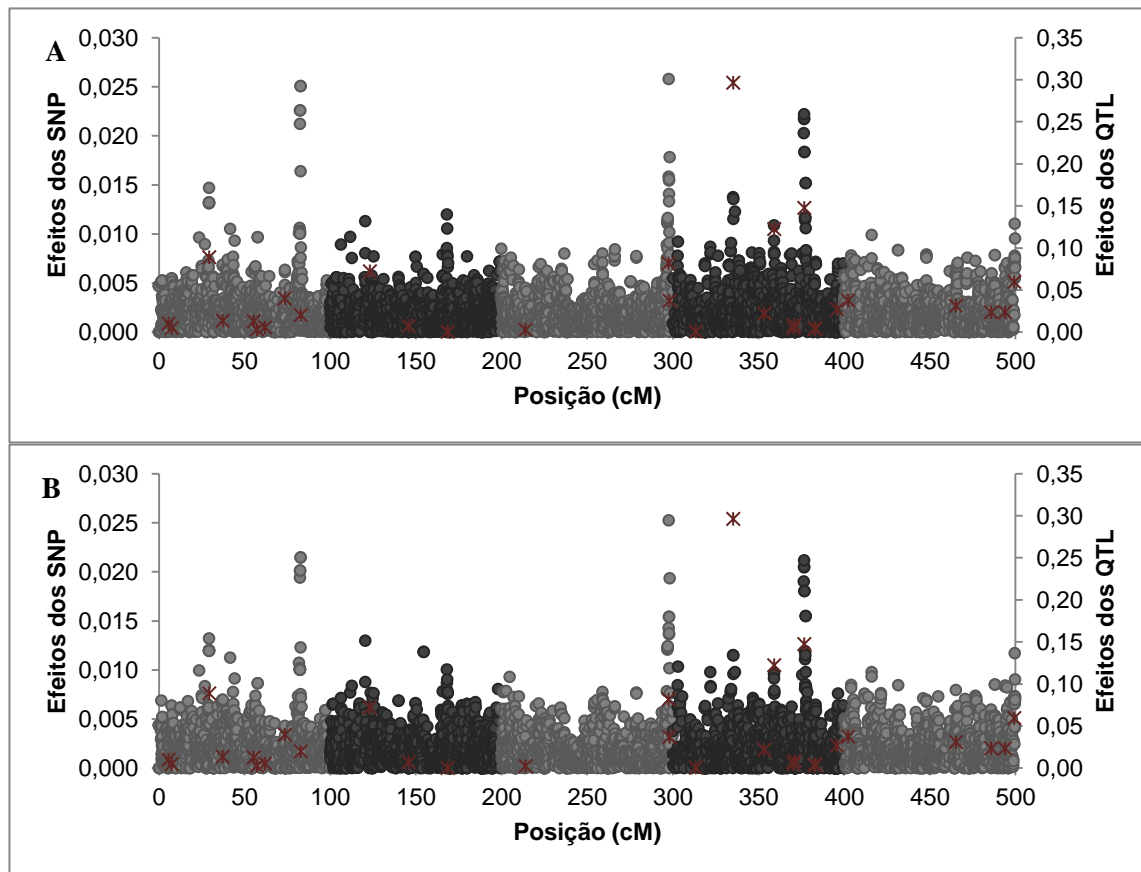


Figura 3 – Soluções dos SNP (● para cromossomos 1, 3, e 5; e ● para cromossomos 2 e 4) e efeitos dos QTL (*) para o fenótipo com herdabilidade 0,40 e com efeito poligênico, utilizando informações de pedigree (Figura A) ou apenas genótipo e fenótipo (Figura B)

Para a característica com herdabilidade 0,40 e com efeito poligênico (Figura 3), a capacidade de identificar QTL por meio dos gráficos das soluções dos efeitos

dos marcadores genéticos foi limitada em relação ao fenótipo sem efeito poligênico e com o mesmo coeficiente de herdabilidade. No cromossomo quatro foram identificados vários QTL, assim como entre o final do cromossomo três e o início do cromossomo quatro, regiões onde foram simulados os QTL de maior efeito para este fenótipo.

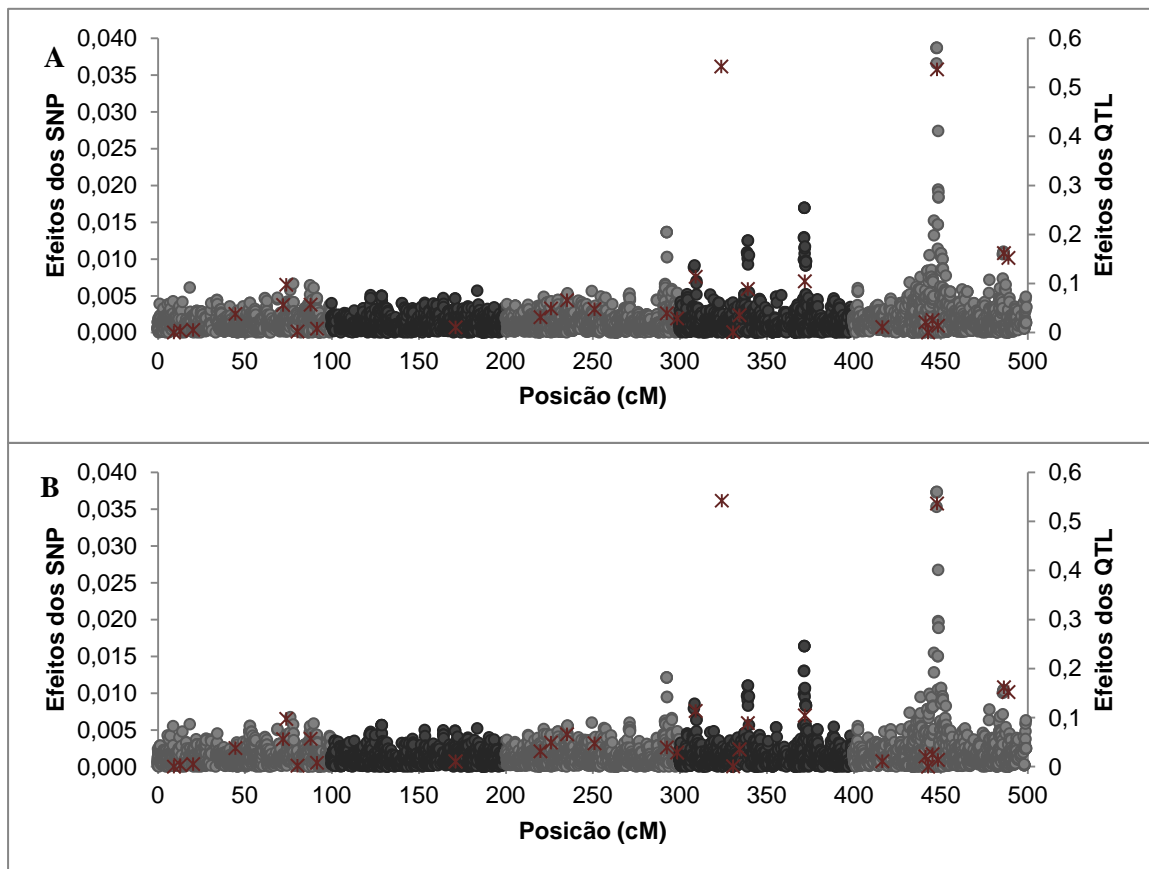


Figura 4 – Soluções dos SNP (● para cromossomos 1, 3, e 5; e ● para cromossomos 2 e 4) e efeitos dos QTL (*) para o fenótipo com herdabilidade 0,25 e sem efeito poligênico, utilizando informações de pedigree (Figura A) ou apenas genótipo e fenótipo (Figura B)

Para a característica com coeficiente de herdabilidade 0,25 e sem efeito poligênico (Figura 4), a eficiência na identificação de QTL foi semelhante à característica com coeficiente de herdabilidade igual a 0,40 e sem efeito poligênico, indicando um QTL de grande efeito na parte central no quinto cromossomo, muitos QTL de pequeno efeito.

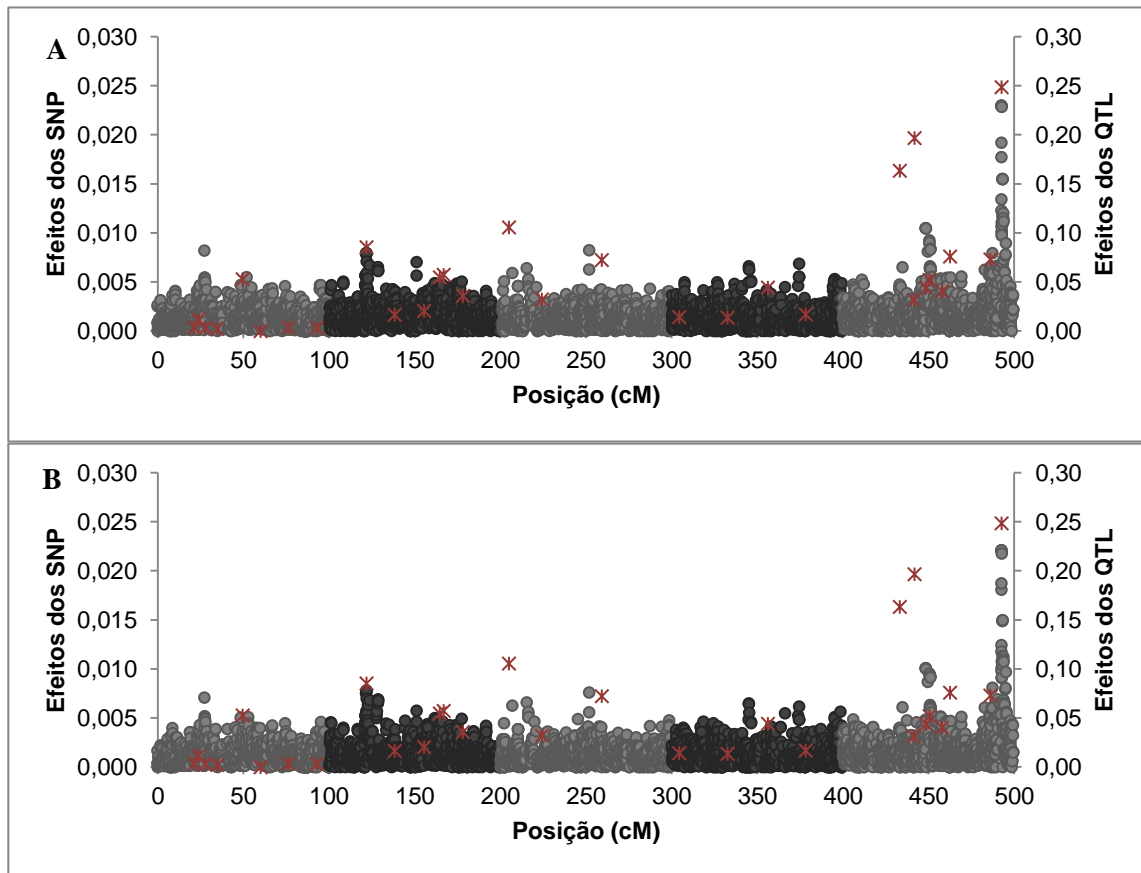


Figura 5 – Soluções dos SNP (● para cromossomos 1, 3, e 5; e ● para cromossomos 2 e 4) e efeitos dos QTL (*) para o fenótipo com herdabilidade 0,25 e com efeito poligênico, utilizando informações de pedigree (Figura A) ou apenas genótipo e fenótipo (Figura B)

Pela análise de associação genômica ampla para o fenótipo com coeficiente de herdabilidade igual a 0,25 e com efeito poligênico (Figura 5), foram identificadas duas regiões com QTL no cromossomo cinco, um central e uma no final do cromossomo, regiões onde foram simulados três QTL de grande efeito. Nas outras regiões não foram identificados marcadores com grande efeito, e assim estas regiões podem ter a presença de QTL de pequeno efeito, conforme os QTL simulados.

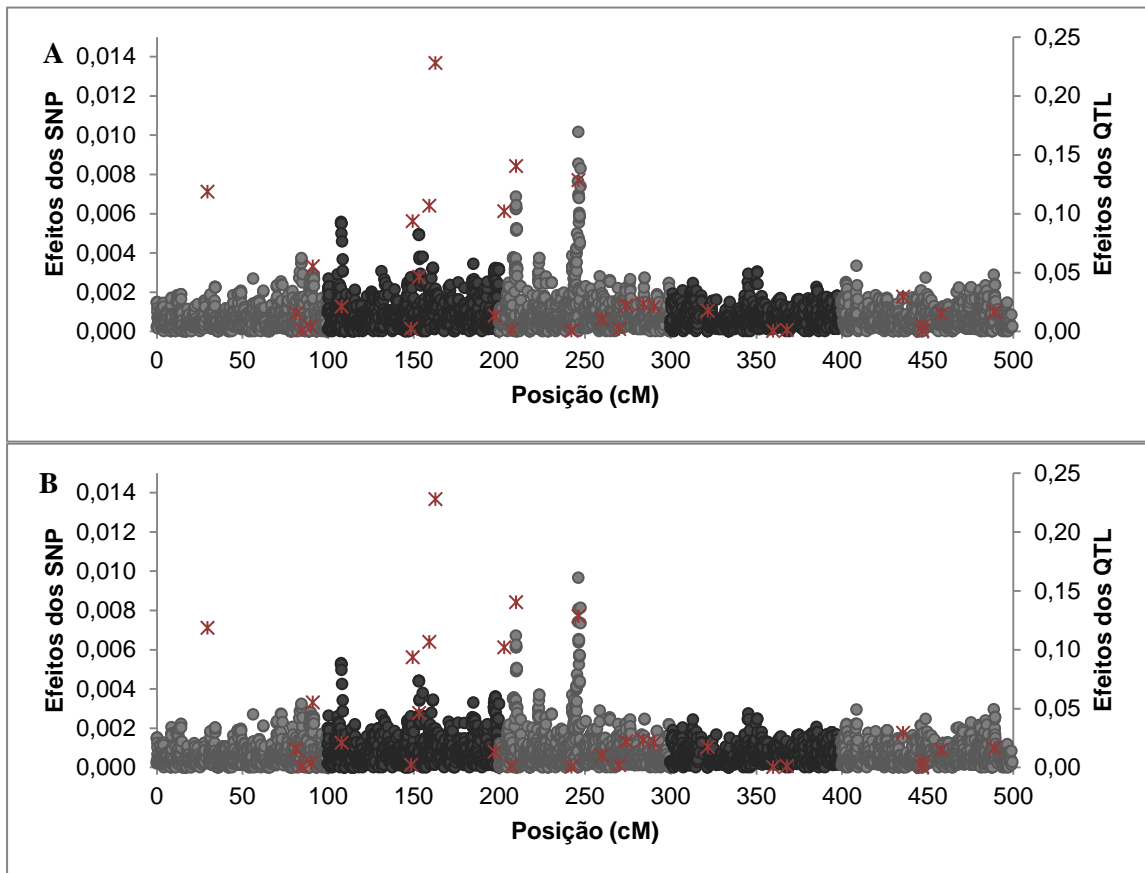


Figura 6– Soluções dos SNP (● para cromossomos 1, 3, e 5; e ● para cromossomos 2 e 4) e efeitos dos QTL (*) para o fenótipo com herdabilidade 0,10 e sem efeito poligênico, utilizando informações de pedigree (Figura A) ou apenas genótipo e fenótipo (Figura B)

Para a característica com herdabilidade 0,10 e sem efeito poligênico (Figura 6), foram identificados QTL nas regiões inicial e intermediária do segundo cromossomo, e nas mesmas regiões do cromossomo três. No primeiro cromossomo foi simulado um QTL que não foi identificado pela análise de associação.

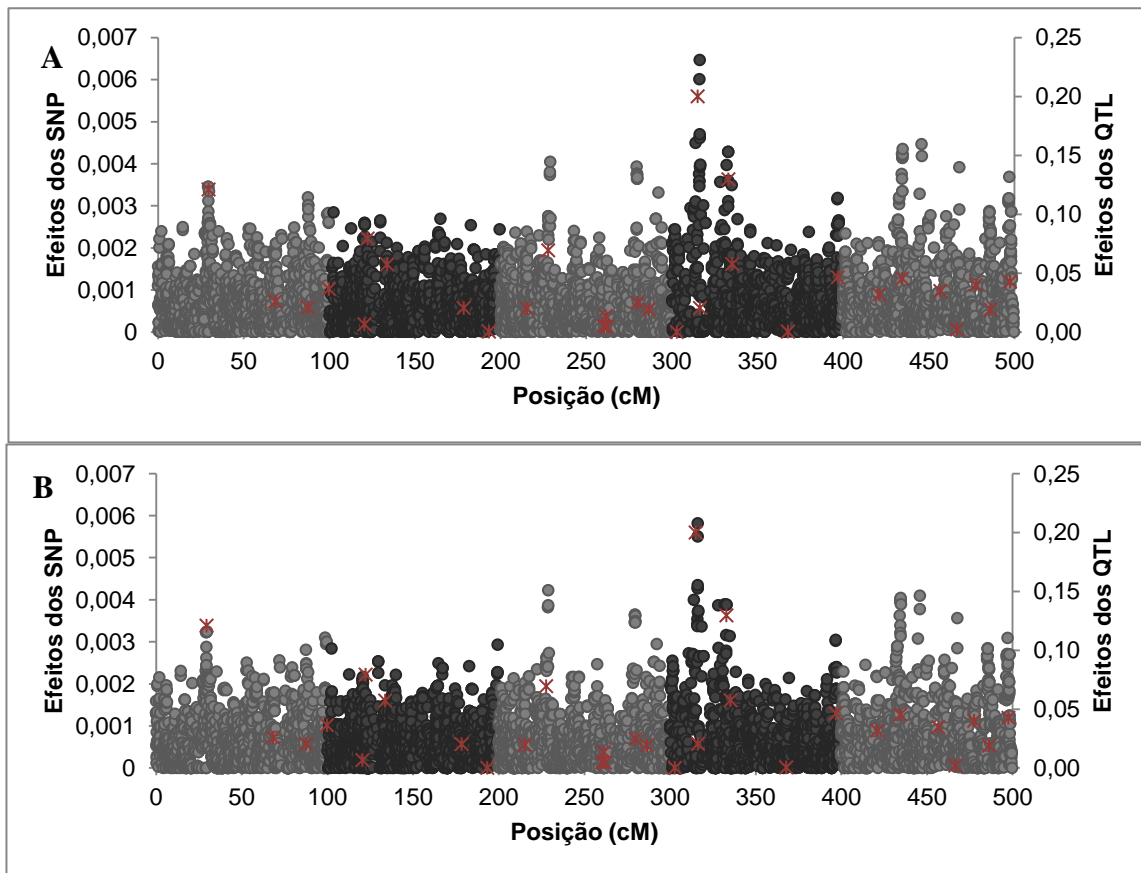


Figura 7 – Soluções dos SNP (● para cromossomos 1, 3, e 5; e ● para cromossomos 2 e 4) e efeitos dos QTL (*) para o fenótipo com herdabilidade 0,10 e com efeito poligênico, utilizando informações de pedigree (Figura A) ou apenas genótipo e fenótipo (Figura B)

Em relação ao fenótipo com coeficiente de herdabilidade 0,10 e com efeito poligênico (Figura 7), houve uma maior dificuldade na identificação de QTL neste caso, sendo identificados apenas um QTL no cromossomo quatro.

Os coeficientes de herdabilidade influenciaram na identificação dos QTL, pois quanto maior a herdabilidade, maiores foram os efeitos estimados para os marcadores genéticos presentes nas regiões próximas ao QTL, e assim com a diminuição do coeficiente de herdabilidade, houve também uma diminuição na capacidade de identificação dos QTLs e no estudo da arquitetura genética da característica, com menores valores nos efeitos dos marcadores genéticos, aumentando a dificuldade de identificação dos QTLs. Quanto menor o coeficiente de herdabilidade, o método ssGBLUP com processo iterativo parece ter maiores benefícios na identificação de QTLs (Figura 8).

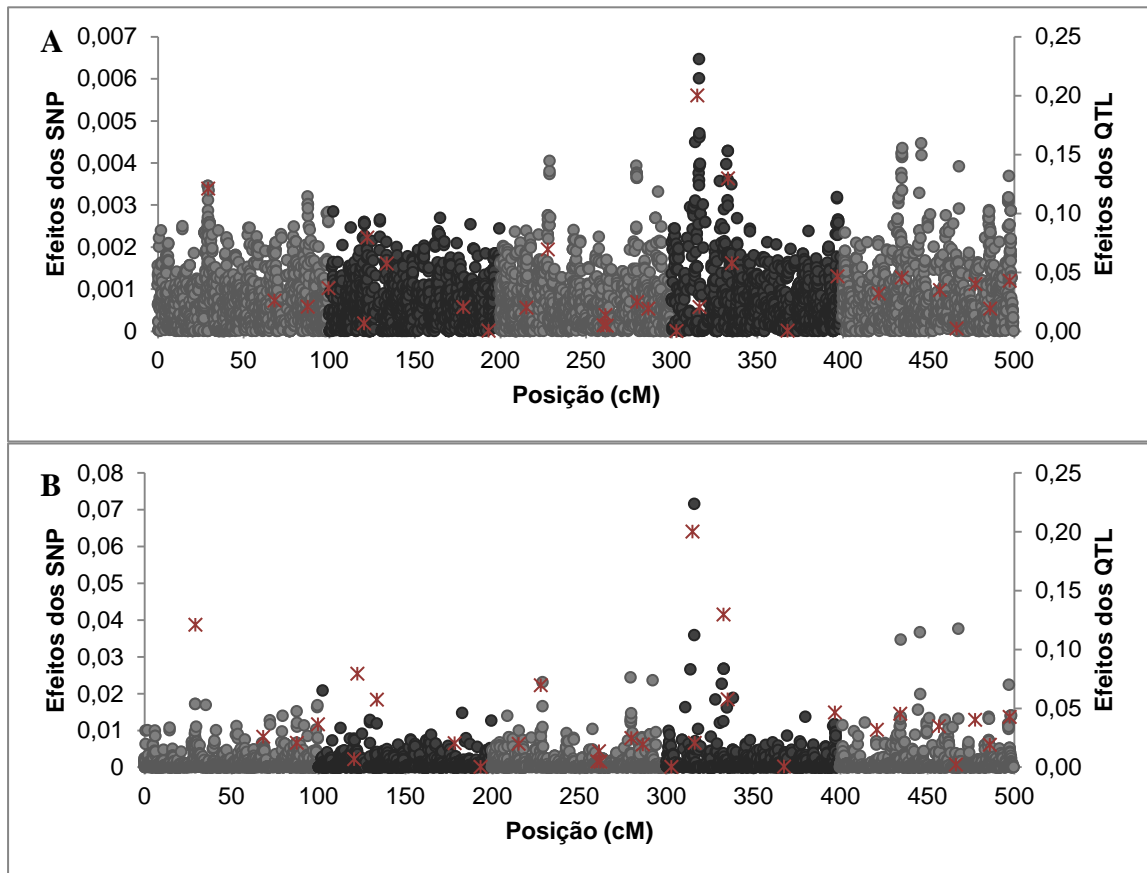


Figura 8 – Soluções dos SNP (● para cromossomos 1, 3, e 5; e ● para cromossomos 2 e 4) e efeitos dos QTL (*) para a primeira (A) e segunda (B) iterações para o fenótipo 4 com informações de pedigree,

3.3.3 Seleção genômica ampla via método GBLUP e abordagem bayesiana para fenótipos com diferentes coeficiente de herdabilidade e efeito poligênico

Na tabela 5 são apresentadas as correlações entre valores genéticos preditos e verdadeiros considerando abordagens clássica (método GBLUP) e bayesiana.

Tabela 5 - Correlações entre valores genéticos preditos e verdadeiros em diferentes métodos de predição considerando o grupo de validação e matriz de relacionamento G

Característica	h^2	h^2_{QTL}	Método	Acurácia (DP)
1	0,40	0,40	GBLUP	0,54 (0,08)
			BayesA	0,52(0,07)
			BayesB ($\pi = 0,7$)	0,52(0,07)
			BayesB($\pi = 0,5$)	0,52(0,07)
			BayesB ($\pi = 0,3$)	0,51(0,07)
			BayesC π	0,52(0,07)
			BLASSO	0,53(0,07)
2	0,40	0,20	GBLUP	0,48 (0,09)
			BayesA	0,69 (0,06)
			BayesB ($\pi = 0,7$)	0,69 (0,06)
			BayesB($\pi = 0,5$)	0,69 (0,06)
			BayesB ($\pi = 0,3$)	0,70 (0,06)
			BayesC π	0,68 (0,05)
			BLASSO	0,69 (0,05)
3	0,25	0,25	GBLUP	0,45 (0,05)
			BayesA	0,49 (0,05)
			BayesB ($\pi = 0,7$)	0,50 (0,05)
			BayesB($\pi = 0,5$)	0,50 (0,05)
			BayesB ($\pi = 0,3$)	0,50 (0,04)
			BayesC π	0,50 (0,04)
			BLASSO	0,51 (0,04)
4	0,25	0,13	GBLUP	0,40 (0,05)
			BayesA	0,47 (0,02)
			BayesB ($\pi = 0,7$)	0,48 (0,02)
			BayesB($\pi = 0,5$)	0,48 (0,02)
			BayesB ($\pi = 0,3$)	0,47 (0,02)
			BayesC π	0,49 (0,02)
			BLASSO	0,49 (0,02)
5	0,10	0,10	GBLUP	0,31 (0,09)
			BayesA	0,30 (0,04)
			BayesB ($\pi = 0,7$)	0,31 (0,04)
			BayesB($\pi = 0,5$)	0,31 (0,04)
			BayesB ($\pi = 0,3$)	0,32 (0,04)
			BayesC π	0,33 (0,03)
			BLASSO	0,34 (0,03)
6	0,10	0,05	GBLUP	0,34 (0,07)
			BayesA	0,37 (0,07)
			BayesB ($\pi = 0,7$)	0,37 (0,07)
			BayesB($\pi = 0,5$)	0,38 (0,07)
			BayesB ($\pi = 0,3$)	0,39 (0,07)
			BayesC π	0,40 (0,08)
			BLASSO	0,41 (0,08)

Foram obtidas correlações entre os valores genéticos verdadeiros e preditos maiores ou semelhantes pelos métodos bayesianos. Para os métodos BayesA e BayesB, foi utilizado um método aproximado, resultando em grande redução do tempo de análise e resultados satisfatórios, pelas acurácias obtidas serem próximas aos métodos BLASSO e BayesC π . Dentre os métodos bayesianos, o BLASSO apresentou as maiores acurácias para todos os fenótipos.

Diversos métodos de estimação dos efeitos dos marcadores genéticos são descritos na literatura. Em relação à predição dos efeitos dos marcadores genéticos, os métodos se diferenciam, principalmente, de acordo os critérios: regularização, que considera a presença de multicolineariedade e o grande número de marcadores genéticos estudados, seleção de variáveis e homogeneidade ou heterogeneidade de variância dos marcadores genéticos (RESENDE et al., 2012). Assim, os diferentes métodos se tornam mais ou menos flexíveis quanto à arquitetura genética das características de interesse.

Dentre os métodos mais frequentemente utilizados, no método GBLUP é considerado que todos os marcadores genéticos apresentam a mesma variância, e não é realizada a regularização e seleção de covariáveis. Em relação à abordagem bayesiana, no método BayesC π é considerado que todos os marcadores genéticos apresentam a mesma variância, ao contrário dos métodos BayesA, BayesB e BLASSO, nos quais é considerado que os marcadores genéticos podem apresentar variância diferentes. Além disso, pelos métodos bayesianos é possível realizar a regularização e seleção de covariáveis, em que no método BayesA os efeitos dos marcadores genéticos devem ser maiores que zero, no método BayesB o número de marcadores genéticos com efeito zero é dada por uma proporção conhecida, no método BayesC π o número de marcadores genéticos com efeito zero é dada por uma proporção desconhecida, e no BLASSO a seleção de covariáveis é realizada por meio de um regressão penalizada (HABIER et al., 2011; DE LOS CAMPOS et al., 2009; MEUWISSEN et al., 2001). Assim, os métodos bayesianos são mais flexíveis em relação à arquitetura genética das características, especialmente o Blasso, no qual não é necessário a fixação da proporção de marcadores com efeito igual a zero, como no BayesB, e por ser considerado que os marcadores genéticos podem apresentar variância diferentes. Em relação ao método BayesC π , estudos

têm demonstrado a eficiência deste método na detecção de regiões associadas aos fenótipos de interesse (WANG, C. et al., 2012).

3.3.4 Avaliação da importância de diferentes números de QTL e pressões de seleção

As acurácias obtidas para fenótipos com diferentes números de QTLs e com coeficiente de herdabilidade igual a 0,40 (Tabela 6) foram pouco diferentes entre os fenótipos, tanto para o grupo de validação (0,67 a 0,72), quanto para o rebanho (0,85 a 0,88). Estas acurácias foram altas, indicativo da qualidade da predição genômica.

Em relação às semelhanças entre o grupo de validação e o rebanho dentro de cada fenótipo, as acurácias mais próximas foram obtidas para o fenótipo com maior número de marcadores, mesmo fenótipo em que foram obtidas as maiores acurácias no grupo de validação.

No rebanho, mas maiores acurácias foram obtidas na primeira iteração, enquanto que no grupo de validação não houve diferença entre a primeira e segunda iteração nos fenótipos com maior número de QTL, sendo que no fenótipo com menor número de QTL foi obtida maior acurácia na primeira iteração.

Tabela 6 - Correlações entre valores genéticos preditos e verdadeiros de fenótipos com diferentes números de QTL e em diferentes iterações considerando grupo de validação e o rebanho

	Característica ¹	i1	i2	i3
Grupo de validação	1	0,68 (0,03)	0,67 (0,03)	0,67 (0,03)
	2	0,70 (0,01)	0,70 (0,02)	0,70 (0,02)
	3	0,72 (0,01)	0,72 (0,01)	0,72 (0,01)
Rebanho	1	0,87 (0,01)	0,85 (0,02)	0,85 (0,02)
	2	0,88 (0,01)	0,86 (0,01)	0,86 (0,01)
	3	0,87 (0,01)	0,86 (0,01)	0,86 (0,01)

¹ it1 a 3=iterações 1 a 3; Característica 1: média de 6 QTL por cromossomo; Característica 2: média de 15 QTL por cromossomo; Característica 3: média de 100 QTL por cromossomo.

As acurácias obtidas no grupo de validação estão de acordo com Wang, H. et al. (2012), em que o processo iterativo no método ssGBLUP é mais eficiente quanto

maior o número de QTL associados ao fenótipo, e resultados divergentes podem estar relacionados à diferenças nas populações.

Tabela 7 - Correlações entre valores genéticos preditos e verdadeiros para cenários com diferentes pressões de seleção e em diferentes iterações considerando grupo de validação e o rebanho

	Característica ¹	i1	i2	i3
Grupo de validação	4	0,68 (0,03)	0,67 (0,03)	0,67 (0,03)
	5	0,72 (0,01)	0,72 (0,01)	0,72 (0,01)
	6	0,68 (0,02)	0,65 (0,04)	0,65 (0,04)
Rebanho	4	0,87 (0,01)	0,85 (0,02)	0,85 (0,02)
	5	0,77 (0,01)	0,77 (0,01)	0,77 (0,01)
	6	0,90 (0,02)	0,88 (0,02)	0,88 (0,02)

¹ it1 a 3=iterações 1 a 3; Característica 4: descarte dos animais por valor genético baixo e substituição de 20% dos macho e 5% das fêmeas; Característica 5: descarte dos animais por idade; Característica 6: descarte dos animais por valor genético baixo e substituição de 60% dos macho e 20% das fêmeas.

As acurácias considerando fenótipos com diferentes pressões de seleção (Tabela 7) foram maiores quanto maior a pressão de seleção no rebanho. No grupo de validação as acurácias foram maiores quanto menor a pressão de seleção.

Quanto maior a pressão de seleção, a variância genética aditiva diminui das gerações mais antigas para as mais recentes. Assim, para o cenário com menor pressão de seleção, as correlações entre os valores genéticos preditos e os verdadeiros foram mais próximas do que nos outros cenários, enquanto que esta correlação é menor nos grupos de validação, em comparação aos valores obtidos para o rebanho, nos outros cenários. Além disso, não houve diferença entre as iterações para o cenário com menor pressão de seleção, enquanto que para os outros cenários a maior acurácia foi obtida na primeira iteração, para ambos os grupo de validação e rebanho.

3.4 Conclusões

Para a predição dos valores genéticos genômicos em uma população sob seleção, as informações de parentesco melhoram a predição dos valores genéticos genômicos. Porém, a melhor habilidade preditiva pode ser obtida quando não são consideradas as informações de parentesco.

A pressão de seleção afeta as acurácias e habilidades preditivas, principalmente pelas mudanças na variância genética aditiva ocorrida nestes casos.

O método GBLUP iterativo foi eficiente para análise de associação genômica ampla, evidenciando regiões de QTL associadas às características de interesse econômico. Além disso, a correlação entre genótipo e fenótipo, possível de ser calculada em dados reais, pode ser um parâmetro para a tomada de decisões, quanto às iterações, nessa população.

O coeficiente de herdabilidade influencia na predição dos valores genéticos genômicos, na habilidade preditiva e nas análises de associação genômica ampla. Assim, quanto maior é a herdabilidade de um fenótipo, há melhorias na identificação dos indivíduos a serem utilizados como reprodutores nas próximas gerações, além de favorecer a identificações de regiões associadas a uma característica de interesse.

Em estudos futuros, deve ser avaliada a influência dos tipos de matrizes de relacionamento em fenótipos com diferentes pressões de seleção para seleção genômica e estudos de associação global. Além disso, a importância das informações de pedigree em rebanhos com diferentes pressões de seleção e em diferentes condições ambientais precisa ser estudada, assim como a aplicação de outros métodos de estimação dos efeitos dos marcadores genéticos, como os métodos bayesianos.

Referências

AGUILAR, I.; MISZTAL, I.; JOHNSON, D.L.; LEGARRA, A.; TSURUTA, S.; LAWLOR, T.J. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. **Journal of Dairy Science**, Champaign, v. 93, p.743–752, 2010.

BRITO, F.V.; BRACCINI NETO, J.; SARGOLZAEI, M.; COBUCI, J.A.; SCHENKEL, F.S. Accuracy of genomic selection in simulated populations mimicking the extent of linkage disequilibrium in beef cattle. **BMC Genetics**, London, v. 12, p. 80, 2011.

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. **Genetics**, Austin, v. 182, p.375-385, 2009.

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

HICKEY, J.M.; GORJANC, G. Simulated data for genomic selection and genome-wide association studies using a combination of coalescent and gene drop methods. **Genes, Genomes and Genetics**, Toronto, v. 2, p. 425-427, 2012.

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

MEUWISSEN, T.H.E. Genomic selection: marker assisted selection on genome-wide scale. **Journal of Animal Breeding and Genetics**, Berlim, v. 124, p. 321-322, 2007.

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

NEVES, H.H.R.; CARVALHEIRO, R.; QUEIROZ, S.A. Adequacy of Different Pseudo-Phenotypes for Model Training and Validation of Genomic Predictions in Beef Cattle. In: WORLD CONGRESS OF GENETICS APPLIED TO LIVESTOCK PRODUCTION, 10., 2014, Vancouver. **Anais eletrônicos...**Vancouver: Organizing committee WCGALP, 2014. Disponível em: https://www.asas.org/docs/default-source/wcgalp-posters/755_paper_9793_manuscript_938_0.pdf?sfvrsn=2. Acesso em: 17 maio 2015.

RESENDE, M.D.V.; SILVA, F.F.; LOPES, P.S.; AZEVEDO, C.F. **Seleção Genômica Ampla (GWS) via Modelos Mistos (REML/BLUP), Inferência Bayesiana (MCMC), Regressão Aleatória Multivariada (RRM) e Estatística Espacial**. Viçosa: Universidade Federal de Viçosa, 2012. 291p.

SARGOLZAEI, M.; SCHENKEL, F. S. QMSim: a largescale genome simulator for livestock. **Bioinformatics**, Oxford, v. 25, p. 680-681, 2009.

WANG, C.; MA, P.; ZHANG, Z.; DING, X.; LIU, J.; FU, W.; WENG, Z.;ZHANG, Q. Comparison of five methods for genomic breeding value estimation for the common dataset of the 15th QTL-MAS Workshop. **BMC Proceedings**, London, v. 6, p. 1-5, 2012.

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

4 ASSOCIAÇÃO GENÔMICA AMPLA PARA COMPOSIÇÃO DE GORDURA INTRAMUSCULAR E CARACTERÍSTICAS DE QUALIDADE DA CARNE E CARÇAÇA EM OVINOS DA RAÇA SANTA INÊS

Resumo

As características de qualidade da carne, carcaça e composição da gordura intramuscular estão relacionadas diretamente com o aumento do consumo destes produtos. Porém, para que estas características sejam mensuradas, é necessário que ocorra o abate dos animais, sendo uma vantagem utilizar informações de milhares de marcadores genéticos para a obtenção dos valores genéticos genômicos e identificações de regiões genômicas associadas às características de interesse econômico. O objetivo deste trabalho foi identificar regiões genômicas associadas às características de qualidade da carne, de carcaça e da composição de gordura intramuscular em ovinos da raça Santa Inês. O banco de dados utilizado neste estudo foi composto por registros de manejo alimentar, consumo, características de qualidade da carne, de carcaça e composição de gordura intramuscular de 179 machos inteiros da raça Santa Inês. As características estudadas foram: área de olho de lombo, espessura de gordura subcutânea, cor, pH ao abate e após 24 horas de resfriamento das carcaças, perdas por cocção, força de cisalhamento, os ácidos graxos cáprico, palmítico e esteárico, peso de carcaça fria, comprimento de carcaça, conformação, peso de pernil. Os dados de qualidade da carne obtidos de ovinos da raça Santa Inês estão dentro dos padrões descritos para esta raça. Foram identificadas diversas regiões genômicas associadas às características estudadas. Para as características de carcaça, algumas das regiões identificadas estão próximas ao gene da miostatina, no cromossomo 2 dos ovinos, sendo que este gene está relacionado com o crescimento muscular.

Palavras-chave: Ácidos graxos; carne; gene posicional; GWAS

Abstract

Meat quality as well as carcass and composition of the intramuscular fat are directly related to an increased consumption of these products. However, to measure these traits it is necessary the slaughter of animals, which is an advantage to use the information from thousands of genetic markers to obtain the genomic breeding values and to identify genomic regions associated with characteristics of economic interest. The objective of this study was to identify genomic regions associated with meat quality, carcass and intramuscular fat composition traits in Santa Inês sheep breed. The database used in this study was consisted of records of feed management, consumption, meat quality, carcass and intramuscular fat composition traits from 179 Santa Inês males. The traits studied were ribeye area, fat thickness, color, pH slaughter and after 24 cooling hours carcass, cooking losses, shear force, the capric, palmitic and stearic fatty acids, cold carcass weight, carcass length, conformation, leg weight. The meat quality data obtained from the Santa Inês sheep are in accordance with the standards described for this breed. Several genomic regions associated with the studied traits were identified. For carcass traits, some of the identified regions are close to the myostatin gene, on chromosome 2, and this gene is related to muscle growth.

Keywords: Fatty acid; meat; positional gene; GWAS

4.1 Introdução

A ovinocultura é uma atividade agropecuária em expansão no Brasil, e apesar do crescimento da produção de carne ovina nos últimos anos e do consumo *per capita* relativamente pequeno, o Brasil realiza importações de carne ovina para abastecer o mercado consumidor, visto que a oferta interna de carne ainda é insuficiente, indicativo de que há mercado interno para aumento na produção de carne ovina.

Neste contexto, a raça Santa Inês é uma das principais de ovinos do Brasil, a qual é bem adaptada às condições climáticas tropicais, apresentando excelente resistência a fatores como ectoparasitas e calor, sendo ainda capazes de se reproduzir o ano todo. Estas características tornam esta raça de grande importância para superação dos desafios associados ao aumento na produção de carne ovina no Brasil, visando à melhoria genética destes animais, ao invés da simples importação e introdução de material genético de clima temperado. A importância disso está na obtenção de índices produtivos satisfatórios, os quais dependem da utilização de genótipos adaptados submetidos ao manejo alimentar adequado. As boas características de adaptação e reprodução da raça Santa Inês não são acompanhadas de boas características de crescimento, qualidade de carcaça e carne, as quais em geral têm níveis de desempenho menores e por vezes insatisfatórios.

Sendo a ovinocultura uma atividade em expansão no Brasil, o aumento do consumo de carne ovina é o principal desafio a ser seguido, e para isso, a melhoria nas características de qualidade da carne é fundamental. Em relação à necessidade de melhoria na qualidade da carne dos ovinos e ao mercado consumidor, é constante o aumento da busca por alimentos com melhor qualidade, tanto nutricional, pelo interesse crescente dos consumidores sobre o efeito benéfico de determinados alimentos à saúde, quanto à apresentação do produto, pois fatores como a coloração, a maciez, o marmoreio e o sabor, influenciam diretamente o interesse do consumidor.

Na busca por melhorias genéticas dos fenótipos de interesse econômico, informações de milhares de marcadores genéticos têm sido utilizadas para identificar regiões genômicas associadas a estas características (GWAS). Para isso, são utilizadas informações obtidas a partir da genotipagem utilizando chips de marcadores genéticos do tipo SNP (*single nucleotide polymorphism*), os quais auxiliam, por meio das análises de associação, em inferências sobre a complexa arquitetura genética das características de interesse econômico (DAETWYLER et al., 2010).

Apesar da importância da ovinocultura e da raça Santa Inês no cenário nacional, e dos estudos de GWAS, poucos estudos desse tipo têm sido realizados em ovinos da raça Santa Inês, especialmente para características de qualidade da carne, gordura intramuscular e carcaça. Assim, o objetivo neste capítulo foi identificar regiões genômicas associadas às características de qualidade da carne, carcaça e ácidos graxos, por meio do método GBLUP e de dados de 54.000 marcadores do tipo SNP (*single nucleotide polimorfismo*).

4.2 Material e métodos

4.2.1 Conjunto de dados

O banco de dados utilizado neste estudo foi composto por registros de manejo alimentar, consumo, características de qualidade da carne e de carcaça de 179 ovinos machos da raça Santa Inês. Os animais foram criados na fazenda de São Gonçalo dos Campos, localizada no município de São Gonçalo dos Campos, Bahia, e pertencente à Universidade Federal da Bahia, UFBA.

Os dados foram coletados de animais confinados do desmame (110 dias em média) ao abate, em baias individuais de 1 m², com alimentação e água fornecidos *ad libitum*. Os animais foram submetidos a diferentes dietas, sendo que todos receberam um concentrado padrão e sal mineral, para balancear a alimentação conforme às exigências nutricionais do grupo de animais avaliados. No primeiro lote, composto por 100 animais, as dietas foram fornecidas duas vezes por dia, e continham caroço de algodão moído ou inteiro e a presença, em diferentes níveis, ou ausência de quitosana.

No segundo lote, composto por 80 ovinos, os ovinos foram divididos em dois grupos experimentais. Metade dos animais receberam dietas com silagem de sorgo com diferentes aditivos, sendo que esta dieta foi fornecida duas vezes por dia. Os outros 40 ovinos receberam dieta padrão de feno e concentrado, sendo que os diferentes manejos alimentares foram definidos pelo número de vezes que as dietas dos animais foram particionadas, ou seja, os animais recebiam a mesma dieta em uma, duas, três ou quatro vezes ao dia.

Antes e durante o abate, foram coletados o peso ao abate, peso da carcaça quente e pH do músculo *Longissimus dorsi*. Após isso, as carcaças foram resfriadas por 24 horas, seguido de mensurações do peso de carcaça fria, pH, comprimento e conformação de carcaça e peso dos cortes comerciais.

As características de qualidade da carne avaliadas foram avaliadas no músculo *Longissimus dorsi*, na região da 12^a e 13^a vertebrae torácicas no lado esquerdo da carcaça, sendo:

- Área de olho de lombo (cm²): foi mensurada com auxílio de papel vegetal e posteriormente analisada com aparelho integrador de área foliar (Li-cor, modelo 3100). O valor final para cada animal foi obtido a partir da média de seis mensurações, as quais foram corrigidas de acordo com um padrão;
- Espessura de gordura subcutânea (mm): foi mensurada com auxílio do paquímetro, a partir da média de três mensurações;
- Cor: a mensuração da coloração da carne foi realizada 24 horas após o abate, utilizando o sistema CIELab, em que foram obtidos valores médios, tomados em três pontos, de L* (teor de luminosidade), a*(teor de vermelho) e b* (teor de amarelo) , realizados como auxílio de um colorímetro (Marca Minolta);
- Perda por cocção e força de cisalhamento: as amostras foram assadas em forno a 320°C até atingirem a temperatura interna de 71°C, medida com termômetro inserido no centro geométrico de cada bife. A medida de perda por cocção foi obtida por diferença entre peso inicial e peso final. Para a mensuração da força de cisalhamento, após o resfriamento dos bifes, foram retiradas três amostras cilíndricas medindo 1,27 cm diâmetro por 2,54 cm de comprimento, as quais foram extraídas paralelamente ao longo do eixo das fibras musculares (WHEELER et al.,1997), e o cisalhamento foi realizado, por

um texturômetro, perpendicularmente à orientação longitudinal das fibras musculares.

- Perfil de ácidos graxos: as extrações de lipídeos e cromatografias foram realizadas no Laboratório de Nutrição e Crescimento Animal da ESAQL/USP. Foram utilizadas amostras de aproximadamente 7 g do músculo *Longissimus dorsi*, sendo os lipídeos extraídos conforme Hara e Radin (1978), por meio de solução de hexano/propanol na proporção 3:2. Após a extração dos lipídeos, as amostras passaram pelo processo de metilação. Os ésteres metílicos de ácidos graxos foram quantificados por cromatografia gasosa, em um equipamento com um detector de ionização de chama e coluna (Supelco) capilar de sílica fundida de 100 m. Os ácidos graxos foram mensurados pelo tempo de retenção destes em relação aos padrões utilizados.

As estatísticas descritivas de número de animais com informações, média, desvio-padrão, valores máximo e mínimo foram obtidos por meio do procedimento UNIVARIATE do SAS[®] para cada característica.

4.2.2 Extração de DNA e genotipagem

Para extração do DNA, foram coletadas amostras de tecido muscular, após o abate e resfriamento das carcaças, e armazenados em freezer -20°C. A extração do DNA foi realizada utilizando proteinase K. Após essa etapa, a concentração e pureza do DNA foi realizada no equipamento NanoDrop[®]. Após a quantificação, foi realizada a diluição adequada do DNA, de acordo com o protocolo da genotipagem, visando a posterior hibridização na matriz do chip Ovine SNP50 BeadChip (Illumina Inc., San Diego, CA). Além da quantificação do DNA, a integridade das amostras de DNA foi analisada por meio de eletroforese em gel de agarose 1%, utilizando um corante GelRedTM para análise do DNA genômico sob iluminação UV. Após a diluição do DNA, e mensurações da quantidade e qualidade deste, as amostras foram submetidas ao procedimento de genotipagem. As extrações de DNA e genotipagem foram realizadas no Laboratório de Biotecnologia da Escola Superior de Agricultura “Luiz de Queiroz” - ESALQ/USP.

4.2.3 Controle de qualidade e análise de associação genômica ampla

As análises de associação genômica ampla foram realizadas no programa Blupf90. O controle de qualidade de amostra e indivíduos foi realizado no mesmo programa, excluindo SNPs e indivíduos com *call rate* menor que 90%, e SNPs com MAF menor que 0,05. O método utilizado foi o GBLUP, e assim a matriz de relacionamento considerada nas análises foi a matriz genômica ou realizada. O modelo considerado foi:

$$y = X\beta + Zu + e,$$

em que: y é o vetor de fenótipos simulados, X é a matriz que relaciona os efeitos fixos aos fenótipos, β é o vetor de efeitos fixos (grupo de contemporâneos formado por manejo alimentar e lote do animal), u é o vetor de efeitos individuais dos animais, Z é a matriz de incidência que relaciona os indivíduos aos fenótipos e e é o vetor de resíduos.

A partir do método GBLUP, os efeitos dos marcadores foram obtidos a partir da equação:

$$\hat{a} = (Z'Z)^{-1}(Z'\hat{u}),$$

em que: \hat{a} é o vetor de efeitos preditos dos marcadores genéticos, Z é a matriz de incidência que relaciona os indivíduos aos fenótipos e u é o vetor de valores genéticos dos animais.

Os gráfico de soluções dos efeitos dos SNPs (Manhattan plot) foram obtidos considerando 8 SNPs adjacentes. Foi realizada uma comparação das regiões identificadas com as informações de QTL disponíveis no banco de dados Animal QTLdb (HU et al., 2013) para ovinos. Além disso, pela posição dos 8 SNPs adjacentes que explicaram a maior parte da variância genética para cada característica, foi possível identificar se estes SNPs se encontram em regiões de genes descritos na literatura, o que foi realizado por meio da ferramenta *Map Viewer* (NCBI).

4.3 Resultados e discussão

As estatísticas descritivas de número de animais mensurados, média, desvio-padrão, mínimo e máximo para as características de qualidade da carne estão na Tabela 8.

Tabela 8 – Estatísticas descritivas para características de qualidade da carne e carcaça.

Característica ¹	N	Média	DP	Mínimo	Máximo
AOL	99	12,42	1,631	6,79	17,10
EGS	98	0,16	0,073	0,1	0,5
pH0	100	6,64	0,138	6,29	7,05
pH24	100	5,72	0,167	5,06	6,09
PPC	98	15,46	3,891	3	29
FC	93	2,39	0,645	0,81	4,79
L*	100	36,68	2,940	30,74	43,72
a*	100	21,32	1,836	16,34	25,31
b*	100	5,40	1,762	2,42	9,77
C10:0	94	0,16	0,030	0,117	0,268
C16:0	94	22,11	1,410	17,82	25,71
C18:0	94	22,23	2,932	17,14	33,26
CC	178	55,88	3,867	41	66
PPER	178	2,56	0,356	1,37	3,90
PA	100	39,63	5,125	20,8	52,0
PCF	178	17,02	2,457	8,50	25,40
CONF	178	2,62	0,429	1,75	3,75

¹N=número de animais avaliados; DP=Desvio padrão; pH0=pH mensurado após o abate dos animais; pH24=pH mensurado após 24 horas de resfriamento da carcaça; EGS=espessura de gordura subcutânea; AOL=área de olho de lombo; PPC=perdas por cocção; FC=força de cisalhamento; L*=luminosidade; a*=teor de vermelho; b*=teor de amarelo; C10:0=ácido cáprico, C16:0=ácido palmítico; C18:0=ácido esteárico; CC=comprimento de carcaça; PPER=peso de pernas; PA=peso ao abate; PCF=peso de carcaça fria; CONF=conformação.

O tamanho da área de olho de lombo está diretamente associado à quantidade de carne na carcaça e com a relação músculo:osso dos cortes comerciais (CEZAR; SOUZA, 2007). Os valores de média, mínimo e máximo, obtidos para área de olho de lombo foram: 12,42, 6,79 e 17,10, respectivamente, valores semelhantes aos resultados encontrados na literatura, com diferenças no manejo alimentar e idade ao abate dos animais. Para a mesma característica mensurada na carcaça, Jucá (2013) obteve os valores 9,06, 3,88 e 16,5, para média, mínimo e máximo, respectivamente, em cordeiros Santa Inês criados em sistema semi-intensivo e abatidos com 424,19 dias de idade em média. Alves et al. 2014, quando avaliados cordeiros inteiros mestiços da raça Santa Inês, que foram submetidos a dietas com ureia de liberação lenta em substituição à ureia convencional, obtiveram medias para área de olho de lombo de 12 a 14, dependendo do tipo de ureia que os animais receberam.

A espessura de gordura subcutânea da carcaça está relacionada com a qualidade de carne, pois durante o resfriamento e transporte das carcaças, a falta de cobertura na carcaça ocasiona em perda de água da carne e no encurtamento das fibras musculares pelo frio, sendo que este segundo fator causa diminuição da maciez da carne (CEZAR;SOUSA, 2010; LAWRIE, 2005; MÜLLER, 1987). Além disso, a quantidade de gordura subcutânea, inter e intramuscular esta diretamente relacionada com a aceitabilidade da carne de cordeiros (SAÑUDO et al., 2000). Para espessura de gordura subcutânea foi obtida média de 0,16, sendo que os valores variaram de 0,1 a 0,5, semelhantes aos resultados obtidos por Alves et al. (2014), no qual a gordura subcutânea variou de 2,48 a 4,25, e por Jucá (2013), que obteve valores para média, mínimo e máximo de 0,10, 0,18 a 0,30, respectivamente.

Para obtenção de boas características de qualidade da carne, como a capacidade de retenção de água, as perdas por cocção e a força de cisalhamento, torna-se importante garantir que o processo de transformação de músculo em carne ocorra, sendo que o pH é uma característica indicativa da eficiência deste processo (BOUTON et al., 1971). Os valores obtidos para o pH inicial e final foram 6,64 e 5,72, respectivamente, valores semelhantes aos obtidos por Oliveira et al. (2004), num estudo com 12 machos inteiros da raça Santa Inês e no qual foram obtidos os valores 6,67 e 5,61 para o pH inicial e final, respectivamente, e também foram semelhantes aos encontrados por Jucá (2013), de 6,63 e 5,46, respectivamente.

A característica perda por cocção está relacionada com o preparo da carne para o consumo. Foram obtidos valores para média, mínimo e máximo de 15,46, 3 e 29, respectivamente, semelhantes aos obtidos por Jucá (2013), com média de 13,46, e menores que outros resultados obtidos por Costa et al. (2011), Vieira et al. (2010) e Bressan et al. (2001), sendo que as diferenças podem estar relacionadas com as metodologias utilizadas.

A maciez da carne está relacionada diretamente com a apreciação do consumidor. Quanto maior a maciez da carne, menor é a medida de força de cisalhamento. A média obtida neste estudo foi 2,39, semelhante ao obtido por Jucá (2013), de 2,27, por Bressan et al. (2001), de 2,5, e menor que os resultados obtidos por Alves et al., 2014, os quais encontraram valores para maciez que variaram 3,25 a 4,00, e por Vieira et al. (2010), que obtiveram valores para força de cisalhamento de 4,6 a 5,3 kgf. Além disso, os resultados obtidos no presente estudo foram maiores que os obtidos por Costa et al. (2011) de 1,45 kgf, sendo que estes

resultados podem estar relacionados com diferenças na metodologia, no manejo alimentar e na idade de abate dos animais.

Outra importante característica de qualidade da carne é a cor, a qual se destaca como principal fator de apreciação consumidor no momento da compra. Dentre os fatores relacionados com a coloração da carne, está o principal pigmento responsável por esta, conhecido como mioglobina, o qual se trata de uma proteína envolvida nos processos de oxigenação do músculo (RENERRE, 1990), e assim, alterações na coloração da carne podem ser indicativos dos processos oxidativos desta. Foram obtidas médias para L^* , a^* e b^* de 36,68, 21,32 e 5,40, respectivamente. Estes valores foram diferentes dos obtidos por Jucá (2013), de 41,76, 16,40 e 6,96, para L^* , a^* e b^* , respectivamente, e assim no presente estudo as carnes apresentaram maior teor de vermelho e menor luminosidade, provavelmente devido às diferenças nas dietas e idade dos animais.

Nas Tabelas 9, 10 e 11 são apresentadas as regiões, em pares de base, associadas às características de qualidade da carne, composição de gordura e carcaça, respectivamente. Além disso, estão descritos nestas tabelas os genes que se encontram nas regiões associadas às características de interesse.

Tabela 9 – Regiões genômicas associadas às características de qualidade da carne

Característica ¹	Cromossomo	Posição	% variância explicada	Gene
AOL	16	33.184.393	0,32	-
	16	33.763.548	0,31	-
	16	33.224.073	0,30	-
EGS	3	21.421.392	0,28	-
	3	21.360.054	0,25	-
	18	58.881.079	0,24	-
PPC	1	164.476.394	0,23	RPL24
	1	164.454.098	0,23	ZBTB11
	1	198.449.920	0,23	ST6GAL1
FC	18	5.508.053	0,30	ADAMTS17
	18	5.455.260	0,29	ADAMTS17
	1	83.548.701	0,26	-
	8	26.576.567	0,24	-
L*	9	97.310.214	0,31	-
	2	130.521.776	0,25	PLEKHA3
	2	130.583.897	0,25	-
	2	130.478.399	0,25	LOC101123351
a*	18	70.382.783	0,30	-
	18	70.320.603	0,29	-
	18	70.402.845	0,27	-
b*	1	125.275.563	0,40	-
	1	37.820.087	0,38	ATG4C
	1	37.557.999	0,36	DOCK7
	1	125.243.619	0,35	-

¹AOL=área de olho de lombo; EGS=espessura de gordura subcutânea; PPC=perdas por cocção; FC=força de cisalhamento; L*=luminosidade; a*=teor de vermelho; b*=teor de amarelo.

Tabela 10 – Regiões genômicas associadas às características de composição de gordura

Característica ¹	Cromossomo	Posição	% variância explicada	Gene
C10:0	14	31.736.852	0,28	-
	14	31.893.971	0,24	-
	10	29.092.520	0,24	FRY
C16:0	2	118.779.167	0,32	-
	2	118.498.709	0,29	ANKAR
	2	118.718.924	0,29	-
C18:0	1	3.577.796	0,30	COL6A3
	1	3.638.674	0,26	COL6A3
	1	25.834.328	0,25	OSBPL9

¹C10:0=ácido cáprico, C16:0=ácido palmítico; C18:0=ácido esteárico.

Tabela 11 – Regiões genômicas associadas às características de carcaça

Característica	Cromossomo	Posição	% variância explicada	Gene
CC	10	55.148.103	0,37	-
	9	24.823.969	0,32	-
	1	74.015.424	0,32	DPYD
PPER	2	118.779.167	0,28	-
	2	118.492.152	0,27	ANKAR
	2	118.498.709	0,26	ANKAR
PA	1	83.548.701	0,28	-
	1	47.719.852	0,23	NEGR1
	1	52.347.468	0,22	ST6GALNAC3
PCF	2	118.492.152	0,34	ANKAR
	2	118.458.759	0,32	OSGEPL1
	2	118.439.018	0,32	ORMDL1
	2	118.779.167	0,31	-
	2	118.498.709	0,31	ANKAR
	2	118.718.924	0,29	-
CONF	8	32.102.849	0,26	LIN28B
	9	91.576.333	0,26	-
	8	32.114.977	0,25	LIN28B
	8	32.071.280	0,23	LIN28B
	14	58.948.271	0,22	-

¹CC=comprimento de carcaça; PPER=peso de pernas; PA=peso ao abate; PCF=peso de carcaça fria; CONF=conformação.

No cromossomo 2 dos ovinos, próximo a região de 118 Mpb, é encontrado o gene da miostatina (MSTN), que é conhecido como fator de crescimento e diferenciação 8 (GDF8), e está relacionado com o crescimento muscular (McPHERRON; LAWLER; LEE, 1997).

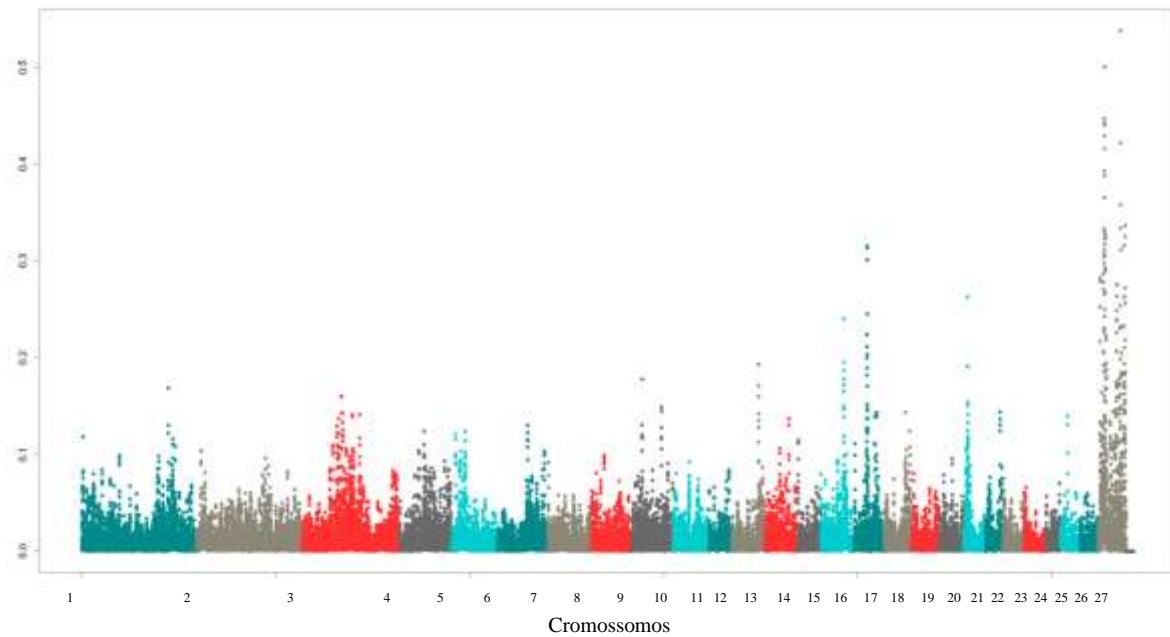


Figura 9 – Soluções dos SNP para a característica área de olho de lombo, considerando a variância explicada por 8 SNPs adjacentes

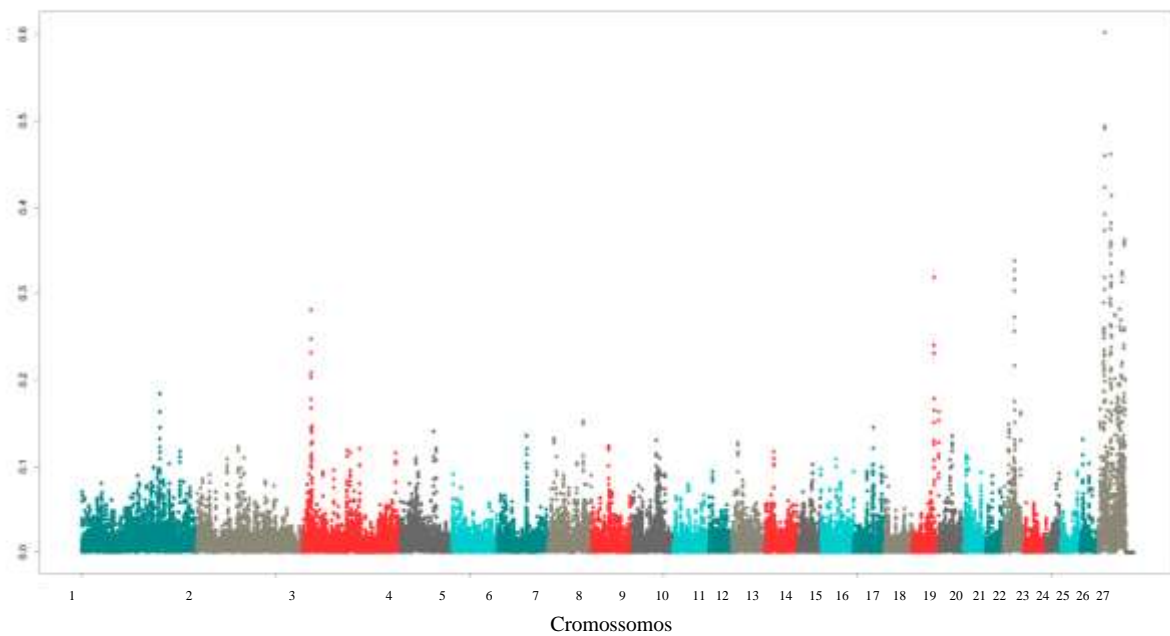


Figura 10 – Soluções dos SNP para a característica espessura de gordura subcutânea, considerando a variância explicada por 8 SNPs adjacentes

Para a característica área de olho de lombo (Figura 9), foram identificadas regiões nos cromossomos 15, 16 e 20. No banco de dados de QTL de ovinos (QTLdb), são descritos QTLs identificados nos cromossomo 2, 7, 9 e 18 para esta característica.

Foram identificadas regiões associadas a característica de espessura de gordura subcutânea (Figura 10) nos cromossomos 3, 18 e 22, sendo que no banco de dados QTLdb estão descritas regiões nos cromossomos 2 e 16 que estão associadas com esta característica. Para a característica perda por cocção, os maiores efeitos de SNPs foram encontrados no cromossomo 1, enquanto que para a força de cisalhamento, os maiores efeitos foram encontrados no cromossomo 18. Para a característica força de cisalhamento, são encontrados QTL descritos no QTLdb nos cromossomos 2 e 18.

Para a característica de luminosidade da carne, foram identificadas regiões nos cromossomos 2 e 9, sendo que no banco de dados de QTL (QTLdb) estão descritos QTL no cromossomo 2 para a mesma característica. Para as características teor de vermelho e teor de amarelo foram identificadas regiões associadas a estas características nos cromossomos 19, para ambas as características, e no primeiro cromossomo para o teor de amarelo. No QTLdb estão descritos QTL presentes no cromossomo 2 para teor de vermelho e de amarelo.

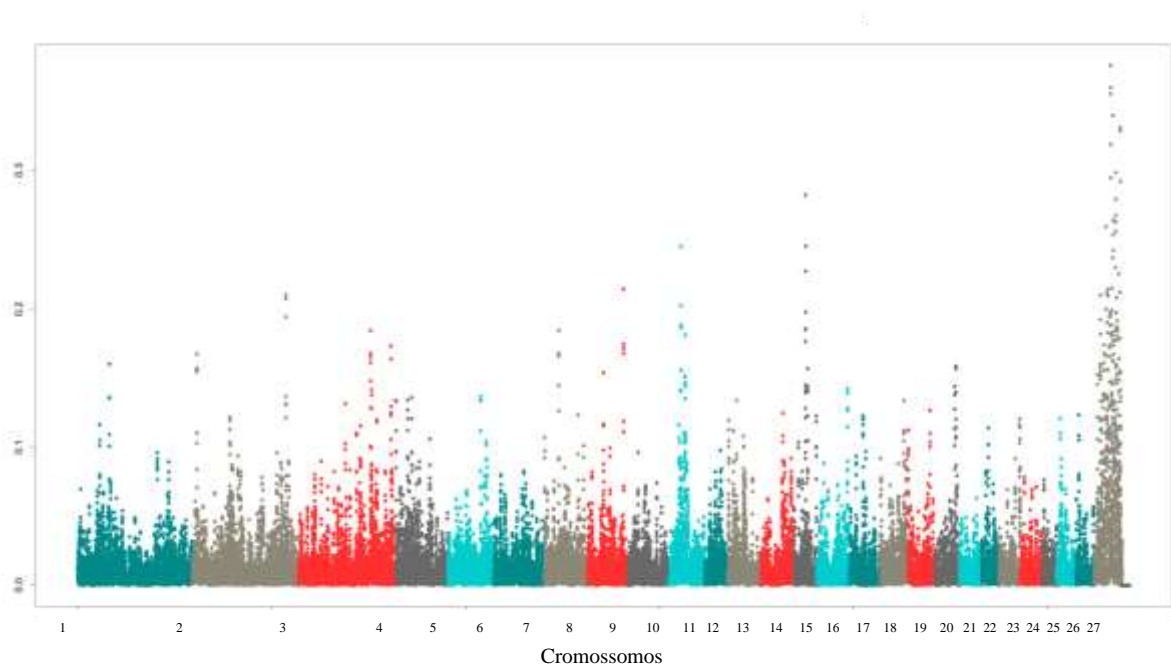


Figura 11 – Soluções dos SNP para a característica ácido cáprico (C10:0), considerando a variância explicada por 8 SNPs adjacentes

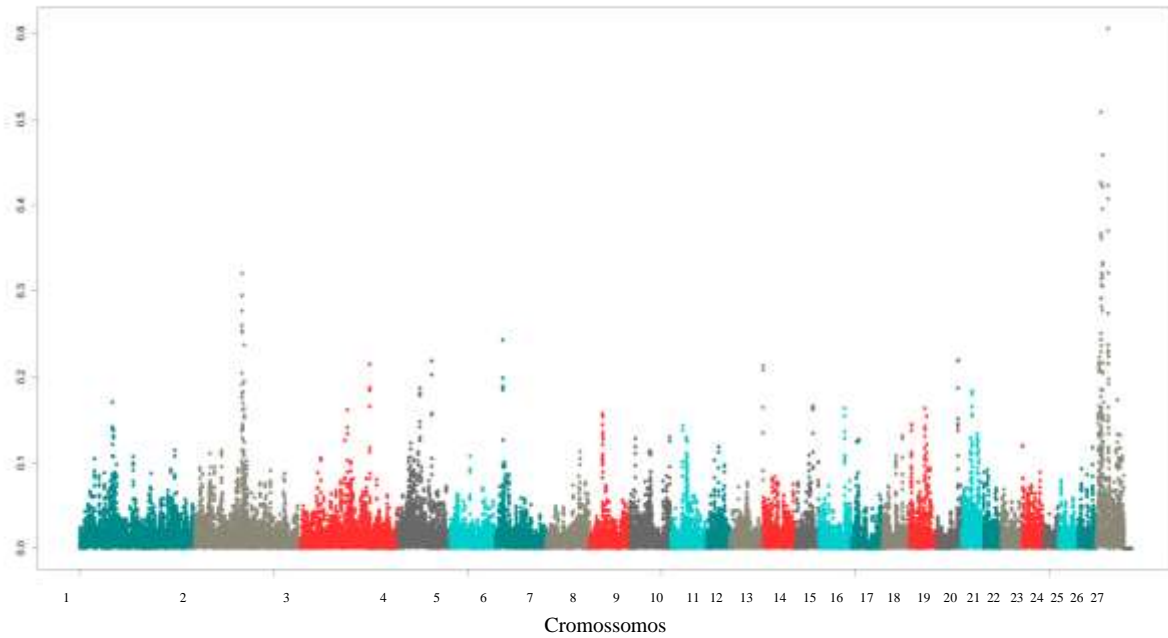


Figura 12 – Soluções dos SNP para a característica ácido palmítico (C16:0), considerando a variância explicada por 8 SNPs adjacentes

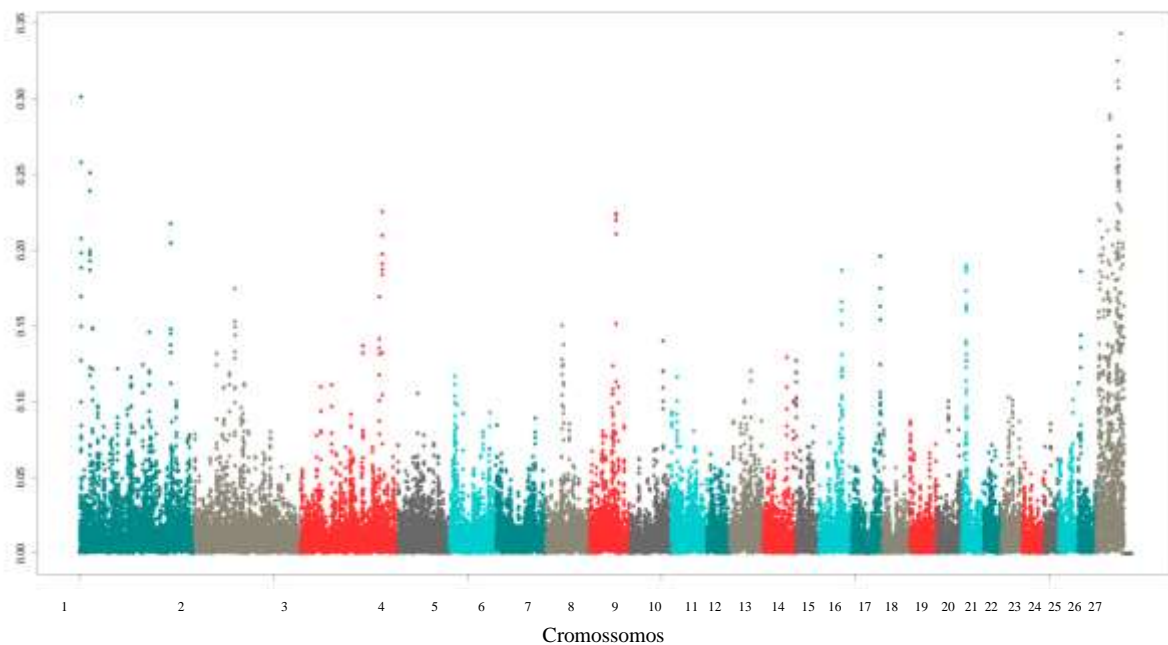


Figura 13 – Soluções dos SNP para a característica ácido esteárico (C18:0), considerando a variância explicada por 8 SNPs adjacentes

Os gráficos de soluções dos efeitos dos SNPs para os ácidos graxos cáprico (Figura 11), palmítico (12) e esteárico (13), são indicativos da presença de QTL nos cromossomos 14, 2 e 1 para cada uma das características, respectivamente. No banco de dados QTLdb são descritos QTL no cromossomo 21 para os ácido graxos palmítico e esteárico.

Para a característica peso de carcaça fria, foram identificadas regiões associadas a esta característica nos cromossomos 2 e 9, e no banco de dados QTLdb regiões nos cromossomos 2 e 9 associados a característica peso de carcaça quente.

Para a característica peso ao abate, foi identificado uma região no cromossomo 1 associada a este fenótipo. No banco de dados QTLdb são descritos vários QTL para características de peso corporal presentes no cromossomo 1, especialmente para pesos mensurados nas semanas 20 e 43 de vida dos animais, semelhante ao presente estudo no qual o peso ao abate foi mensurado próximo da trigésima semana de vida dos cordeiros. Para peso do pernil, tanto pela análise de associação, quanto pelo banco de dados QTLdb, foram identificados QTL no segundo cromossomo. Para comprimento de carcaça, foram identificadas regiões nos cromossomos 1, 9 e 10 que explicam mais da variância genética para esta característica do que as outras regiões, sendo que na literatura (QTLdb), são descritos QTL nos cromossomos 1 e 2 de ovinos para a mesma característica. Para a característica de conformação, foram identificadas possíveis regiões associadas a este fenótipo nos cromossomos 8 e 9.

4.4 Conclusões

Os dados de qualidade da carne de ovinos da raça Santa Inês estão dentro dos padrões para machos inteiros desta raça e espécie, validando estes dados para que sejam submetidos a análises de associação genômica ampla. Nas análises de associação genômica ampla, foram identificadas diversas regiões que possivelmente estão associadas às características estudadas. Em estudos futuros, as relações entre os genótipos e os diferentes manejos nutricionais devem ser avaliadas, ou seja, a interação entre alimentação e genótipo.

Referências

- ALVES, E.M.; PEDREIRA, M.S.; MOREIRA, B.S.; FREIRE, L.D.R.; LIMA, T.R.; SANTOS-CRUZ, C.L. Carcass characteristics of sheep fed diets with slow-release urea replacing conventional urea. **Acta Scientiarum**, Maringá, v. 36, p. 303-310, 2014.
- BOUTON, P.E.; HARRIS, P.V.; SHORTHOSE, W.R. Effects of ultimate pH upon the water-holding capacity and tenderness of mutton. **Journal of Food Science**, Malden, v. 36, p. 435-439, 1971.
- BRESSAN, M.C.; PRADO, O.V.; PÉREZ, J.R.O.; LEMOS, A.L.S.C.; BONAGURIO, S. Efeito ao abate de cordeiros Santa Inês e Bergamácia sobre as características físico-químicas da carne. **Ciência e Tecnologia de Alimentos**, Campinas, v. 21, p. 293-303, 2001.
- CEZAR, M.F.; SOUSA, W.H. Proposta de avaliação e classificação de carcaças de ovinos deslanados e caprinos. **Tecnologia & Ciência Agropecuária**, João Pessoa, v. 4, p. 41-51, 2010.
- CEZAR, M.F.; SOUSA, W.H. **Carcaças ovinas e caprinas**. Uberaba: Editora agropecuária tropical, 2007. 231p.
- COSTA, R.G.; SANTOS, N.M.; SOUSA, W.H.; QUEIROGA, R.C.R.E.; AZEVEDO, P.S.; CARTAXO, F.Q. Qualidade física e sensorial da carne de cordeiros de três genótipos alimentados com rações formuladas com duas relações volumoso:concentrado. **Revista Brasileira de Zootecnia**, Viçosa, v. 40, p. 1781-1787, 2011.
- DAETWYLER, H.D.; PONG-WONG, R.; VILLANUEVA, B.; WOOLLIAMS, J.A. The Impact of Genetic Architecture on Genome-Wide Evaluation Methods. **Genetics**, Austin, v. 185, p. 1021–1031, 2010.
- HARA, A.; RADIN, N.S. Lipid extraction of tissues with low-toxicity solvent. **Analytical Biochemistry**, San Diego, v.90, p.420-426, 1978.
- HU, Z.L.; PARK, C.A.; WU, X.L.; REECY, J.M. Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era. **Nucleic Acids Research**, Oxford, v. 41, p.871-D87, 2013.

JUCÁ, A. **Características de desenvolvimento, morfologia e carcaça em ovinos Santa Inês**. 2013. 97p. Tese (Doutorado em Ciência Animal nos Trópicos) - Escola de Medicina Veterinária e Zootecnia, Universidade Federal da Bahia, Salvador, 2013.

LAWRIE, R.A. **Ciência da carne**. 6. ed. Porto Alegre: Artmed, 2005. 384p.

NCBI. Map Viewer. Disponível em:

http://www.ncbi.nlm.nih.gov/projects/mapview/map_search.cgi?taxid=9940&build=10.0 . Acesso em: 25 set. 2015

MCPHERRON, A.C.; LAWLER, A.M.; LEE, S.J. Regulation of skeletal muscle mass in mice by a new TGF- β superfamily member. *Nature*, New York, v.387, p.83-90, 1997.

MÜLLER, L. **Normas para avaliação de carcaças e concurso de carcaças de novilhos**. 2. ed. Santa Maria: Universidade Federal de Santa Maria, 1987. 32p.

OLIVEIRA, I.; SILVA, T.J.P.; FREITAS, M.Q.; TORTELLY, R.; PAULINO F.O. Caracterização do processo de Rigor Mortis em músculos de cordeiros e carneiros da raça Santa Inês e maciez da carne. **Acta Scientiae Veterinariae**, Porto Alegre, v. 32, p. 25-31, 2004.

RENERRE, M. Review: factors involved in the discoloration of beef meat. **International Journal of Food Science and Technology**, London, v. 25, p. 613-630, 1990.

SAÑUDO, C., ENSER, M.E., CAMPO, M.M., NUTE, G.R.; MARÍA, G.; SIERRA, I.; WOOD, J.D. Fatty acid composition and sensory characteristic of lamb carcasses from Britain and Spain. **Meat Science**, Barking, v. 54, p. 339-346, 2000.

VIEIRA, T.R.L.; CUNHA, M.G.G.; GARRUTTI, D.S.; DUARTE, T.F.; FÉLEX, S.S.S.; Pereira FILHO, J.M.; MADRUGA, M.S. Propriedades físicas e sensoriais da carne de cordeiros Santa Inês terminados em dietas com diferentes níveis de caroço de algodão integral (*Gossypium hirsutum*). **Ciência e Tecnologia de Alimentos**, Campinas, v. 30, p. 372-377, 2010.

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

WHEELER, T. L.; SHACKELFORD, S. D.; JOHNSON, L. P.; MILLER, M. F.; MILLER, R. K.; KOOHMARAIE, M. A comparison of Warner-Bratzler shear force assessment within and among institutions. **Journal of Animal Science**, Champaign, v. 75, p. 2423-2432, 1997.

ANEXOS

Anexo A – Programação para simular um fenótipo, utilizado no programa QMSim, com coeficiente de herdabilidade 0,40 e sem efeito poligênico, e com baixa pressão de seleção na população de estudo

```

/*****
** Global parameters **
*****/
title = "Característica 1 - tese (SFNP2015)";
nrep = 10;           //Number of replicates
h2 = 0.40;          //Heritability
qtlh2 = 0.40;       //QTL heritability
phvar = 1.0;        //Phenotypic variance

/*****
** Historical population **
*****/
begin_hp;
  hg_size = 1000[0]
           1000[1000] //Size of the historical generations
           200 [2020];
// nmfhg = 500;      //Number of males in the first generation
// nmlhg = 100;      //Number of males in the last generation
end_hp;

/*****
** Populations **
*****/
begin_pop = "expand_pop"; //expansion
begin_founder;
  male [n = 100, pop = "hp"];
  female [n = 100, pop = "hp"];
end_founder;
ls = 5;           //Litter size
pmp = 0.5 /fix;  //Proportion of male progeny
ng = 8;          //Number of generations
md = rnd_ug;     // random union of gametes
dr = 1 1 /e;     //key parameter for population expansion
//generating summary data for the expanded population
begin_popoutput;
  data;
  stat;
end_popoutput;
end_pop;

```

```

begin_pop = "recent_pop1"; //recent population
begin_founder;
  male [n = 640, pop = "expand_pop", gen = 8];
  female [n = 32000, pop = "expand_pop", gen = 8];
end_founder;
ls = 1; //Litter size
pmp = 0.5; //Proportion of male progeny
ng = 5; //Number of generations
md = rnd; //Mating design
sr = 0.6; //Replacement ratio for sires
dr = 0.2; //Replacement ratio for dams
sd = ebv/h; //Selection design
cd = ebv/l; //Culling design
ebv_est = blup /true_av;
begin_popoutput;
  stat; //Statistics on simulated data
  data; //Individual's data (except their genotypes)
  allele_freq; //Allele frequencies
  genotype /snp_code; //Genotype data
end_popoutput;
end_pop;

```

```

begin_pop = "recent_pop2"; //recent population
begin_founder;
  male [n = 50, pop = "recent_pop1", gen = 5];
  female [n = 1350, pop = "recent_pop1", gen = 5];
end_founder;
ls = 1; //Litter size
pmp = 0.5; //Proportion of male progeny
ng = 5; //Number of generations
md = rnd; //Mating design
sr = 0.2; //Replacement ratio for sires
dr = 0.05; //Replacement ratio for dams
sd = ebv/h; //Selection design
cd = ebv/l; //Culling design
ebv_est = blup /true_av;
begin_popoutput;
  stat; //Statistics on simulated data
  data; //Individual's data (except their genotypes)
  allele_freq; //Allele frequencies
  genotype /snp_code; //Genotype data
  ld /maft 0.1 /gen 0 5; //LD data
end_popoutput;

```

end_pop;

```

/*****
**      Genome      **
*****/
begin_genome;
  begin_chr = 5;
  chrln = 100;      //Chromosome length
  nmloci = 5000;    //Number of markers
  mpos = even;      //Marker positions
  nma = all 2;      //Number of marker alleles
  maf = eql;        //Marker allele frequencies
  nqloci = 150;     //Number of QTL
  qpos = rnd;       //QTL positions
  nqa = all 2;      //Number of QTL alleles
  qaf = eql;        //QTL allele frequencies
  qae = rndg 0.4;   //QTL allele effects
end_chr;
  select_seg_loci /maft 0.01 /nmrk 7500 /nqtl 30; //Select segregating loci in the last historical
generation: default=none
  mmutr = 1e-5 /recurrent; //Marker mutation rate
  qmutr = 1e-5 /recurrent; //QTL mutation rate
end_genome;

/*****
**      Output options      **
*****/
begin_output;
  linkage_map;
  allele_effect;
  hp_stat;          //Brief statistics on historical population
  monitor_hp_homo; //Mean marker and QTL homozygosity of the historical population
end_output;

```