

BRUNA LIMA BARBOSA

**ASSOCIAÇÃO E SELEÇÃO GENÔMICA AMPLA EM POPULAÇÕES
SIMULADAS DE OVINOS**

TERESINA – PIAUÍ

2020

BRUNA LIMA BARBOSA

**ASSOCIAÇÃO E SELEÇÃO GENÔMICA AMPLA EM POPULAÇÕES
SIMULADAS DE OVINOS**

Tese apresentada ao Programa de Pós-Graduação em
Ciência Animal da Universidade Federal do Piauí como
requisito para obtenção do título de Doutor em Ciência
Animal

Área de concentração: Produção animal

Orientador: Prof. Dr. José Lindenberg Rocha Sarmento

Coorientador: Prof. Dr. Luiz Fernando Brito

TERESINA – PIAUÍ

2020

FICHA CATALOGRÁFICA
Universidade Federal do Piauí
Biblioteca Comunitária Jornalista Carlos Castello Branco
Serviço de Processos Técnicos

B238a Barbosa, Bruna Lima.
 Associação e seleção genômica ampla em populações
 simuladas de ovinos. / Bruna Lima Barbosa. – 2020.
 106 f.: il.

 Tese (Doutorado) – Universidade Federal do Piauí, Pós-
 Graduação em Ciência Animal, Teresina, 2020.
 “Orientador: Prof. Dr. José Lindenberg Rocha Sarmiento.”

 1. Acurácia e predição. 2. GWAS. 3. *Ovis aries*. 4. Predição
 genômica. 5. ssGBLUP. 6. ssGWAS. I. Barbosa, Bruna Lima.
 II. Título.

CDD 636.3

Bibliotecário: Gésio dos Santos Barros – CRB-3/1469

**ASSOCIAÇÃO E SELEÇÃO GENÔMICA AMPLA EM POPULAÇÕES
SIMULADAS DE OVINOS**

BRUNA LIMA BARBOSA

Tese aprovada em: 12/11/2020.

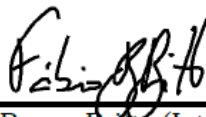
Banca examinadora:



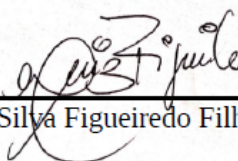
Prof. Dr. José Lindenberg Rocha Sarmiento (Presidente) / DZO/CCA/UFPI



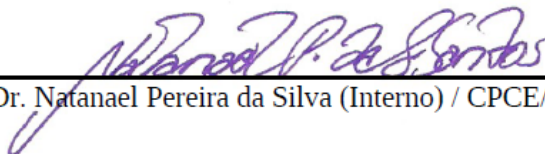
Prof. Dr. Luiz Fernando Brito (Co-orientador) / *Purdue University*



Prof. Dr. Fábio Barros Brito (Interno)/ CCN/UFPI



Prof. Dr. Luiz Antônio Silva Figueiredo Filho (Externo)/ IFMA



Prof. Dr. Natanael Pereira da Silva (Interno) / CPCE/UFPI



Prof. Dr. Gleyson Vieira dos Santos (Externo) / UEMA

AGRADECIMENTOS

Agradeço ao Programa de Pós-graduação em Ciência Animal pela oportunidade de cursar o doutorado, à CAPES pela bolsa concedida durante este período e à Universidade Federal do Piauí, minha segunda casa desde a Graduação, pela possibilidade de aprendizado e formação pessoal e profissional.

Ao professor José Elivalto Guimarães Campelo pela paciência, motivação e ensinamentos. Ao meu orientador, professor José Lindenberg Rocha Sarmiento, pela compreensão, conhecimento repassado e todo o apoio concedido e, ao meu coorientador, Luiz Fernando Brito pelas palavras de motivação, humildade e sabedoria. Grandes mestres!

À Deus, toda honra e Glória, e à Nossa Senhora pelo amparo em todos os momentos. Força e luz nas batalhas da vida.

À minha mãe, Maria Senhora Lima Barbosa, dedico tudo que conquistei e tudo que almejo em minha vida. É tudo pela senhora, mãe!

Agradeço ao meu amigo Luciano Sena pela ajuda e companheirismo de sempre, em todos os momentos. Ao André Campelo pela sabedoria, ensinamentos, motivação e humildade compartilhada. Vocês são exemplos a serem seguidos.

Max Brandão e Romuere pelos auxílios estatísticos e todos os ensinamentos. Sempre disponíveis a ajudar.

Aos meus amigos da Pós-graduação: Lilian, Abigail, Artur, Tatiana, Laylson, João Lopes e Geandro pelos momentos de alegria, tristeza, compreensão, apoio e aprendizado. À Débora Carvalho, amiga de luta e de oração, que sempre me diz e guardo sempre comigo: Deus não dá uma benção pela metade! Sou grata também às amigas Ana Paula e Dayane que me motivaram, auxiliaram e apoiaram na finalização desse trabalho. Ao meu amigo de batalha, em meios aos inúmeros obstáculos de nossos trabalhos, Expedito Ulisses, meu muito obrigado!

Aos amigos de vida, Natália, Bruno e Jaqueline pela torcida de sempre! Márcia, Lháis, Lívia, Luana, Verônica, Rosana, Gláucia e Joanna por me apoiarem e estarem sempre me motivando a conquistar meus objetivos. À minhas primas, Jéssica e Samantha, pelo amor e incentivo! Amo todos vocês!

Ao meu pai, José Barbosa e minha irmã, Bianca Lima, por me ajudarem nos períodos conturbados! Amo vocês!

Por fim, sou grata a todos que, direta ou indiretamente, torceram para que esta etapa de minha vida fosse concluída.

Deus os abençoe!

Sumário

LISTA DE FIGURAS.....	vi
LISTA DE TABELAS.....	viii
RESUMO.....	9
ABSTRACT.....	10
1. INTRODUÇÃO GERAL.....	11
2. REVISÃO DE LITERATURA.....	12
2.1 O uso da Genômica na produção animal	12
2.2 Limitações e desafios para aplicação da genômica na produção animal	16
2.3 Estudos genômicos com dados simulados.....	17
2.4. <i>ssGBLUP</i> e <i>ssGWAS</i>	18
2.5. Abordagem para detecção de QTLs em GWAS	21
3. REFERÊNCIAS.....	22
CAPÍTULO I – Associação genômica ampla em populações simuladas de ovinos...28	
RESUMO.....	29
ABSTRACT.....	30
1. Introdução.....	31
2. Material e métodos.....	31
3. Resultados.....	38
4. Discussão.....	61
5. Conclusões	62
Referências.....	64
CAPÍTULO II - Seleção genômica ampla em populações simuladas de ovinos66	
Resumo.....	67
Abstract.....	68
1. Introdução.....	69
2. Materiais e métodos.....	70
3. Resultados.....	75
4. Discussão.....	80
5. Conclusões.....	82
Referências.....	83
CONSIDERAÇÕES FINAIS.....	85
ANEXOS	86

LISTA DE FIGURAS

CAPÍTULO I

- Figura 1. *Manhattan plots* com abordagem do p -valor exato para dados simulados de ovinos utilizando painéis de alta, média e baixa densidade e duas quantidades de QTLs. Representação da janela do cromossomo 9 referente aos cenários com 15 mil informações de genótipos e fenótipos.....42
- Figura 2. Gráficos Quantil-quantil (Q-Qplot) para o $-\log_{10}(p\text{-valor exato})$ de dados simulados de ovinos utilizando painéis de alta, média e baixa densidade e duas quantidades de QTLs. Representação da janela do cromossomo 9 referente aos cenários com 15 mil informações de genótipos e fenótipos.....43
- Figura 3. *Manhattan plots* com abordagem do p -valor exato para dados simulados de ovinos utilizando painéis de alta, média e baixa densidade e duas quantidades de QTLs. Representação da janela do cromossomo 2 referente aos cenários com 15 mil fenótipos e 7.500 informações de genótipos.....44
- Figura 4. Gráficos Quantil-quantil (Q-Qplot) para o $-\log_{10}(p\text{-valor exato})$ de dados simulados de ovinos utilizando painéis de alta, média e baixa densidade e duas quantidades de QTLs. Representação da janela do cromossomo 2 referente aos cenários com 15 mil fenótipos e 7.500 informações de genótipos.....45
- Figura 5. *Manhattan plots* com abordagem do p -valor exato para dados simulados de ovinos utilizando painéis de alta, média e baixa densidade e duas quantidades de QTLs. Representação da janela do cromossomo 25 referente aos cenários com 5 mil informações de genótipos e fenótipos.....47
- Figura 6. Gráficos Quantil-quantil (Q-Qplot) para o $-\log_{10}(p\text{-valor exato})$ de dados simulados de ovinos utilizando painéis de alta, média e baixa densidade e duas quantidades de QTLs. Representação da janela do cromossomo 25 referente aos cenários com 5 mil informações de genótipos e fenótipos48
- Figura 7. *Manhattan plots* com abordagem do p -valor exato para dados simulados de ovinos utilizando painéis de alta, média e baixa densidade e duas quantidades de QTLs. Representação da janela do cromossomo 16 referente aos cenários com 5 mil fenótipos e 2.500 informações de genótipos.....49
- Figura 8. Gráficos Quantil-quantil (Q-Qplot) para o $-\log_{10}(p\text{-valor exato})$ de dados simulados de ovinos utilizando painéis de alta, média e baixa densidade e duas quantidades de QTLs. Representação da janela do cromossomo 16 referente aos cenários com 5 mil fenótipos e 2.500 informações de genótipos.....50
- Figura 9. *Manhattan plots* com abordagem do p -valor exato para dados simulados de ovinos utilizando painéis de alta, média e baixa densidade e duas quantidades de QTLs. Representação da janela do cromossomo 2 referente aos cenários com 2mil informações de genótipos e fenótipos.....52

- Figura 10. Gráficos Quantil-quantil (Q-Qplot) para o $-\log_{10}(p\text{-valor exato})$ de dados simulados de ovinos utilizando painéis de alta, média e baixa densidade e duas quantidades de QTLs. Representação da janela do cromossomo 2 referente aos cenários com 2mil informações de genótipos e fenótipos.....53
- Figura 11. *Manhattan plots* com abordagem do p -valor exato para dados simulados de ovinos utilizando painéis de alta, média e baixa densidade e duas quantidades de QTLs. Representação da janela do cromossomo 9 referente aos cenários com 2 mil fenótipos e 1 mil informações de genótipos.....54
- Figura 12. Gráficos Quantil-quantil (Q-Qplot) para o $-\log_{10}(p\text{-valor exato})$ de dados simulados de ovinos utilizando painéis de alta, média e baixa densidade e duas quantidades de QTLs. Representação da janela do cromossomo 9 referente aos cenários com 2 mil fenótipos e 1 mil informações de genótipos.....55
- Figura 13. *Manhattan plots* com abordagem do p -valor exato para dados simulados de ovinos utilizando painéis de alta, média e baixa densidade e duas quantidades de QTLs. Representação da janela do cromossomo 1 referente aos cenários com 500 informações de genótipos e fenótipos.....57
- Figura 14. Gráficos Quantil-quantil (Q-Qplot) para o $-\log_{10}(p\text{-valor exato})$ de dados simulados de ovinos utilizando painéis de alta, média e baixa densidade e duas quantidades de QTLs. Representação da janela do cromossomo 1 referente aos cenários com 500 informações de genótipos e fenótipos.....58
- Figura 15. *Manhattan plots* com abordagem do p -valor exato para dados simulados de ovinos utilizando painéis de alta, média e baixa densidade e duas quantidades de QTLs. Representação da janela do cromossomo 11 referente aos cenários com 500 fenótipos e 250 informações de genótipos.....59
- Figura 16. Gráficos Quantil-quantil (Q-Qplot) para o $-\log_{10}(p\text{-valor exato})$ de dados simulados de ovinos utilizando painéis de alta, média e baixa densidade e duas quantidades de QTLs. Representação da janela do cromossomo 11 referente aos cenários com 500 fenótipos e 250 informações de genótipos.....60

LISTA DE TABELAS

REVISÃO DE LITERATURA

Tabela 1. Alguns genes candidatos associados a características de carcaça e qualidade de carne identificados por meio de estudos de associação genômica ampla (GWAS)15

CAPÍTULO I

Tabela 1. Cenários para análises de Associação Genômica Ampla com dados simulados para área de olho de lombo em *Ovis aries*.....35

Tabela 2. Estimativas de componentes de variância, herdabilidades e erros padrão usando o ssGWAS com 2.014 QTLs para os cenários com três painéis de densidades alta, média e baixa e populações com tamanhos e número de animais genotipados distintos.....39

Tabela 3. Estimativas de componentes de variância, herdabilidades e erros padrão usando o ssGWAS com 114 QTLs para os cenários com três painéis de densidades alta, média e baixa e populações com tamanhos e número de animais genotipados distintos.....40

CAPÍTULO II

Tabela 1. Cenários utilizados para seleção genômica com dados simulados em ovinos.....73

Tabela 2. A média das estimativas de componentes de variância, herdabilidade e erro padrão (entre parênteses) obtidas com uso do método ssGBLUP com 2.014 QTLs para os cenários com dois painéis de densidades moderada e baixa e populações com tamanhos e número de animais genotipados distintos.....76

Tabela 3. Médias das Correlações e viés dos cenários com painéis de 12k e 50k marcadores SNPs simulados.....77

Tabela 4. Análise de variância para os efeitos de densidade dos painéis, tamanho populacional e para a interação entre ambos.....78

Tabela 5. Acurácia da predição (Corr) e desvios padrão (SD) para cenários com tamanhos populacionais com diferentes quantidades de animais genotipados na população de treinamento.....79

RESUMO

BARBOSA, Bruna Lima. **Associação e seleção genômica ampla em populações simuladas de ovinos**. 2020. Tese (Doutorado em Ciência Animal) – Universidade Federal do Piauí, Teresina, Piauí, 2020.

A utilização de dados simulados é útil para o desenvolvimento de métodos genômicos e avaliação de cenários que seriam mais complexos com o uso de dados reais. Neste estudo, objetivou-se avaliar a eficiência da detecção de QTLs e acurácias de predição genômica, com uso de diferentes tamanhos populacionais, densidade de *Chip* de SNPs e estratégias de genotipagem. No Capítulo 1, os QTLs foram detectados por estudo de associação genômica ampla com uso do método *single-step* GWAS e abordagem do *p*-valor exato, com informações geradas por meio de simulações baseadas em parâmetros de dados reais de AOL. Painéis simulados de SNPs de três densidades diferentes foram utilizados com base na estrutura do genoma ovino. Animais das três últimas gerações constituíram 48 cenários propostos de acordo com o painel de SNPs, quantidade de QTLs e número de animais genotipados. O limite de significância para um conjunto de marcadores foi obtido pela correção de *Bonferroni* e taxa de falsa descoberta. Verificou-se que quanto maior a quantidade de animais genotipados e SNPs no painel, maior a possibilidade de detecção de associações verdadeiras entre os SNPs significativos e os QTLs simulados. Com 2.014 QTLs, os cenários com o painel de alta densidade proporcionaram maior quantidade de detecções verdadeiras. Com 114 QTLs, geralmente, observou-se pelo menos uma associação verdadeira, com picos maiores nos cenários com maior quantidade de animais. Portanto, a abordagem do *p*-valor exato possibilitou a detecção de pelo menos uma associação verdadeira entre SNPs e QTLs simulados. No Capítulo 2, dois painéis de SNPs (50K e 12K) foram utilizados. Diferentes tamanhos de populações de referência e validação compuseram seis cenários diferentes para cada painel. O método *single-step* GBLUP foi utilizado para a estimação dos efeitos dos SNPs. As estimativas de herdabilidade variaram de 0,17 a 0,38. Os maiores valores de acurácia foram obtidos com uso das estruturas de treinamento com maior quantidade de animais genotipados, enquanto os menores valores deste parâmetro foram observados nas populações com menos animais genotipados na estrutura de treinamento. Diferenças significativas foram observadas entre os diferentes tamanhos da população de treinamento, enquanto que nos cenários com o treinamento e com menor quantidade de informações de genótipos e fenótipos, não foi verificada diferença significativa entre os resultados obtidos com uso dos dois painéis. Há uma tendência de aumento da acurácia com o acréscimo de animais genotipados na população de treinamento ou referência.

Palavras-chave: acurácia de predição; GWAS; *Ovis aries*; predição genômica; ssGBLUP; ssGWAS.

ABSTRACT

BARBOSA, Bruna Lima. **Genome-wide association and selection in simulated population of sheep.** 2020. Thesis (Doctor of Philosophy in Animal Science) – Universidade Federal do Piauí, Teresina, 2020.

The use of simulated data is essential for the development of genomic methods and evaluation of scenarios that would be unfeasible using real data. In this study, we aimed to assess the detection of QTLs associated with a phenotype like the loin eye area (LEA), as well as the efficiency of genomic selection in sheep (*Ovis aries*) for LEA, using simulated scenarios with different genetic structures, population sizes, and number of genotyped animals. In Chapter 1, QTLs were detected by genome-wide association study using the single-step GBLUP method and the exact p-value approach, with information generated through simulations based on parameters of real data of LEA. Simulated SNP panels of three different densities were used based on the structure of the sheep genome. Animals of the three last generations composed 48 scenarios proposed based the SNP panel, number of QTLs, and number of genotyped animals. The significance of SNP effects was obtained by the Bonferroni correction and false discovery rate. We observed that larger number of genotyped animals and higher density of SNP panels increased the possibility of detection of true associations between significant SNPs and simulated QTLs. When 2,014 QTLs were used, the scenarios with high density panel allowed a higher number of true detections. In general, when 114 QTLs were used, we observed at least one true association with higher peaks is scenarios containing larger number of animals. Therefore, the exact p-value approach allowed the detection of at least one true association between SNPs and simulated QTLs. In Chapter 2, two SNP panels (12k and 50k) were used. Different sizes of reference and validation populations composed six different scenarios for each panel. The single-step GBLUP method was used to estimate SNP effects on LEA. The prediction accuracy was obtained using the Pearson correlation between genomic estimated breeding values (GEBVs) and true breeding values (TBVs), as well as the regression coefficients, and bias of prediction. Heritability estimates ranged from 0.17 to 0.38. The highest values were obtained using training sets with larger number of genotyped animals, whereas the lowest values were observed when training populations containing lower number of genotyped animals were used. Significant differences were observed between scenarios with different training population sizes; however, no significant differences were observed among the results of scenarios with lower training population sizes, considering the two SNP panels evaluated. There is a trend of increasing the prediction accuracy with the addition of genotyped animals either in the training or reference populations.

Keywords: accuracy; genomic prediction; GWAS; *Ovis aries*; SNPs; ssGBLUP.

1 INTRODUÇÃO GERAL

As informações oriundas de marcadores moleculares têm sido utilizadas em programas de melhoramento genético de diferentes espécies, em conjunto com informações fenotípicas e de pedigree (SIMIANER, 2016). O advento de painéis de polimorfismos de nucleotídeo simples (SNPs) e o desenvolvimento de métodos estatísticos para a inclusão de dados de SNPs em avaliações genéticas resultaram na implementação da seleção genômica ampla (GWS), proposta por Meuwissen, Hayes e Goddard (2001). Além da aplicação na GWS, os painéis de SNPs tornaram possível a detecção de QTL (*loci* de características quantitativas) e o melhor entendimento sobre a arquitetura genética de características de importância econômica, por meio de estudos de associação genômica ampla (GWAS) (DEKKERS, 2012).

A GWS envolve a estimação do efeito de todos os SNPs, simultaneamente, considerando seus efeitos como aleatórios (MEUWISSEN; HAYES; GODDARD, 2016), de modo que os QTLs estejam em desequilíbrio de ligação (LD) com pelo menos um SNP. Os métodos desenvolvidos para GWS diferem, essencialmente, nas pressuposições feitas sobre a distribuição dos efeitos dos marcadores (DEKKERS, 2012). A maioria dos métodos disponíveis para a realização de predições genômicas envolve a execução de várias etapas e, portanto, são conhecidos como métodos *multi-step*, entre os quais se destacam os métodos Bayesianos e BLUP genômico (GBLUP) (LEGARRA et al., 2014).

No entanto, os métodos *multi-step* apresentam alta complexidade por derivar valores genéticos genômicos envolvendo várias etapas, além de possibilitar a avaliação do mérito genômico apenas para animais genotipados (PÉRTILE et al., 2016). Um método alternativo, conhecido como *single-step* ou ssGBLUP, foi desenvolvido por Legarra et al. (2009), Misztal et al. (2009); Aguilar et al. (2010) e Christensen e Lund (2010), para possibilitar também a estimação de valores genéticos genômicos para animais não genotipados.

Desde sua implementação, o método ssGBLUP tem tido grande aceitação por pesquisadores, de modo que, há anos é o método mais utilizado mundialmente para GWS (MISZTAL, 2016). Apesar do método *single-step* ser mais intensivo computacionalmente, o mesmo tem apresentado maior acurácia em relação aos demais métodos, quando aplicado em situações em que nem todos os animais são genotipados, que é a realidade da maioria dos rebanhos (LEGARRA et al., 2014; PÉRTILE et al., 2016), cujo uso em análises de Associação Genômica Ampla (GWAS) também tem se mostrado eficiente.

A GWAS busca compreender a estrutura genética dos indivíduos baseada na expressão de genes que estejam associados a características produtivas. Seu princípio é baseado nas variações presentes nos genomas de indivíduos geneticamente relacionados, também conhecidas como marcadores moleculares, sendo os SNPs os mais utilizados pela eficiência na identificação de QTLs que afetam o fenótipo (WANG et al., 2012; LI et al., 2017). As análises realizadas em GWAS buscam a associação de *locus* de características quantitativas e qualitativas associadas com características complexas e poligênicas (MEUWISSEN; HAYES; GODDARD, 2001). O conhecimento de genes associados a características de interesse proporcionou grandes avanços na produção animal.

A tese em questão está estruturada de acordo com as normas estabelecidas pelo Programa de Pós-Graduação em Ciência Animal da Universidade Federal do Piauí (UFPI), sob a resolução 001/03 – CCMCA. O trabalho está apresentado conforme a seguinte estrutura: uma introdução geral e revisão de literatura sobre as informações mais relevantes do tema abordado; Capítulo I “Associação genômica ampla em populações simuladas de ovinos” e Capítulo II “Seleção genômica ampla em populações simuladas de ovinos”. Por fim, são apresentadas as Considerações Finais, reforçando os principais aspectos desta pesquisa, assim como sugestões para sua continuação.

2 REVISÃO DE LITERATURA

2.1 O uso da Genômica na produção animal

A eficiência da inclusão de informações genômicas para o progresso genético de características economicamente importantes já foi demonstrado em estudos de características diretamente relacionadas ao aumento da produção ou melhoria de qualidade de carne em bovinos de corte (ONOGI et al., 2014; GORDO et al., 2016; MEHRBAN et al., 2019), suínos (MEUWISSEN et al., 2014; SAMORÈ; FONTANESI, 2016; SONG et al., 2019), ovinos de corte (AUVRAY et al., 2014; BRITO et al., 2017) e frangos de corte (LIU et al., 2014), por exemplo. No entanto, a seleção genômica tem resultado em maiores benefícios em bovinos leiteiros, pois esta categoria apresenta a cadeia produtiva melhor estruturada e maior quantidade de animais (de vários países) nas populações de referência (MEUWISSEN; HAYES; GODDARD, 2016).

Nos inúmeros resultados disponíveis na literatura científica e em situações reais de programas de melhoramento genético, a eficiência da seleção genômica tem sido demonstrada por proporcionar o aumento da acurácia de predição do valor genético

quando comparado aos métodos de seleção tradicionais (MEUWISSEN; HAYES; GODDARD, 2016), permitindo assim, identificar e selecionar animais melhoradores com maior acurácia, a fim de aumentar a produtividade nos sistemas de produção, principalmente em características de interesse econômico, como tamanho da ninhada em suínos (FORNI et al., 2011), peso corporal e rendimento de peito em frangos (CHEN et al., 2011), peso da carcaça quente em ovinos (DAETWYLER et al., 2012), características de carcaça e peso ao ano em bovinos (MEHRBAN et al., 2019).

Outras vantagens da GWS dizem respeito à diminuição do intervalo de geração, redução nos custos relacionados a manutenção do animal no rebanho e testes de progênie, devido à possibilidade de avaliar o mérito genômico de maneira acurada em animais jovens, visto que a coleta de informações precoces no animal permite definir novas estratégias de melhoramento, visando impulsionar o ganho genético, para assim, agilizar o processo de seleção e melhorar a confiabilidade das estimativas (IBTISHAM et al., 2017). No entanto, é importante atentar para o fato de que forte seleção pode levar a redução da variância genética na população e aumento de correlações genéticas indesejáveis de características produtivas, podendo reduzir os ganhos com acurácia ao longo do tempo (HIDALGO et al., 2020).

De modo geral, a seleção genômica apresenta maior vantagem para características de difícil mensuração, como aquelas relacionadas à carcaça e qualidade de carne. Portanto, a inclusão de informações resultantes de marcadores moleculares na avaliação genética pode contribuir significativamente para o melhoramento genético destas características (NAVAJAS, 2014). Diversos programas de melhoramento genético de bovinos de corte (MEUWISSEN; HAYES; GODDARD, 2016; SARMENTO; SENA, 2017), ovinos de corte (SWAN et al., 2014; BRITO et al., 2017; B+LNZ GENETICS, 2017), suínos (TOPIGS NORSVIN, 2017) e frangos de corte (WOLC et al., 2016), por exemplo, já incluem informações genômicas na avaliação genética rotineira para características de carcaça e/ou qualidade de carne.

No âmbito dos resultados oriundos de estudos de associação genômica ampla, há inúmeros relatos de genes ou regiões genômicas associados a características diretamente ligadas a rendimento de carcaça e qualidade de carne, em diferentes espécies, que podem ser verificados na plataforma *AnimalQTLdb* (<https://www.animalgenome.org>), por exemplo (HU; PARK; REECY, 2019). Em bovinos de corte, há um dos maiores números de resultados de GWAS para características de carcaça e qualidade de carne (SHARMA

et al., 2015). Alguns exemplos de genes candidatos para estas características, identificados em diferentes espécies com uso de GWAS, são apresentados na Tabela 1.

Tabela 1. Alguns genes candidatos associados a características de carcaça e qualidade de carne identificados por meio de estudos de associação genômica ampla (GWAS)

Espécie/Categoria	Gene(s)	Característica	Estudo	
Bovinos de corte	<i>ARFGAP3, CLSTN2 e DPYD</i>	AOL	Santana et al. (2015)	
	<i>OSBPL3, SUDS3, RARRES1 e VEPH1</i>	EG		
	<i>ALKBH3, HSD17B12</i>	AOL	Silva et al. (2017)	
	<i>PLAG1, CAPN5, MYO7A e XKR4</i>	EG		
	<i>CHCHD7, LYPLA e LYN</i>	EG	Hay e Roberts (2018)	
	<i>GNPDA2</i>	AOL		
	<i>ELOVL5, ESSRG, PCYT1A, ABC5, ABC6 e ABC10</i>	PAG	Lemos et al. (2016)	
	<i>PLAG1</i>	PCQ	Fernandes Júnior et al. (2016)	
	<i>SORCS2, AQP3, AQP7, CDC42BPA, ASIP e ACSS2</i>	EG		
	<i>CDKN2A, CDKN2B, SLC38A1 e SLC38A2</i>	AOL		
Suínos	<i>FOXA3, AP2S1, FKRFP, NPASI e ATP6V1G1</i>	MS	Silva et al. (2019)	
	<i>PLTP, TNNC2 e GPAT2</i>	AOL		
	<i>TKT, FNDC5 e CHRND</i>	EG	Iqbal et al. (2015)	
	<i>PRKCA e SMN1</i>	PC		
	<i>MX1 e MX2</i>	EG	Hérault et al. (2018)	
	<i>PRKAA1, CIDEA, ABCD2, FABP7, ELOVL4 e ACAA2</i>	EG		
	Aves de corte	<i>LCORL, LAP3, LDB2 e TAPT1</i>	PC e PCE	Liu et al. (2013)
		<i>TYRO3, MGST1 e COL1A2</i>	COC	Sun et al. (2013)
		<i>RET, NPPB e SREBF1</i>	DGA	
		<i>CBLN2 e HPGDS</i>	GPER	Zhang et al. (2015)
<i>SETD2 e ANKRD46</i>		GPEI		
Ovinos de corte	<i>ALDOA, STK32B e FAM190A</i>	MAR	Wang et al. (2015)	
	<i>CREB1, STEAP4, CTBP1 e RIP140</i>	DGC	Xu et al. (2017)	
	<i>NDUFAF2</i>	EG	Hernandez et al. (2018)	
	<i>ACACA</i>	PM		
	<i>DGAT2, TRHDE, TPH2, ME1, C6, C7, UBE3D, PARP14 e MRPS30</i>	PAG	Rovadoski et al. (2018)	

AOL: Área de olho de lombo; EG: Espessura de gordura; PC: Peso de carcaça; PCQ: Peso de carcaça quente; PCE: Peso de carcaça eviscerada; PAG: Perfil de ácidos graxos; MS: musculatura; COC: coloração da carne; DGA: deposição de gordura abdominal; GPER: conteúdo de gordura na perna; GPEI: conteúdo de gordura no peito; DGC: deposição de gordura na cauda; PM: profundidade muscular; MAR: marmoreio.

2.2 Limitações e desafios para aplicação da genômica na produção animal

Embora os benefícios da aplicação da genômica já estejam bem documentados na literatura, há alguns aspectos que podem comprometer a viabilidade da utilização de informações genômicas para as finalidades relatadas anteriormente (seleção e associação). De acordo com Sharma et al. (2015), algumas das seguintes medidas podem ser úteis para evitar problemas com resultados de estudos genômicos: definição detalhada do estudo antes de sua realização (escolha de animais, metodologia estatística e painel de genotipagem, por exemplo); e controle de qualidade adequado dos dados a serem utilizados nas análises genômicas. No entanto, a maior parte dos fatores é de ordem estrutural e/ou financeira, de modo que não estão sob o controle dos pesquisadores.

O principal problema para implementação de GWS é a necessidade de se ter grandes populações de treinamento para a obtenção de equações de predição com capacidade de relacionar SNPs com informações fenotípicas de forma mais acurada (BLASCO; TORO, 2014). Isto porque, com exceção de bovinos leiteiros e suínos, na maior parte das situações relacionadas a outras espécies de animais de produção, os efetivos populacionais são bem menores, há carência de coleta consistente de fenótipos e controle zootécnico, assim como grandes limitações para investimentos em genotipagem da cadeia produtiva (BLASCO; TORO, 2014; SHARMA et al., 2015; MEUWISSEN; HAYES; GODDARD, 2016), principalmente em países subdesenvolvidos (MRODE et al., 2018, 2019). Em espécies como caprinos e ovinos, cujos animais apresentam baixo valor econômico, em comparação a bovinos, por exemplo, os custos de genotipagem apresentam a maior relevância na decisão sobre a viabilidade de implantação de seleção genômica em programas de melhoramento (BLASCO; TORO, 2014; RUPP et al., 2016).

No que diz respeito à viabilidade de estudos de associação genômica, os mesmos fatores limitantes relatados acima são válidos. De acordo com Sharma et al. (2015), os principais desafios para GWAS incluem a escolha adequada de uma população homogênea e a contabilização da estratificação populacional existente. Uma alternativa interessante para contornar algumas limitações do uso de dados reais, para verificar a viabilidade de aplicação da genômica, consiste no uso de análises de dados simulados, pois as simulações demandam apenas recursos computacionais e definições adequadas de métodos estatísticos (DAETWYLER et al., 2013).

2.3. Estudos genômicos com dados simulados

Os estudos genômicos por meio de simulações têm se tornado cada vez mais importantes para os avanços nas metodologias aplicadas ao melhoramento animal. Devido às limitações referentes a GWAS e GWS, especialmente pelo alto custo para obtenção e incorporação de informações genômicas através de *chips* de SNPs em pesquisas e programas de melhoramento genético, informações advindas de simulações são importantes. O potencial de métodos estatísticos para adequação e detecção de QTLs também pode ser aproveitado com a simulação de dados (HICKEY; GORJANC, 2012).

Em estudos de associação genômica ampla e na seleção genômica ampla, o sucesso para estimação de efeitos de SNPs e predição da acurácia depende, dentre outros fatores, do desequilíbrio de ligação entre os marcadores e QTLs devido ao desconhecimento, geralmente, das mutações causais responsáveis pela variância fenotípica. Assim, o nível de LD entre o marcador e a mutação pode capturar a variância do marcador (MEUWISSEN; HAYES; GODDARD, 2016). Outros fatores como tamanho efetivo populacional, herdabilidade da característica, estrutura genética da população e número de animais fenotipados e genotipados também afetam o poder de detecção de QTLs e acurácia da predição (VAN DEN BERG et al., 2013).

Diversos estudos têm incorporado a informação simulada em suas estratégias de pesquisa. Pértile et al. (2016) realizaram, através de dados simulados, tomando como referência informações de ovinos Santa Inês, para análises de GWAS e GWS sob diferentes cenários de estudo. Os autores concluíram que as informações de parentesco melhoraram a predição dos valores genéticos genômicos e maiores estimativas de herdabilidade favorecem a identificação de regiões associadas a características de interesse.

Takeda, Uemoto e Satoh (2020) em um estudo simulado com 300 touros genotipados e diferentes quantidades de progênies fenotipadas com diferentes herdabilidades, tamanho efetivo populacional e número de QTLs, usando como base uma população de bovinos para verificar o poder de detecção de QTLs, através do ssGWAS, verificaram que o poder de detecção dos QTLs aumentou com o aumento do número de informações das progênies desses touros em detrimento da herdabilidade ou da quantidade de QTLs utilizados.

Uma comparação entre os métodos GBLUP em um e dois passos em populações simuladas de bovinos de corte foi realizada por Piccoli et al. (2018), que observaram que valores genômicos diretos e GEBVs preditos pelos procedimentos de GBLUP testados

apresentaram precisões muito semelhantes, sem viés significativo nas previsões de GEBV de procedimentos de uma ou duas etapas.

Araujo et al. (2021) utilizaram dados simulados de populações ovinas geneticamente diversas a fim de verificar a acurácia e o viés das previsões genômicas, usando o ssGBLUP, com análises de SNPs individuais e haplótipos, diferentes densidades de painéis e métodos de construção de blocos de haplótipos. Concluíram que os modelos com base em haplótipos, que poderiam melhor capturar o LD entre SNPs e QTLs, não melhoraram os valores dos GEBVs.

Dessa forma, o uso de dados simulados tem sido utilizado frequentemente por pesquisadores, pois possibilita testes replicáveis baseados em cenários reais para o teste de hipóteses a baixo custo, podendo ser ainda investigados efeitos a longo prazo da seleção, inviáveis com o uso de dados reais (DAETWYLER et al., 2013).

2.4 *ssGBLUP e ssGWAS*

Os procedimentos em várias etapas ou *Multi-step*, onde apenas os animais com genótipos são incluídos no modelo (VANRADEN et al., 2008), podem ser relativamente complexos e envolver a dupla contagem das informações genômicas, quando ocorre a inclusão de genótipos de pais e progênie na análise (MISZTAL; LOURENCO; LEGARRA, 2020). Legarra et al. (2014) demonstraram que, durante essas múltiplas etapas, podem ser gerados vieses, o que torna a sua eficácia comprometida para decisões práticas oriundas de estudos genômicos, via seleção genômica.

Desde sua proposição, os métodos em etapa única, que incorporam informações de genótipos, fenótipos e pedigree na mesma rodada de análises, tornaram o processo de avaliação genômica mais simples e possibilitaram a extensão das informações genômicas a animais não-genotipados. Assim, como geralmente há uma pequena porção de genótipos nas avaliações genéticas, a combinação do parentesco por pedigree e genômico surgiu como uma alternativa para esses cenários (LORENCO et al., 2020), possível através de uma distribuição conjunta dos valores genéticos dos animais genotipados ou não (LEGARRA et al., 2009).

O desenvolvimento de uma matriz híbrida, a matriz H, que combina a matriz de parentesco baseada no pedigree (A), considerada como uma informação *a priori*, com a matriz de relacionamento genômico (G), como um parentesco observado, possibilitou que a informação dos SNPs fosse extrapolada aos animais não-genotipados, como uma projeção do mérito genético. Esta matriz é complexa e exige que G seja invertível, sendo

a sua inversa mais simples, a chamada matriz \mathbf{H}^{-1} (AGUILAR et al., 2010; CHRISTENSEN; LUND, 2010). A compatibilidade das informações de pedigree e genômica é importante para evitar vieses e aumentar as acurácias nas avaliações via ssGBLUP, especialmente sob forte seleção (VITEZICA et al., 2011).

Assim, o método denominado de BLUP genômico em passo único, o ssGBLUP (LEGARRA et al., 2009; MISZTAL et al., 2009; AGUILAR et al., 2010; CHRISTENSEN; LUND, 2010), tornou-se consolidado em estudos genômicos de várias espécies, devido a sua simplicidade e acurácia, especialmente para características que não sejam relacionadas a produção de leite (MISZTAL; LOURENCO; LEGARRA, 2020). Pode-se considerar o seguinte modelo abaixo:

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

em que: y é o vetor de observações para cada característica; β é o vetor de efeitos fixos (grupo contemporâneo e co-variável); α é o vetor de efeitos genéticos aditivos, com $\alpha \sim N(0, \mathbf{A}\sigma_a^2)$ na abordagem baseada no pedigree ou $\alpha \sim N(0, \mathbf{H}\sigma_a^2)$ na abordagem genômica, em que σ_a^2 é a variância genética aditiva e \mathbf{H} é a matriz de parentesco que combina os parentescos baseados em informações de pedigree e genômicas; \mathbf{X} é a matriz de incidência dos efeitos fixos; \mathbf{Z} é a matriz de incidência dos efeitos genéticos aditivos; e \mathbf{e} é o vetor de efeitos residuais, $e \sim N(0, \mathbf{I}\sigma_e^2)$.

A inversa da matriz \mathbf{H} normalmente é calculada de acordo com a proposição de Aguilar et al. (2010) e Christensen e Lund (2010), como:

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

onde \mathbf{A}_{22}^{-1} a inversa da matriz de parentesco baseado no pedigree para os animais genotipados.

Algumas vantagens do método ssGBLUP, além da simplicidade, fácil implementação e acurácia maior que a de outros métodos como os de múltiplas etapas (LOURENCO et al., 2014), já apresentadas aqui, incluem o ajuste simultâneo das informações genômicas e estimativas de outros efeitos no modelo (LEGARRA et al., 2014), o que permite o entendimento do viés de pré-seleção e isso auxilia a não haver perda de informação (LOURENCO et al., 2014). Também são permitidas a inclusão de informações de características complexas e multi-característica ao modelo (WANG et al.,

2014). Seguir as pressuposições do método em questão é crucial para o delineamento do estudo genômico apresentar simplicidade em sua aplicação (LOURENCO et al., 2020).

Mesmo apresentando facilidades, os estudos genômicos com a utilização do ssGBLUP possuem desafios. A definição de formas de validação não afetadas por seleção e a redução que a mesma provoca nas variâncias genéticas ao longo do processo seletivo, bem como de estimativas que elucidem todas as informações genômicas utilizadas em seleção podem ser citadas (MISZTAL; LOURENCO; LEGARRA, 2020). Além disso, esse método possui limitações em relação a pequena quantidade de animais genotipados incluídos na análise, bem como para características com QTLs de efeito maior, esta última podendo não ser uma informação verídica, já que o método assume a igualdade nas proporções de variâncias genéticas explicadas pelos marcadores (WANG et al., 2014).

Nesse contexto, Wang et al. (2012) propuseram a inclusão de pesos ponderando os efeitos estimados dos marcadores, de acordo com a variância genética explicada, denominado de *Weighted-ssGBLUP* ou BLUP genômico de passo único ponderado. Este método estima variâncias diferentes para os SNPs e isso possibilita melhor precisão das estimativas em estudos com pequeno número de fenótipos e genótipos, bem como para características afetadas por QTLs de grandes efeitos (WANG et al., 2012). Cinco métodos de inclusão de pesos no WssGBLUP foram testados por ZHANG et al. (2016), mostrando melhoras nos acurácias dos GEBVs.

O ssGBLUP também pode ser utilizado para a estimação dos efeitos de SNPs, detecção de QTLs e genes associados a características de importância econômica através do ssGWAS (WANG et al., 2012), por meio do desequilíbrio de ligação entre os SNPs ou possíveis mutações causais de características de relevância. A ponderação dos efeitos dos SNPs também pode ser aplicada a esse método, sendo o WssGWAS uma abordagem iterativa que promove o aumento dos pesos de SNPs com maiores efeitos e a redução dos que possuem pequenos efeitos à média (WANG et al., 2014).

Zhang et al. (2016) também citaram que o uso de janelas genômicas com ponderações pode detectar QTLs desconhecidos quando poucos as análises apresentarem poucos genótipos. Os mesmos autores também afirmaram que o uso de pesos para os SNPs é importante especialmente quando o estudo apresenta características influenciadas por poucos QTLs, sendo, assim, o ssGBLUP tradicional eficiente para a maioria das características poligênicas de interesse no melhoramento.

2.5. Abordagem para detecção de QTLs em GWAS

A identificação de SNPs significativamente associados a características importantes para o melhoramento animal de várias espécies, se faz por meio de testes de hipóteses clássicos, como o EMMA (*efficient mixed-model association*), pelo cálculo de p -valores, da proporção de variância genética explicada pelos SNPs ou uso de janelas fixas, móveis ou blocos de SNPs, formados com base no desequilíbrio de ligação, por exemplo (CHEN; STEIBEL; TEMPELMAN, 2017; WANG et al. 2014). No entanto, muitas dessas descobertas podem ser associações falso-positivas ou falso-negativas, consideradas associações espúrias (LI et al., 2021). Além disso, geralmente, não se testam a possibilidade dos resultados das associações serem casuais, por meio de p -valores ou de probabilidades e isso pode representar um problema (AGUILAR et al., 2019). Logo, os resultados de GWAS devem ser minuciosamente analisados, antes de se assumir uma possível associação como efeito causal ou significativa (FRAGOMENI et al., 2014).

Em estudos de GWAS, comumente utiliza-se, como forma de verificar a solução de hipóteses nulas, a declaração de um limiar de significância que explique a variância genética dos marcadores próximos. Além disso, pode-se utilizar esquemas por meio de iterações e da escolha por janelas entre os marcadores, o que funciona bem, porém de forma arbitrária e, conseqüentemente, de mais complexa interpretação e comparação entre os estudos (AGUILAR et al., 2019).

Os SNPs dentro de um mesmo segmento podem ser altamente correlacionados, dificultando a identificação dos efeitos individuais de cada marcador, pois estes podem estar atuando conjuntamente sobre a característica. O uso de janelas genômicas para testar as significâncias detectadas em GWAS surgiu como uma alternativa (LI et al., 2021). Nesta abordagem, por exemplo, o efeito detectado pode ser constante na população ou gerações ou, caso contrário, não poderá ser extrapolado para populações ou amostras externas por ser apenas verdadeiro para aquela amostra, assim sendo considerados resultados não confiáveis, devendo assim serem experimentados em populações com estruturas semelhantes (FRAGOMENI et al., 2014).

A determinação do correto limiar do p -valor para a significância de associações genômicas representa uma das principais opções para minimizarem a detecção de falsos positivos e falsos negativos em análises de GWAS. Para determinar o limiar de significância estatística em GWAS, vários métodos e medidas têm sido utilizados, dentre os quais se destacam a correção de *Bonferroni* e a taxa de falsa descoberta (*false discovery rate*, FDR), sendo que esta última medida controla a proporção esperada de falsos

positivos entre as hipóteses nulas rejeitadas e é menos conservativa do que a correção de *Bonferroni* (KALER; PURCELL, 2019).

Aguilar et al. (2019) propuseram uma metodologia de *p*-valores frequentistas implementada na estrutura do método ssGWAS que pode ser utilizado em estudos com características ou modelos complexos, bem como para grandes tamanhos populacionais. A obtenção das estimativas dos marcadores e valores genéticos genômicos (com *p*-valores associados) pode ser realizada através da inversa das equações de modelos mistos e o *p*-valor do marcador é obtido em um único passo de GWAS.

De modo geral, as variantes testadas em um determinado estudo dependem de fatores específicos de cada população avaliada, como o padrão de desequilíbrio de ligação e frequência do alelo menos frequente. Portanto, o limiar adequado para a significância em GWAS pode variar para diferentes populações e espécies (KALER; PURCELL, 2019).

3. REFERÊNCIAS

AGUILAR, I. et al. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. **Journal of Dairy Science**, v. 93, n. 2, p. 743–752, 2010.

AGUILAR, I. et al. Frequentist *p*-values for large-scale-single step genome-wide association, with an application to birth weight in American Angus cattle. **Genetics Selection Evolution**, v. 51, n. 28, 2019.

ARAUJO, A. et al. A Comprehensive Comparison of Haplotype-Based Single-Step Genomic Predictions in Livestock Populations With Different Genetic Diversity Levels: A Simulation Study. **Frontiers in Genetics**, v. 12:729867, 2021.

AUVRAY, B. et al. Genomic prediction of breeding values in the New Zealand sheep industry using a 50K SNP chip. **Journal of Animal Science**, v. 92, n. 10, p. 4375–4389, 2014.

B+LNZ GENETICS, **Research and development**. 2017, 56 p. Disponível em: https://www.blzgenetics.com/files/1507838033_BLG-RD-2017-Web.pdf>. Acesso em: 17 de novembro de 2019.

BLASCO, A.; TORO, M. A. A short critical history of the application of genomics to animal breeding. **Livestock Science**, v. 166, p. 4-9, 2014.

BRITO, L. F. et al. Prediction of genomic breeding values for growth, carcass and meat quality traits in a multi-breed sheep population using a HD SNP chip. **BMC Genetics**, v. 18, n. 7, 2017.

- CHEN, C. Y. et al. Effect of different genomic relationship matrices on accuracy and scale. **Journal of Animal Science**, v. 89, n. 9, p. 2673–2679, 2011.
- CHEN, C.; STEIBEL, J. P.; TEMPELMAN, R. J. Genome-Wide Association Analyses Based on Broadly Different Specifications for Prior Distributions, Genomic Windows, and Estimation Methods. **Genetics**, v. 206, n. 4, p. 1791–1806, 2017.
- CHRISTENSEN, O. F.; LUND, M. S. Genomic prediction when some animals are not genotyped. **Genetics Selection Evolution**, v. 42, n. 2, p. 1-8, 2010.
- DAETWYLER, H. D. et al. Accuracy of pedigree and genomic predictions of carcass and novel meat quality traits in multi-breed sheep data assessed by cross-validation. **Genetics Selection Evolution**, v. 44, n. 33, p. 1-11, 2012.
- DAETWYLER, H. D. et al. Genomic prediction in animals and plants: simulation of data, validation, reporting, and benchmarking. **Genetics**, v. 193, n. 2, p. 347–365, 2013.
- DEKKERS, J. C. M. Application of Genomics Tools to Animal Breeding. **Current Genomics**, v. 13, p. 207-212, 2012.
- FRAGOMENI, Bde. et al. Changes in variance explained by top SNP windows over generations for three traits in broiler chicken. **Frontiers in Genetics**, v. 5, n. 332, 2014.
- FERNANDES JÚNIOR, G. A. et al. Genome scan for postmortem carcass traits in Nellore cattle. **Journal of Animal Science**, v. 94, n. 10, p. 4087–4095, 2016.
- FORNI, S.; AGUILAR, I.; MISZTAL, I. Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. **Genetics Selection Evolution**, v. 43, n. 1, 2011.
- GORDO, D. G. M. et al. Genetic parameter estimates for carcass traits and visual scores including or not genomic information. **Journal of Animal Science**, v. 94, p. 1821–1826, 2016.
- HAY, E. H.; ROBERTS, A. Genome-wide association study for carcass traits in a composite beef cattle breed. **Livestock Science**, v. 213, p. 35-43, 2018.
- HÉRAULT, F. et al. Combined GWAS and LDLA approaches to improve genome-wide quantitative trait loci detection affecting carcass and meat quality traits in pig. **Meat Science**, v. 135, p. 148-158, 2018.
- HERNANDEZ, D. G. et al. Analysis of single nucleotide polymorphisms variation associated with important economic and computed tomography measured traits in Texel sheep. **Animal**, v. 12, p. 915-922, 2018.
- HICKEY, J. M., GORJANC, G. Simulated Data for Genomic Selection and Genome-Wide Association Studies Using a Combination of Coalescent and Gene Drop Methods. **G3 Genes, Genomics, Genetics**, v. 2, 2012.
- HIDALGO, J. Changes in genetic parameters for fitness and growth traits in pigs under genomic selection. **Journal of Animal Science**, v. 98, n. 2, p. 1 - 12, 2020.
- HU, Z. L.; PARK, C. A.; REECY, J. M. Building a livestock genetic and genomic information knowledgebase through integrative developments of Animal QTLdb and CorrDB. **Nucleic Acids Research**, v. 47, D701–D710, 2019.

- IBTISHAM F. et al. Genomic selection and its application in animal breeding. **Thai Journal of Veterinary Medicine**, v. 47, p. 301–310, 2017.
- IQBAL, A. et al. Genome-wide Association Study to Identify Quantitative Trait Loci for Meat and Carcass Quality Traits in Berkshire. **Asian-Australasian Journal of Animal Science**, v. 28, n. 11, p. 1537-1544, 2015.
- KALER, A. S.; PURCELL, L. C. Estimation of a significance threshold for genome-wide association studies. **BMC Genomics**, v. 20, n. 618, 2019.
- LEGARRA, A.; AGUILAR, I.; MISZTAL, I. A relationship matrix including full pedigree and genomic information. **Journal of Dairy Science**, v. 92, n. 9, p. 4656-4663, 2009.
- LEGARRA, A.; AGUILAR, I.; MISZTAL, I. A relationship matrix including full pedigree and genomic information. **Journal of Dairy Science**, v. 92, p. 4656-4663, 2009.
- LEGARRA, A. et al. **Single step, a general approach for genomic selection.** *Livestock Science*, v. 166, p. 54-65, 2014.
- LEMOS, M. V. et al. Genome-wide association between single nucleotide polymorphisms with beef fatty acid profile in Nellore cattle using the single step procedure. **BMC Genomics**, v. 17, n. 213, p. 1-16, 2016.
- LI, H. et al. An efficient unified model for genome-wide association studies and genomic selection. **Genetics Selection Evolution**, v. 49, n. 64, p. 1-8, 2017.
- LI, J. et al. Tests of association based on genomic windows can lead to spurious associations when using genotype panels with heterogeneous SNP densities. **Genetics Selection Evolution**, v. 53, n. 45, 2021.
- LIU, R. et al. Genome-wide association study identifies Loci and candidate genes for body composition and meat quality traits in Beijing-You chickens. **PLoS One**, v. 8, n. 4, e61172, 2013.
- LIU, T. et al. Accuracy of genomic prediction for growth and carcass traits in Chinese triple-yellow chickens. **BMC Genetics**, v. 15, n. 110, p. 1-8, 2014.
- LOURENCO, D. A. et al. Are evaluations on young genotyped animals benefiting from the past generations? **Journal of Dairy Science**, v. 97, n. 6, p. 3930–3942, 2014.
- LOURENCO, D. et al. Single-Step Genomic Evaluations from Theory to Practice: Using SNP Chips and Sequence Data in BLUPF90. **Genes**, v. 11, n. 7:790, 2020.
- MEHRBAN, H. et al. Comparison of conventional BLUP and single-step genomic BLUP evaluations for yearling weight and carcass traits in Hanwoo beef cattle using single trait and multi-trait models. **PLoS One**, v. 14, n. 10, 2019.
- MEUWISSEN, T. et al. On the distance of genetic relationships and the accuracy of genomic prediction in pig breeding. **Genetics Selection Evolution**, v. 46, n. 49, 2014.
- MEUWISSEN, T.; HAYES, B. J.; GODDARD M. E. Prediction of total genetic values using genome-wide dense marker maps. **Genetics**, v. 157, p. 1819-1829, 2001.
- MEUWISSEN, T.; HAYES, B. J.; GODDARD, M. E. Genomic selection: A paradigm shift in animal breeding. **Animal Frontiers**, v.6, n.1, p. 6-14, 2016.

- MISZTAL, I.; LEGARRA, A.; AGUILAR, I. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. **Journal of Dairy Science**, v. 92, p. 4648–4655, 2009.
- MISZTAL, I. Is genomic selection now a mature technology? **Journal of Animal Breeding and Genetics**, v. 133, p. 81–82, 2016.
- MISZTAL, I.; LOURENCO, D.; LEGARRA, A. Current status of genomic evaluation. **Journal of Animal Science**, v. 98, n. 4, p. 1–14, 2020.
- MRODE, R. et al. Genomic selection and use of molecular tools in breeding programs for indigenous and crossbred cattle in developing countries: current status and future prospects. **Frontiers in Genetics**, v. 9, n. 694, 2019.
- MRODE, R. et al. Invited review: Genomic selection for small ruminants in developed countries: how applicable for the rest of the world? **Animal**, v. 12, n. 7, p. 1333-1340, 2018.
- NAVAJAS, E.A. Animal breeding and genetics: DNA Markers and Marker-Assisted Selection in the Genomic Era. **Encyclopedia of Meat Sciences**, v.1, n.2, p.12-18, 2014.
- ONOGI, A. et al. Genomic prediction in Japanese black cattle: Application of a single-step approach to beef cattle. **Journal of Animal Science**, v. 92, p. 1931-1938, 2014.
- PÉRTILE, S. F. N. et al. Seleção e associação genômica ampla para o melhoramento genético animal com uso do método ssGBLUP. **Pesquisa Agropecuária Brasileira**, v.51, n.10, p.1729-736, 2016.
- PICCOLI, M. L. et al. A comprehensive comparison between single- and two-step GBLUP methods in a simulated beef cattle population. **Canadian Journal of Animal Science**, v. 98, n. 565–575, 2018.
- ROVADOSCKI, G. A. et al. Estimates of genomic heritability and genome-wide association study for fatty acids profile in Santa Inês sheep. **BMC Genomics**, v. 19, n. 375, p. 1-14, 2018.
- RUPP, R. et al. Genomic application in sheep and goat breeding. **Animal Frontiers**, v. 6, n. 1, p. 39-44, 2016.
- SAMORÈ, A. B.; FONTANESI, L. Genomic selection in pigs: state of the art and perspectives. **Italian Journal of Animal Science**, v. 15, n. 2, p. 211-232, 2016.
- SANTANA, M. H. A. et al. A genome wide association mapping study using ultrasound-scanned information identifies potential genomic regions and candidate genes affecting carcass traits in Nellore cattle. **Journal of Animal Breeding and Genetics**, v. 132, p. 420-427, 2015.
- SARMENTO, J.L.R.; SENA, L.S. Avanços no melhoramento genético na América Latina e perspectivas futuras. **Archivos Latinoamericanos de Produccion Animal**, v.25, n. 1-2, p.81-89, 2017.
- SHARMA, A. et al. Stories and Challenges of Genome Wide Association Studies in Livestock — A Review. **Asian Australasian Journal of Animal Science**, v. 28, n. 10, p. 1371-1379, 2015.

- SILVA, R. M. O. et al. Genome-Wide Association Study for Carcass Traits in an Experimental Nelore Cattle Population. **PloS one**, v. 12, n. 1, e0169860, 2017.
- SILVA, R. P. et al. Genomic regions and enrichment analyses associated with carcass composition indicator traits in Nellore cattle. **Journal of Animal Breeding and Genetics**, v. 136, p. 118-133, 2019.
- SIMIANER, H. Genomic and other revolutions – why some technologies are quickly adopted and others are not. **Animal Frontiers**, v. 6, n. 1, p. 53-58, 2016.
- SONG, H. et al. Using different single-step strategies to improve the efficiency of genomic prediction on body measurement traits in pig. **Frontiers in Genetics**, v. 9, n. 730, p. 1-10, 2019.
- SUN, Y. et al. The identification of 14 new genes for meat quality traits in chicken using a genome-wide association study. **BMC Genomics**, v. 14, n. 458, 2013.
- SWAN, A. A. et al. Genomic evaluations in the Australian sheep industry. **In: World Congress on Genetics Applied to Livestock Production**, 10, 2014, Vancouver. Proceedings... 2014.
- TAKEDA, M.; UEMOTO, Y.; SATOH, M. Effect of genotyped bulls with different numbers of phenotyped progenies on quantitative trait loci detection and genomic evaluation in a simulated cattle population. **Animal Science Journal**, v. 91, n. 1, e13432, 2020.
- TOPIGS NORSVIN. **We. Annual Magazine 2017**. Disponível em: <<http://noticias.porcisan.es/wp-content/uploads/WeMag17-Topigs-Norsvin-LR.pdf>>. Acesso em: 11 de março de 2020.
- VAN DEN BERG, I.; FRITZ, S.; BOICHARD, D. QTL fine mapping with Bayes C(π): a simulation study. **Genetics Selection Evolution**, v. 45, n.19, 2013.
- VANRADEN, P. M. Efficient Methods to Compute Genomic Predictions. **Journal of Dairy Science**. v. 91, p. 4414–4423, 2008.
- VITEZICA, Z. G. Bias in genomic predictions for populations under selection. **Genetics Research (Camb)**, v. 93, n. 5, p. 357–366, 2011.
- WANG, H. et al. Genome-wide association mapping including phenotypes from relatives without genotypes. **Genetics Research (Cambridge)**, v. 94, n. 2, p. 73-83, 2012.
- WANG, H. et al. Genome-wide association mapping including phenotypes from relatives without genotypes in a single-step (ssGWAS) for 6-week body weight in broiler chickens. **Frontiers in Genetics**, v. 5, n. 134, 2014.
- WANG, H. et al. Genome-Wide Specific Selection in Three Domestic Sheep Breeds. **PLoS One**, v. 10, n. 6, e0128688, 2015.
- WOLC, A. et al. Implementation of genomic selection in the poultry industry. **Animal Frontiers**, v. 6, n. 1, p. 23-31, 2016.

XU, S. S. et al. Genome-wide association analysis identifies the genetic basis of fat deposition in the tails of sheep (*Ovis aries*). **Animal Genetics**, v. 48, n. 5, p. 560-569, 2017.

ZHANG, T. et al. Genome-wide association study of meat quality traits in chicken. **Genetic Molecular Research**, v. 14, n. 3, p. 10452–10460, 2015.

ZHANG, X. et al. Weighting strategies for single-step genomic BLUP: an iterative approach for accurate calculation of GEBV and GWAS. **Frontiers in Genetics**, v. 7, n. 151, p. 1-14, 2016.

CAPÍTULO I

Associação genômica ampla em populações simuladas de ovinos

Associação genômica ampla em populações simuladas de ovinos

BARBOSA, Bruna Lima. **Associação genômica ampla em populações simuladas de ovinos**. 2020. Tese (Doutorado em Ciência Animal) – Universidade Federal do Piauí, Teresina, Piauí, 2020.

RESUMO: O objetivo com esta pesquisa foi verificar a eficiência na detecção de QTLs, usando o método *ssGWAS* com a abordagem do *p*-valor exato, no genoma simulado de ovinos, em cenários com tamanhos populacionais e número de animais genotipados distintos (15.000, 7.500, 5.000, 2.500, 2.000, 1.000, 500 e 250), quantidade de QTLs (2.014 e 114) e utilizando painéis de SNPs de diferentes densidades (600K, 50K e 12K). As informações de pedigree, genótipo e fenótipo foram geradas com base na herdabilidade e variância fenotípica de dados reais de AOL. Um painel de alta densidade foi simulado e outros dois, de média e baixa densidade, foram obtidos a partir deste, sendo todos utilizados nas análises posteriores. Foram avaliados 48 cenários propostos de acordo com o painel de SNPs, quantidade de QTLs e número de animais com informações genotípicas disponíveis. Os limites de significância dos efeitos dos SNPs foram obtidos pela correção de *Bonferroni* e taxa de falsa descoberta numa escala $-\log_{10}$ do *p*-valor (5%). As posições dos SNPs associados foram comparadas com a dos QTLs que determinavam a característica. Foi possível identificar vários marcadores próximos a QTLs e, em alguns casos, com alta associação entre eles. Verificou-se que quanto maior a quantidade de animais genotipados e SNPs no painel, maior a possibilidade de detecção de associações entre os SNPs significativos e os QTLs simulados. Com 2.014 QTLs, os cenários com o painel de alta densidade proporcionaram maior quantidade de detecções. Com 114 QTLs, geralmente, observou-se pelo menos uma associação, com picos maiores nos cenários com maior quantidade de animais. Os cenários com 15 mil e 7.500 genótipos com as duas quantidades de QTLs simulados, especialmente no painel de alta densidade, se destacaram na obtenção de associações entre SNPs significativos e QTLs. Portanto, a abordagem do *p*-valor exato possibilitou a detecção de pelo menos uma associação entre SNPs e QTLs simulados.

Palavras-chave: *ssGWAS*, *p*-valor exato, dados simulados, detecção de QTLs.

Genome-wide association in simulated sheep populations

BARBOSA, Bruna Lima. **Genome-wide association in simulated population of sheep.** 2020. Thesis (Doctor of Philosophy in Animal Science) – Universidade Federal do Piauí, Teresina, 2020.

ABSTRACT: The objective of this study was to verify the efficiency in the detection of QTLs using the ssGWAS method with the exact p-value approach, in the simulated sheep genome, in scenarios with different population sizes and numbers of genotyped animals (15,000, 7,500, 5,000, 2,500, 2,000, 1,000, 500 and 250), number of QTLs (2,014 and 114) and using panels of SNPs of different densities (600K, 50K and 12K). Pedigree, genotype, and phenotype data were generated based on estimates of heritability and phenotypic variance of real data of LEA. A high-density panel was simulated and two others, of medium and low density, were obtained from it, all of which were used in the subsequent analyses. 48 scenarios proposed based on SNP panel, number of QTLs, and number of animals with genotypic information were used. The significance of SNP effects was obtained using the Bonferroni correction and the false discovery rate in a -log₁₀ scale of the p-value (5%). The positions of estimated SNPs were compared to those of simulated QTLs. Therefore, it was possible to identify several markers near QTLs and, in some cases, with high association among them. It was observed that a larger higher number of genotyped animals and SNPs in the panel result in a higher probability of detection of associations between significant SNPs and simulated QTLs. The scenarios in which the high-density panel and 2,014 QTLs were used allowed a higher number of true detections. Overall, in the scenarios where 114 QTLs were used, at least one true association was observed, with higher peaks in scenarios that used larger number of animals. The scenarios that used 15,000 and 7,500 genotypes with the two numbers of simulated QTLs, specially the high density one, stood out in the obtention of associations between significant SNPs and QTLs. Thus, the exact p-value approach allowed the detection of at least one association between SNPs and simulated QTLs.

Keywords: ssGWAS, exact *p*-valor, simulated data, detection of QTLs.

1. Introdução

O advento das informações moleculares através de painéis de SNPs tornaram possível a detecção de QTL, estimação dos efeitos de marcadores e o melhor entendimento sobre a arquitetura genética de características de importância econômica, por meio de estudos de associação genômica ampla (GWAS) (DEKKERS, 2012). A compreensão da estrutura genética dos indivíduos é baseada na expressão de genes que estejam associados a características produtivas, sendo os SNPs os mais utilizados devido a eficiência na identificação de QTLs que afetam o fenótipo (WANG et al., 2012; LI et al., 2017).

As análises realizadas em GWAS buscam a associação de *locus* de características quantitativas e qualitativas associadas com características complexas e poligênicas (MEUWISSEN; HAYES; GODDARD, 2001). Vários fatores são importantes parâmetros para tornar mais provável a detecção de um QTL por meio de GWAS, dentre eles o conhecimento da estrutura da característica avaliada, o desequilíbrio de ligação entre os marcadores e o QTL, informações de fenótipos e pedigree, herdabilidade da característica, entre outros (VAN DEN BERG et al., 2013).

Os procedimentos em passo único, como o *ssGWAS* (*single-step GWAS*), têm mostrado eficiência na detecção de *loci* associados a características quantitativas. Aguilar et al. (2019) propuseram o método *ssGWAS* com *p*-valores exatos (*exact single-step GWAS*), apresentando ser uma estratégia eficiente de detecção de QTLs, por lidar com problemas como a falta de um limiar específico para a determinação de regiões significativamente associadas com os fenótipos de interesse.

A utilização de simulação de dados representa uma das principais alternativas economicamente viáveis para gerar dados genômicos para realizar pesquisas de GWAS. Portanto, o objetivo com esta pesquisa foi verificar a eficiência na detecção de QTLs, usando o método *ssGWAS* com a abordagem do *p*-valor exato, em cenários com diferentes tamanhos populacionais, número de QTLs que determinam a característica e número de indivíduos genotipados.

2. Material e métodos

2.1. Simulações

O presente estudo não necessitou de aprovação da Comissão de Ética no Uso de Animais (CEUA/UFPI), da Universidade Federal do Piauí devido ao fato dos dados serem gerados computacionalmente através de simulações.

2.1.1 Estrutura da população

As simulações foram realizadas com o *software* QMSim versão 1.10 (SARGOLZAEI; SCHENKEL, 2009), sendo conduzidas em três passos: O primeiro consistiu na simulação da população histórica, composta por 1.500 animais, em iguais proporções entre os sexos, quantidade essa que se manteve constante até a geração 1.000. Após as 20 gerações seguintes, esse valor decaiu para 600 animais, também em iguais proporções entre os sexos. Esta etapa foi realizada com o intuito de gerar o desequilíbrio de ligação na população histórica a partir do efeito de mutação-deriva nas mesmas.

No segundo passo, a formação da população fundadora, foram selecionados os 600 indivíduos da última geração da população histórica, com igual proporção entre machos e fêmeas (300 animais de cada sexo), os quais foram acasalados segundo os critérios: 3 progênies por fêmea por geração, 8 gerações de acasalamento aleatório com união aleatória de gametas, crescimento exponencial das fêmeas e sem descarte de animais. O processo de expansão da população, a terceira etapa, foi realizado a partir de animais da população fundadora, sendo 400 machos e 10.000 fêmeas, da última geração da população anterior. Foram consideradas duas progênies por fêmea e 8 gerações de acasalamento aleatório. A taxa de reposição de machos e fêmeas foi de 60% e 20%, respectivamente. Os animais foram selecionados e descartados com base no fenótipo. Ao final desse processo, 170.010 indivíduos foram obtidos, dos quais 60.000 animais, provenientes das três últimas gerações, foram selecionados através dos maiores fenótipos para a formação dos cenários que foram avaliados.

2.1.2 Genoma

Um painel de alta densidade contendo 576.595 SNPs foi simulado, de acordo com informações disponíveis no programa *SNPchiMp* versão 3 (NICOLAZZI et al. 2015) e, a partir dos arquivos gerados, foram formados dois painéis de densidades distintas, um de média (50K) e um de baixa densidade (12K) contendo cada um, 52.326 e 11.385 SNPs,

respectivamente. As informações sobre número de SNPs por cromossomo foram baseadas nos trabalhos de Bolormaa et al. (2015) para o painel de baixa densidade e Alvarenga et al. (2018) para o painel de média densidade. A variância fenotípica de 5,61 e herdabilidade de 0,30 foram considerados como parâmetros globais na simulação, sendo obtidas no trabalho de Sena (2020), no intuito de simular uma característica com a arquitetura genética semelhante a AOL. A herdabilidade dos QTLs utilizada foi de 0,18 (60% da herdabilidade de AOL), assumindo que os efeitos dos QTLs têm contribuição expressiva na variância fenotípica da característica.

O genoma foi constituído de 26 cromossomos autossômicos com um tamanho total de 2.656 centiMorgans, baseado em informações do genoma ovino contidas na plataforma *NCBI* (<https://www.ncbi.nlm.nih.gov/genome/?term=Ovis+aries>), *Assembly Oar_rambouillet* versão 1.0. Foi considerada a presença de 114 e 2.014 QTLs, em cenários diferentes, obtidos a partir de informações da plataforma *SheepQTLdb*, (<https://www.animalgenome.org/cgi-bin/QTLdb/OA/index>), onde constavam 12 QTLs reais referentes a AOL, localizados nos cromossomos 2, 6, 7, 9 e 18. Considerou-se a existência de pelo menos um QTL nos cromossomos que não continham essa informação na plataforma, sendo assim, distribuídos de forma proporcional e gerados por simulações diferentes, com efeito de distribuição gama e parâmetro *Shape* 0,4. A taxa de mutação recorrente para marcadores e QTLs foi de 1×10^{-9} . A taxa de genótipos faltantes foi de 1×10^{-2} e de erros de genotipagem de 5×10^{-3} . Os parâmetros utilizados nas simulações para obtenção dos dados foram apresentados no ANEXO I.

2.1.3 Cenários de estudo

Os indivíduos obtidos nas três últimas gerações de acasalamento aleatório das simulações foram selecionados para compor cada cenário proposto neste estudo. De acordo com as informações genotípicas, fenotípicas e de pedigree, os animais foram agrupados em 48 cenários, sendo oito destes relacionados a cada um dos painéis utilizados e as duas quantidades de QTLs.

A partir de 15 mil animais selecionados aleatoriamente das três últimas gerações com o painel simulado de 600K, selecionou-se, primeiramente, os marcadores que formaram o painel de média densidade e, a partir deste, o de baixa densidade, tomando como base o arquivo de marcadores e a posição no cromossomo aos quais estão localizados (arquivo de mapa) por meio de uma seleção aleatória. Posteriormente, os 15 mil animais do painel de alta densidade funcionaram como base para a formação dos

cenários com animais que continham todas as informações de genótipo, fenótipo e pedigree, sendo constituídos de metade dos genótipos dos ovinos e todas as informações de fenótipo e pedigree, contendo 500, 2.000, 5.000 e 15.000 animais. De cada um destes, formou-se outros cenários contendo os mesmos números de animais com fenótipo, porém com metade do número de animais genotipados, constando de 250, 1.000, 2.500 e 7.500 animais, conforme apresentado na Tabela 1.

Tabela 1. Cenários para análises de Associação Genômica Ampla com dados simulados para área de olho de lombo em *Ovis aries*

Painel	Cenário	Quantidade de QTL	Nº de animais (genótipos e fenótipos)
600k	1	2.014	15000 ^a
	2		7500 ^b
	3		5000 ^a
	4		2500 ^b
	5		2000 ^a
	6		1000 ^b
	7		500 ^a
	8		250 ^b
	9	114	15000 ^a
	10		7500 ^b
	11		5000 ^a
	12		2500 ^b
	13		2000 ^a
	14		1000 ^b
	15		500 ^a
	16		250 ^b
50k	17	2.014	15000 ^a
	18		7500 ^b
	19		5000 ^a
	20		2500 ^b
	21		2000 ^a
	22		1000 ^b
	23		500 ^a
	24		250 ^b
	25	114	15000 ^a
	26		7500 ^b
	27		5000 ^a
	28		2500 ^b
	29		2000 ^a
	30		1000 ^b
	31		500 ^a
	32		250 ^b
12k	33	2.014	15000 ^a
	34		7500 ^b
	35		5000 ^a
	36		2500 ^b
	37		2000 ^a
	38		1000 ^b
	39		500 ^a
	40		250 ^b
	41	114	15000 ^a
	42		7500 ^b
	43		5000 ^a
	44		2500 ^b
	45		2000 ^a
	46		1000 ^b
	47		500 ^a
	48		250 ^b

^a Todos os animais contém informações de genótipos, fenótipo e pedigree. ^b Mesma quantidade de animais com informações de fenótipo e pedigree, do cenário imediatamente anterior, porém com metade dessa quantidade de animais genotipados.

2.2 Análises estatísticas

Os bancos de dados gerados com 10 repetições das simulações foram editados com auxílio do *software* R versão 3.6.1 (R CORE TEAM, 2019) para utilização nos programas da família BLUPf90 (MISZTAL et al. 2019). Os animais das três últimas gerações da população recente (15.000 animais de ambos os sexos) foram selecionados para as análises de GWAS, de acordo com cada um dos cenários propostos. O controle de qualidade dos genótipos foi feito com base nos critérios *Call rate* (<95%), *call rate* por animal (<0,90), Frequência do alelo menos frequente (MAF <0,01) e desvios de equilíbrio de *Hardy-Weinberg* ($p < 10^{-15}$), utilizando o *software* PreGSf90 (MISZTAL et al., 2019). Não houve exclusão de dados por meio destes critérios de controle de qualidade.

O *software* Airemlf90 (MISZTAL et al., 2019) foi utilizado, com as informações dos marcadores, fenótipos e pedigree, para obter os componentes de variância na população, no intuito de verificar se a herdabilidade na população avaliada estava próximo do desejado (aproximadamente 0,30).

Foi utilizado o método GWAS em único passo (*single-step* GWAS), (WANG et al., 2012), para realização das estimativas dos efeitos dos marcadores sobre a característica avaliada. Essa metodologia baseia-se no uso da matriz de parentesco \mathbf{H} , resultante da combinação da matriz \mathbf{A} (parentesco via pedigree) e matriz \mathbf{G} (parentesco genômico).

No modelo animal unicaracterístico foram considerados como efeitos fixos o sexo dos animais, com dois níveis, e geração, com três níveis, e como efeitos aleatórios, os animais e o erro aleatório, com níveis variando de acordo com o cenário. Para as estimativas de componentes de (co) variância e valores genéticos foi utilizado o modelo:

$$y = \mathbf{X}\beta + \mathbf{Z}\alpha + \mathbf{e}$$

em que: y é o vetor de fenótipos; β é o vetor de efeitos fixos; α é o vetor aleatório de efeitos genéticos aditivos, \mathbf{X} é a matriz de incidência dos efeitos fixos; \mathbf{Z} é a matriz de incidência dos efeitos genéticos aditivos; \mathbf{e} é o vetor aleatório de efeitos residuais, $e \sim N(0, \mathbf{I}\sigma_e^2)$. A inversa da matriz \mathbf{H} foi calculada de acordo com Aguilar et al. (2010), como:

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

Sendo, \mathbf{A}^{-1} a inversa da matriz de parentesco com base em informações de pedigree, \mathbf{G}^{-1} a inversa da matriz de parentesco genômico e \mathbf{A}_{22}^{-1} a inversa da matriz de parentesco baseada em pedigree para os animais genotipados.

2.2.1 Análises de Associação genômica em passo único com p-valor exato (ssGWAS de p-valor exato)

As análises foram realizadas nos programas BLUPf90 e POSTGSF90 com a utilização da função "snp_p_value" (MISZTAL et al., 2019), para obter os GEBVs e calcular os efeitos dos marcadores, respectivamente.

De acordo com método proposto por Aguilar et al. (2019), os p-valores para cada i-ésimo marcador ($pval_i$) foram obtidos por meio de:

$$pval_i = 2 \left(1 - \Phi \left(\left| \frac{\hat{a}_i}{sd(\hat{a}_i)} \right| \right) \right)$$

Sendo Φ a função normal padrão cumulativa, \hat{a}_i o efeito dos marcadores e sd o desvio padrão associado ao efeito dos marcadores.

Os resultados das GWAS foram plotados em gráficos do tipo *Manhattan plot* com auxílio do pacote *qqman* (TURNER, 2014) do *software* R e o valor de significância para o efeito dos SNPs em relação a característica avaliada foi obtido pelas correções para testes múltiplos de *Bonferroni* (DUNN, 1961) e FDR (*False Discovery Rate*) (BENJAMINI; HOCKBERG, 1995), por cromossomo, numa escala $-\log_{10}$ considerando uma probabilidade de 5% (α), sendo para o teste de *Bonferroni*:

$$pvalor = \frac{0,05}{k}$$

em que k é o número total de marcadores (testes estatísticos).

$$pvalor = \frac{pR \times 0,05}{k}$$

em que pR indica o *rank* do p-valor do marcador, para o teste FDR.

O FDR foi executado após a realização da correção de *Bonferroni*, sendo utilizado o número de marcadores que passaram neste teste para o cálculo da taxa de falsos positivos pelo FDR.

As distribuições de p -valores obtidos pela metodologia de Aguilar et al. (2019) foram representadas em *Manhattan plots* os quais foram comparados quanto a eficiência de detecção dos QTLs nos diferentes cenários de estudo de acordo com as significâncias encontradas a partir dos limiares estabelecidos.

3. Resultados

3.1 Estimativas dos componentes de variância e herdabilidade

A média das estimativas de componentes de variância e herdabilidade das 10 repetições simuladas com 2.014 QTLs e três painéis de densidades alta, média e baixa para populações com tamanhos e número de animais genotipados distintos estão apresentadas na Tabela 2. Os cenários simulados com o painel de alta densidade apresentaram estimativas de herdabilidade que variaram de 0,21 a 0,27, enquanto com uso dos painéis de média densidade essas estimativas variaram de 0,21 a 0,26 e foram obtidos valores de 0,17 a 0,25 ao se utilizar os painéis de baixa densidade. Portanto, quanto mais SNPs, mais os cenários se aproximaram da herdabilidade simulada (0,30).

As variâncias genéticas aditivas estimadas foram diferentes das simuladas, porém dentro do esperado (valor simulado de 1,68), com valores de 1,15 a 1,48 com uso do painel de alta densidade, 1,17 a 1,54 com o painel de densidade média, e 0,96 a 1,37 com o de baixa densidade. Já as variâncias residuais variaram de 3,95 a 4,39, 4,27 a 4,45 e de 4,33 a 4,59 para os cenários com as três densidades de SNPs, na mesma ordem mencionada anteriormente. Note que os valores iniciais simulados para a variância genética aditiva e residual foram de 1,68 e 3,93, respectivamente.

As estimativas de herdabilidade dos cenários simulados com 114 QTLs variaram de 0,21 a 0,35, de 0,21 a 0,29 e de 0,16 a 0,35 para os painéis de 600k, 50k e 12k, respectivamente, maiores que às herdabilidades referentes aos cenários com maior quantidade de QTLs (Tabela 3). As variâncias genéticas aditivas obtidas com uso dos diferentes painéis variaram de 1,21 a 1,97 (600k), de 1,16 a 1,60 (50k) e de 0,92 a 1,98 (12k). As variâncias residuais estimadas foram de 3,66 a 4,41 para os cenários com o painel de alta densidade, de 3,93 a 4,46 para o painel de densidade moderada e de 3,61 a 4,68 para o painel de baixa densidade.

Tabela 2. Estimativas de componentes de variância, herdabilidades e erros padrão usando o ssGWAS com 2.014 QTLs para os cenários com três painéis de densidades alta, média e baixa e populações com tamanhos e número de animais genotipados distintos

C1			C2			C3			C4			C5			C6			C7			C8		
σ_a^2	σ_e^2	h^2	σ_a^2	σ_e^2	h^2	σ_a^2	σ_e^2	h^2	σ_a^2	σ_e^2	h^2	σ_a^2	σ_e^2	h^2	σ_a^2	σ_e^2	h^2	σ_a^2	σ_e^2	h^2	σ_a^2	σ_e^2	h^2
600k																							
1,31	4,28	0,23	1,32	4,27	0,23	1,29	4,36	0,23	1,34	4,26	0,24	1,18	4,39	0,21	1,15	4,40	0,21	1,48	3,95	0,27	1,48	3,98	0,26
(0,07)	(0,07)		(0,10)	(0,10)		(0,14)	(0,13)		(0,23)	(0,23)		(0,27)	(0,27)		(0,48)	(0,49)		(0,89)	(0,89)		(1,76)	(1,73)	
C17			C18			C19			C20			C21			C22			C23			C24		
50K																							
1,24	4,36	0,22	1,21	4,39	0,22	1,25	4,32	0,22	1,29	4,27	0,23	1,17	4,45	0,21	1,28	4,28	0,23	1,29	4,40	0,22	1,54	4,34	0,26
(0,07)	(0,06)		(0,10)	(0,10)		(0,13)	(0,13)		(0,22)	(0,22)		(0,26)	(0,27)		(0,48)	(0,48)		(0,90)	(0,90)		(1,79)	(1,79)	
C33			C34			C35			C36			C37			C38			C39			C40		
12k																							
1,02	4,57	0,18	1,02	4,56	0,18	1,04	4,54	0,19	0,96	4,59	0,17	1,09	4,54	0,19	1,16	4,45	0,21	1,13	4,56	0,20	1,37	4,33	0,25
(0,11)	(0,12)		(0,09)	(0,10)		(0,12)	(0,12)		(0,19)	(0,20)		(0,23)	(0,25)		(0,43)	(0,44)		(0,81)	(0,83)		(1,59)	(1,59)	

σ_a^2 : variância genética aditiva; σ_e^2 : variância residual; h^2 : herdabilidade. Valores médios de 10 repetições das simulações.

Tabela 3. Estimativas de componentes de variância, herdabilidades e erros padrão usando o ssGWAS com 114 QTLs para os cenários com três painéis de densidades alta, média e baixa e populações com tamanhos e número de animais genotipados distintos

C9			C10			C11			C12			C13			C14			C15			C16		
600k																							
σ_a^2	σ_e^2	h^2	σ_a^2	σ_e^2	h^2	σ_a^2	σ_e^2	h^2	σ_a^2	σ_e^2	h^2	σ_a^2	σ_e^2	h^2	σ_a^2	σ_e^2	h^2	σ_a^2	σ_e^2	h^2	σ_a^2	σ_e^2	h^2
1,24	4,35	0,22	1,21	4,41	0,21	1,28	4,36	0,23	1,30	4,34	0,23	1,24	4,34	0,22	1,41	4,19	0,25	1,70	3,73	0,31	1,97	3,66	0,35
(0,07)	(0,07)		(0,10)	(0,10)		(0,14)	(0,13)		(0,23)	(0,23)		(0,27)	(0,27)		(0,49)	(0,48)		(0,90)	(0,89)		(1,80)	(1,77)	
C25			C26			C27			C28			C29			C30			C31			C32		
50K																							
1,18	4,42	0,21	1,16	4,41	0,21	1,18	4,43	0,21	1,17	4,46	0,21	1,23	4,33	0,22	1,59	4,09	0,28	1,39	4,03	0,26	1,60	3,93	0,29
(0,07)	(0,06)		(0,10)	(0,10)		(0,13)	(0,13)		(0,22)	(0,22)		(0,26)	(0,26)		(0,49)	(0,48)		(0,88)	(0,88)		(1,71)	(1,81)	
C41			C42			C43			C44			C45			C46			C47			C48		
12k																							
0,97	4,62	0,21	0,97	4,62	0,17	0,96	4,60	0,17	0,92	4,68	0,16	0,99	4,58	0,18	1,08	4,48	0,19	1,69	3,79	0,31	1,98	3,61	0,35
(0,06)	(0,06)		(0,09)	(0,10)		(0,11)	(0,12)		(0,19)	(0,20)		(0,23)	(0,25)		(0,42)	(0,44)		(0,83)	(0,81)		(1,69)	(1,54)	

σ_a^2 : variância genética aditiva; σ_e^2 : variância residual; h^2 : herdabilidade. Valores médios de 10 repetições das simulações.

3.2 ssGWAS pela abordagem do *p*-valor exato

As figuras de 1, 3, 5, 7, 9, 11, 13 e 15 representam os *Manhattan plots* gerados para os cenários com diferentes tamanhos populacionais, painéis de alta, média e baixa densidade e duas quantidades de QTLs do genoma simulado de ovinos. A abordagem do *p*-valor exato para o cenário com 15 mil informações de genótipos e fenótipos apresentou maior quantidade de associações detectadas nas estruturas populacionais com 2.014 QTLs, por apresentarem mais SNPs significativos próximos a QTLs distribuídos ao longo de todo o cromossomo nove (Figura 1). Foram observadas associações em ambos os limiares de *Bonferroni* e FDR. Com a redução da densidade de SNPs referente ao painel, houve redução na quantidade de detecções, bem como nos picos de associações, de acordo com os *p*-valores.

Com relação aos cenários com 114 QTLs e a mesma quantidade de informações de genótipos e fenótipos (15 mil), para o mesmo cromossomo supracitado, observou-se maiores picos de SNPs significativos pelos dois limiares, especialmente entre 40Mb e 50Mb e entre 90Mb e 100Mb. Isto indica que os SNPs associados ao fenótipo se localizam nas regiões central e terminal do cromossomo nove, próximos a QTLs. Houve também redução na quantidade de associações detectadas, conforme a redução da densidade de SNPs. Foram identificados QTLs nas mesmas posições dos SNPs associados, o que indica forte associação entre os marcadores significativos e os QTLs simulados, assumidos como posições verdadeiras (Figura 1).

Os cenários com 15 mil informações de fenótipo e 7.500 genótipos foram representados pelo cromossomo dois. Picos significativos entre marcadores e QTLs foram observados no cenário com 600K e 2.104 QTLs, localizados da região central e porção terminal do cromossomo. Muitos QTLs foram verificados na mesma posição ou próximo, especialmente em relação aos picos com SNPs significativos nas regiões com 15Mb, 23Mb e 25Mb. Nos painéis de média e baixa densidade e 2.104 QTLs não foram observados SNPs significativos pelos limiares adotados. Com relação aos cenários com 114 QTLs, observou-se que o cenário referente ao painel de 600K apresentou os mesmos picos observados no cenário com o uso deste painel e 2.014 QTLs, porém com menor quantidade de QTLs nas proximidades ou na mesma posição dos picos (Figura 3).

Comparando os cenários com 15 mil genótipos e fenótipos àqueles que apresentaram metade das informações de genótipos, observou-se um padrão similar entre os cenários referentes a 2.014 QTLs, com SNPs associados significativamente a QTLs distribuídos por toda extensão do cromossomo (Figuras 1 e 3).

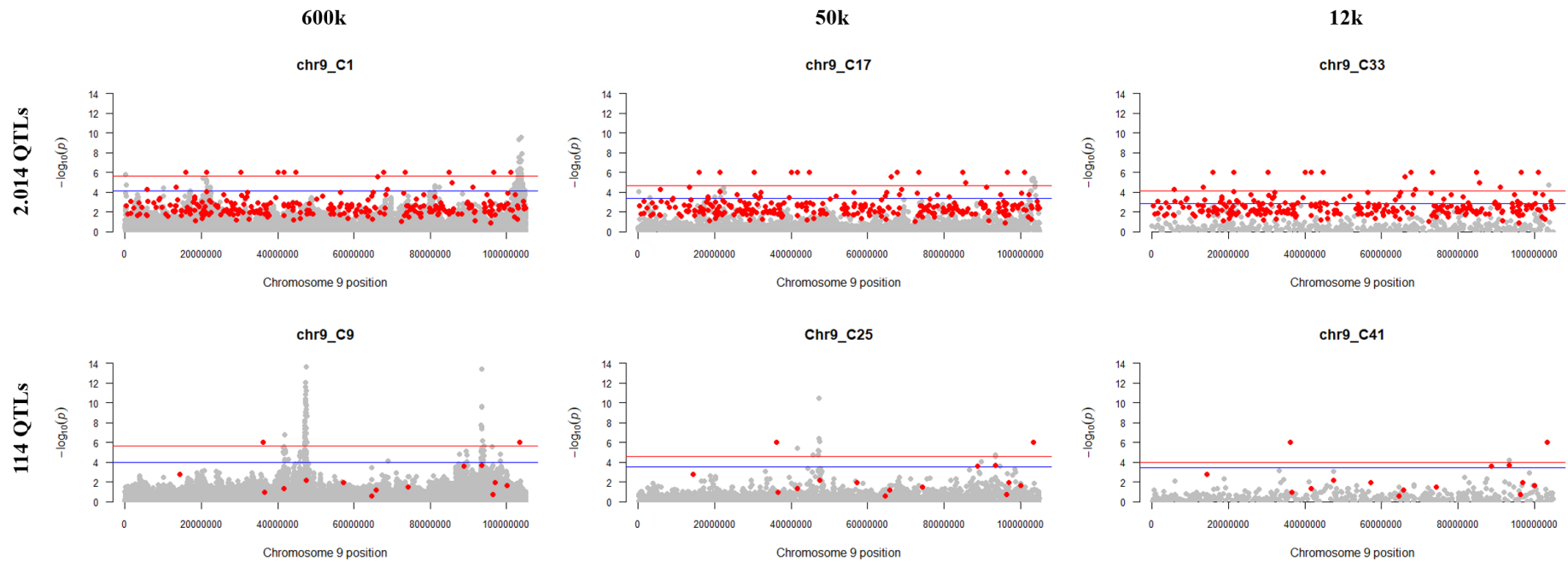


Figura 1. *Manhattan plots* com abordagem do p valor exato para dados simulados de ovinos utilizando painéis de alta, média e baixa densidade e duas quantidades de QTLs. Representação da janela do cromossomo 9 referentes aos cenários com 15 mil informações de genótipos e fenótipos. Nas linhas estão representados os cenários para 2.014 e 114 QTLs e nas colunas para os painéis de 600k, 50k e 12k, respectivamente. O eixo y representa o $-\log_{10}$ (p-valor exato) e o eixo x mostra as posições dos marcadores em pares de bases (bp). Os pontos de cor cinza representam os valores estimados dos SNPs e os pontos vermelhos representam os QTLs simulados. A linha vermelha corresponde ao limiar de *Bonferroni* e a linha azul ao limiar de FDR.

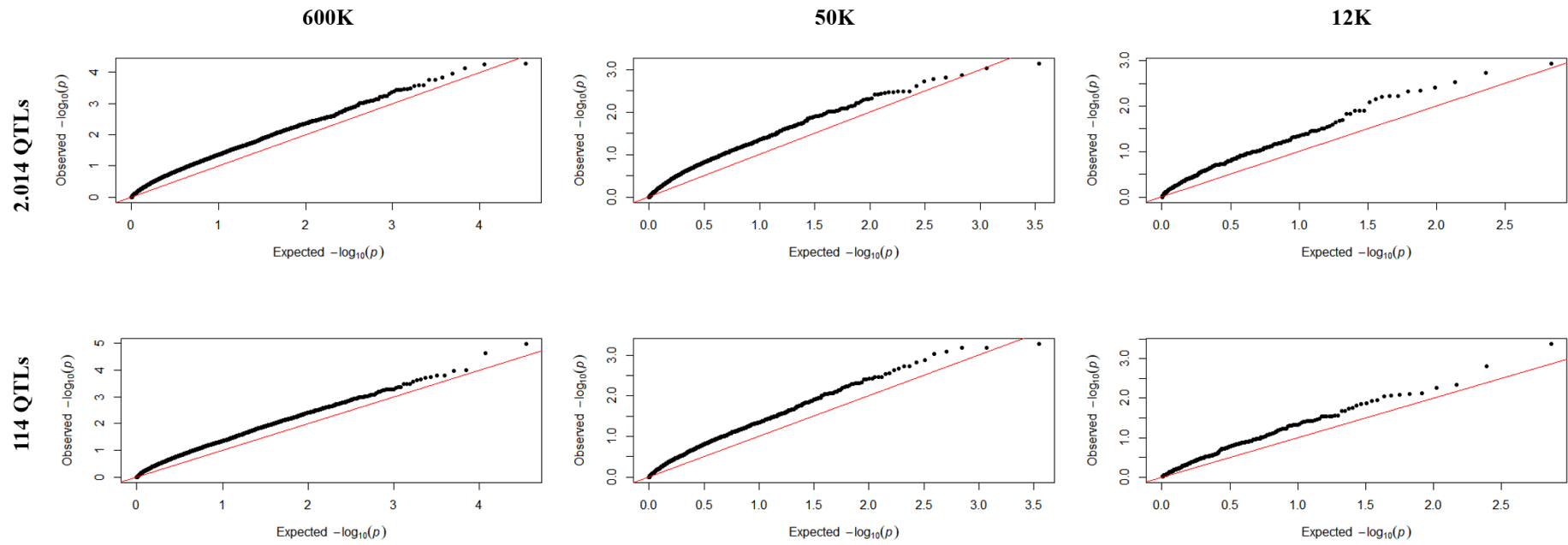


Figura 2. Gráficos Quantil-quantil (Q-Qplot) para o $-\log_{10}(p)$ -valor exato) de dados simulados de ovinos utilizando painéis de alta, média e baixa densidade e duas quantidades de QTLs. Representação da janela do cromossomo 9 referente aos cenários com 15 mil informações de genótipos e fenótipos.

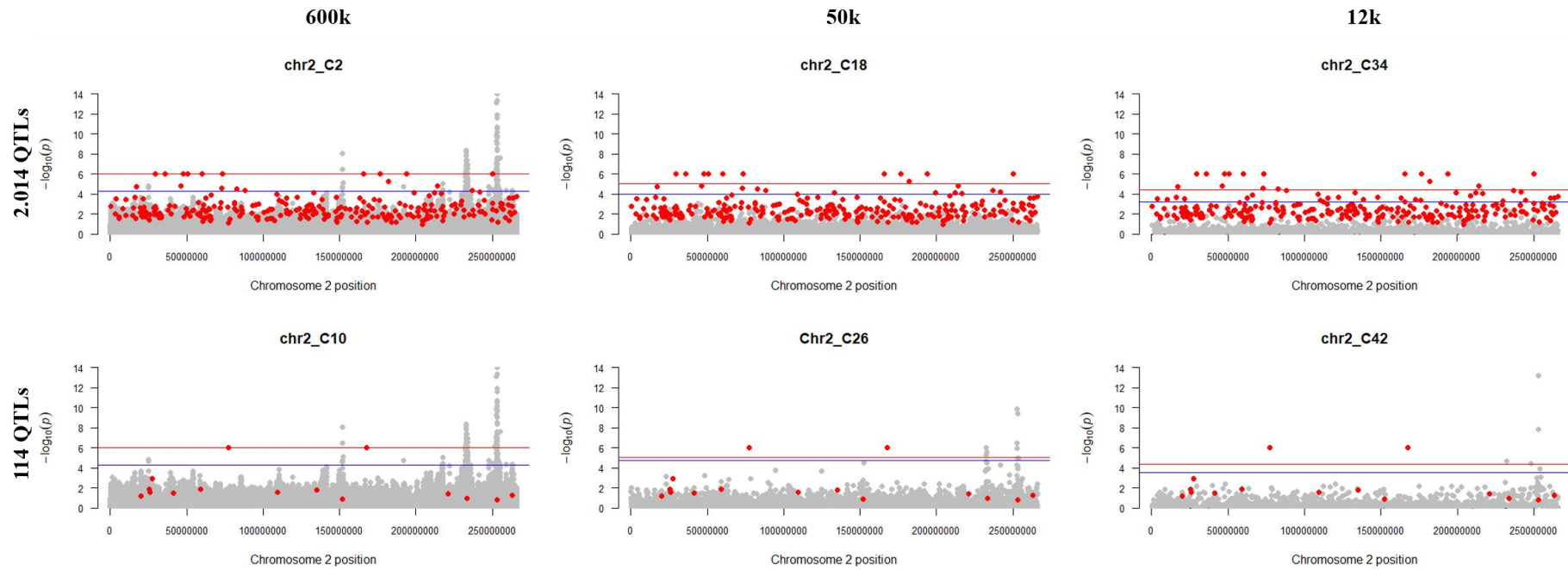


Figura 3. *Manhattan plots* com abordagem do p valor exato para dados simulados de ovinos utilizando painéis de alta, média e baixa densidade e duas quantidades de QTLs. Representação da janela do cromossomo 2 referentes aos cenários com 15 mil fenótipos e 7.500 informações de genótipos. Nas linhas estão representados os cenários para 2.014 e 114 QTLs e nas colunas para os painéis de 600k, 50k e 12k, respectivamente. O eixo y representa o $-\log_{10}(p)$ (p-valor exato) e o eixo x mostra as posições dos marcadores em pares de bases (bp). Os pontos de cor cinza representam os valores estimados dos SNPs e os círculos vermelhos representam os QTLs simulados. A linha vermelha corresponde ao limiar de *Bonferroni* e a linha azul indica o limiar de FDR.

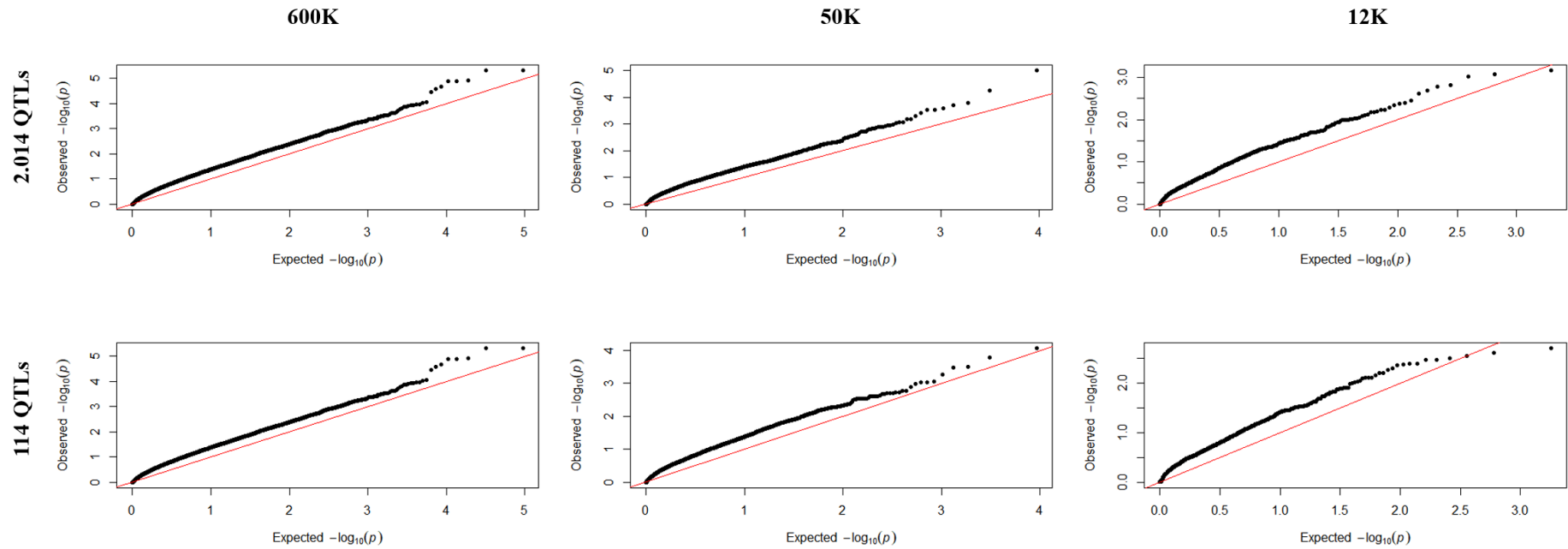


Figura 4. Gráficos Quantil-quantil (Q-Qplot) para o $-\log_{10}(p)$ -valor exato) de dados simulados de ovinos utilizando painéis de alta, média e baixa densidade e duas quantidades de QTLs. Representação da janela do cromossomo 2 referente aos cenários com 15 mil fenótipos e 7.500 informações de genótipos.

Quanto aos 114 QTLs, verificou-se picos com as associações mais evidentes que nos cenários com todos os genótipos. Os painéis de 50k e 12k apresentaram picos menores que os observados no painel de maior densidade, porém localizados na região terminal do cromossomo, aproximadamente 25Mb, próximos a QTLs. Assim, verificou-se que o método capturou os efeitos dos SNPs pelo p -valor exato, principalmente no painel de 600k para as duas quantidades de QTLs e, para 114 QTLs, por todas as densidades pelos dois limiares (Figura 2).

Com 5 mil informações de genótipos e fenótipos observou-se no painel 600k e 2.014 QTLs, cromossomo 25, picos de associação na região entre 30Mb e 35Mb, sendo detectada uma associação entre um QTL de maior efeito com os SNPs com maiores p valores. Nos painéis de 50k e 12k não foram detectadas associações. Com 114 QTLs, a abordagem do p -valor exato capturou o mesmo pico detectado com a maior quantidade de QTLs, porém com poucas associações consideradas verdadeiras pelo fato de terem apenas três QTLs simulados nesse cromossomo. Os cenários com painel de média e baixa densidade apresentaram associações significativas, com maior quantidade no cenário com o painel de 50k (Figura 3).

Nos *Manhattan plots* para 2.500 genótipos e 5 mil fenótipos verificou-se SNPs significativos associados a QTLs apenas nos cenários com o painel de alta densidade com as duas quantidades de QTLs. Detecções nos cenários com 114 QTLs foram verificadas para todas as densidades de painéis, com associações verdadeiras apenas nos painéis de alta e média densidade com um QTL na região central do cromossomo 16 (Figura 4).

Em uma comparação entre os cenários com 5 mil informações de genótipos e fenótipos com as estruturas populacionais com metade dos genótipos e mesma quantidade de fenótipos (Figuras 3 e 4), verificou-se que a abordagem do p -valor exato capturou melhor as associações dos SNPs considerados significativos com os QTLs nos cenários com o painel de alta densidade, sendo que o cromossomo 25 apresentou um dos maiores p -valores para SNPs entre todos os cenários.

Gráficos Quantil-Quantil (*Q-Q plots*) mostram uma comparação entre os p -valores observados e esperados de acordo com uma distribuição nula. Estes apresentaram menor evidência de desvio para os cenários com maior quantidade de informações de genótipos e fenótipos (Figuras 2, 4, 6, 8, 10, 12, 14 e 16).

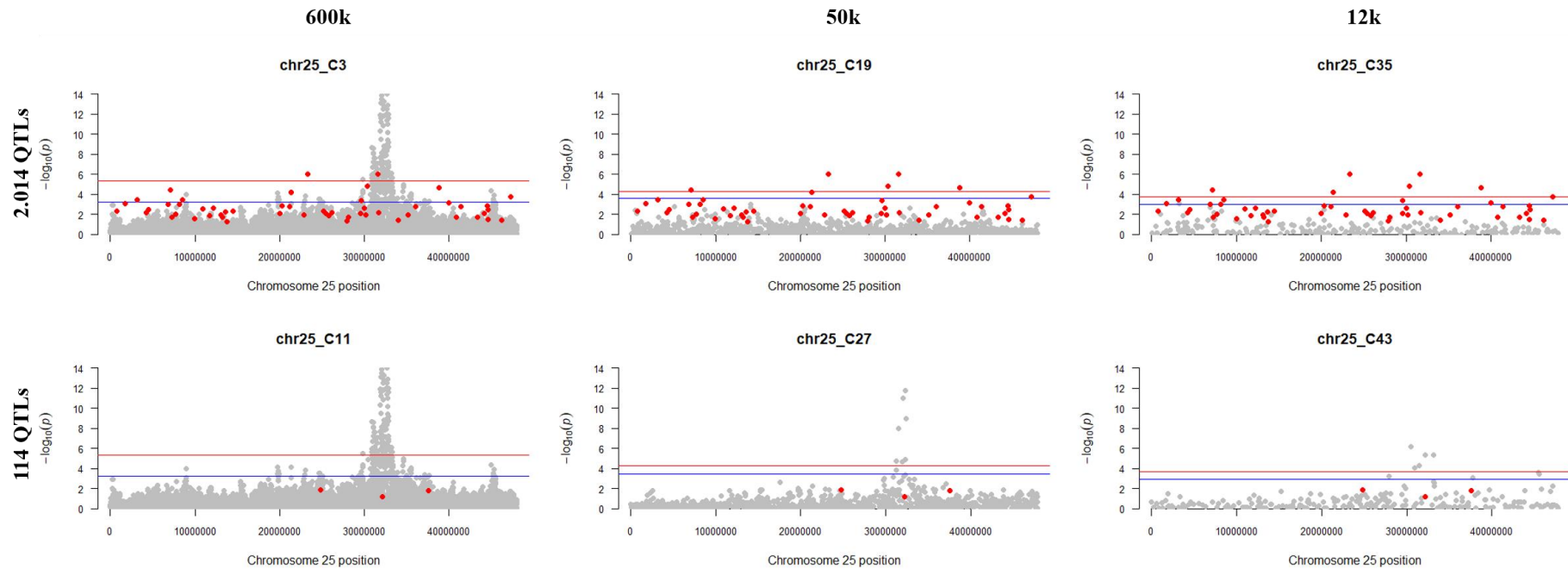


Figura 5. *Manhattan plots* com abordagem do p valor exato para dados simulados de ovinos utilizando painéis de alta, média e baixa densidade e duas quantidades de QTLs. Representação da janela do cromossomo 25 referentes aos cenários com 5 mil informações de genótipos e fenótipos. Nas linhas estão representados os cenários para 2.014 e 114 QTLs e nas colunas para os painéis de 600k, 50k e 12k, respectivamente. O eixo y representa o $-\log_{10}(p)$ (p-valor exato) e o eixo x mostra as posições dos marcadores em pares de bases (bp). Os círculos cinzas representam os valores estimados dos SNPs e os círculos vermelhos, os QTLs simulados. A linha vermelha corresponde ao limiar de *Bonferroni* e a azul ao FDR.

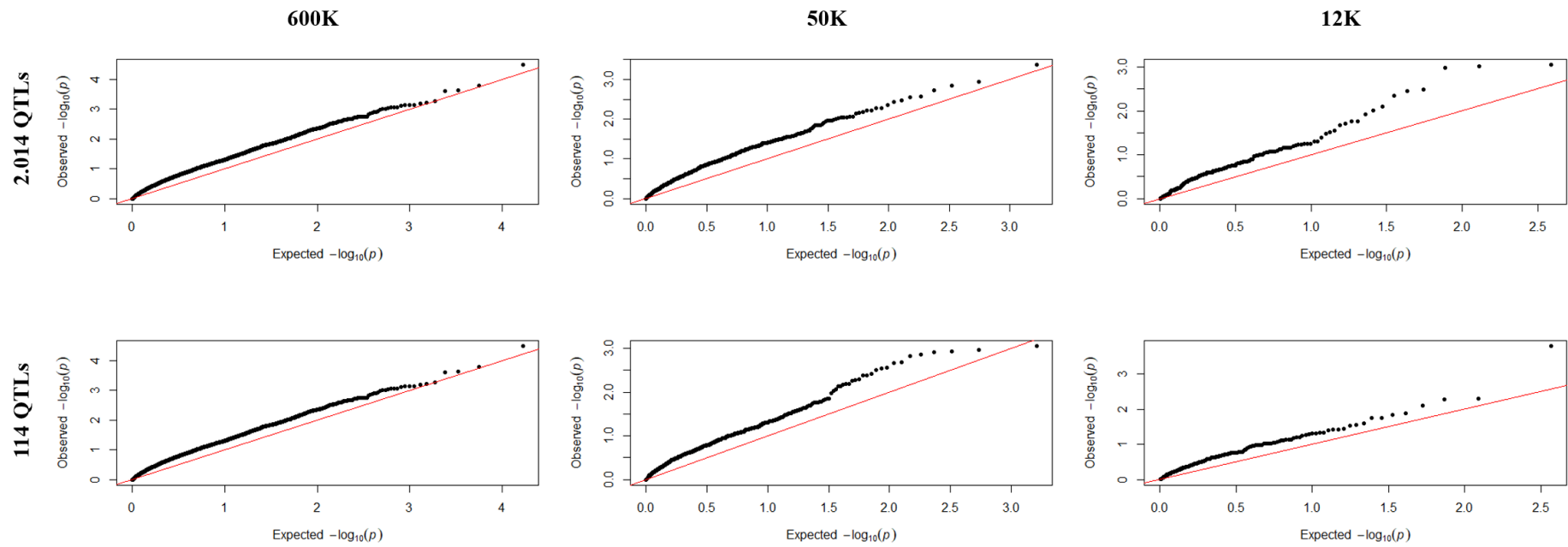


Figura 6. Gráficos Quantil-quantil (Q-Qplot) para o $-\log_{10}(p)$ -valor exato) de dados simulados de ovinos utilizando painéis de alta, média e baixa densidade e duas quantidades de QTLs. Representação da janela do cromossomo 25 referente aos cenários com 5 mil informações de genótipos e fenótipos

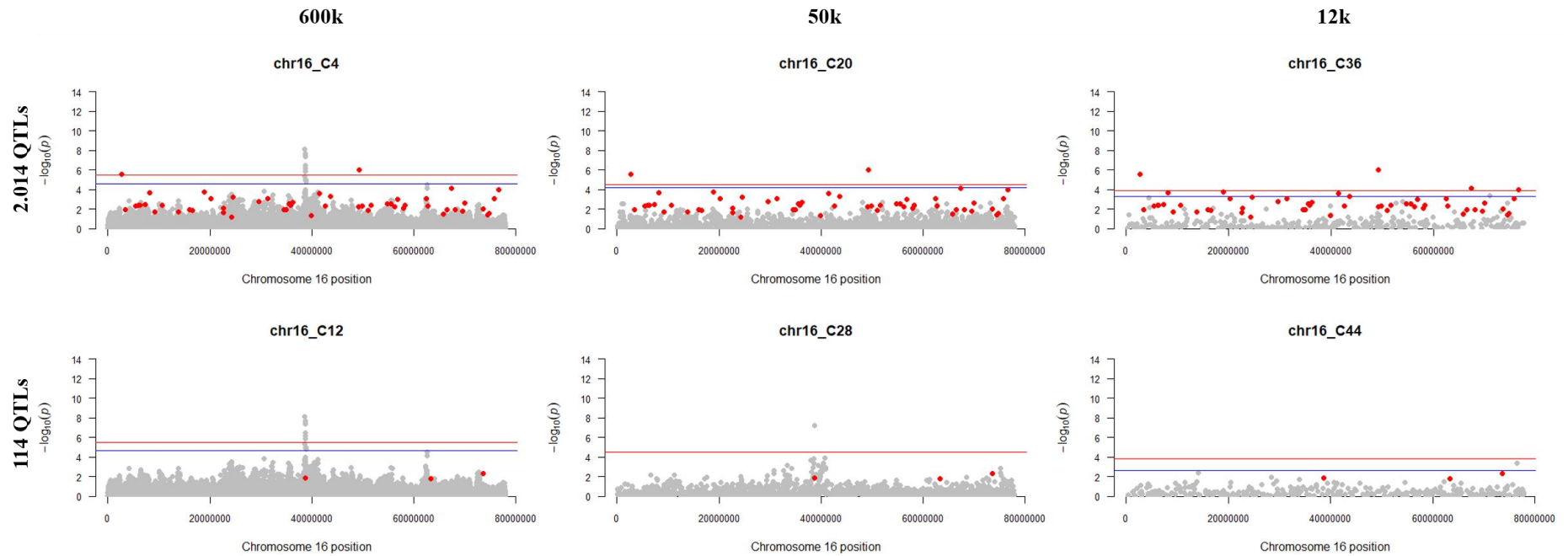


Figura 7. *Manhattan plots* com abordagem do p valor exato para dados simulados de ovinos utilizando painéis de alta, média e baixa densidade e duas quantidades de QTLs. Representação da janela do cromossomo 16 referente aos cenários com 5 mil fenótipos e 2.500 informações de genótipos. Nas linhas estão representados os cenários para 2.014 e 114 QTLs e nas colunas para os painéis de 600k, 50k e 12k, respectivamente. O eixo y representa o $-\log_{10}(p)$ (p-valor exato) e o eixo x mostra as posições dos marcadores em pares de bases (bp). Os círculos cinzas representam os valores estimados dos SNPs e os círculos vermelhos, os QTLs simulados. A linha vermelha corresponde ao limiar de *Bonferroni* e a azul ao FDR.

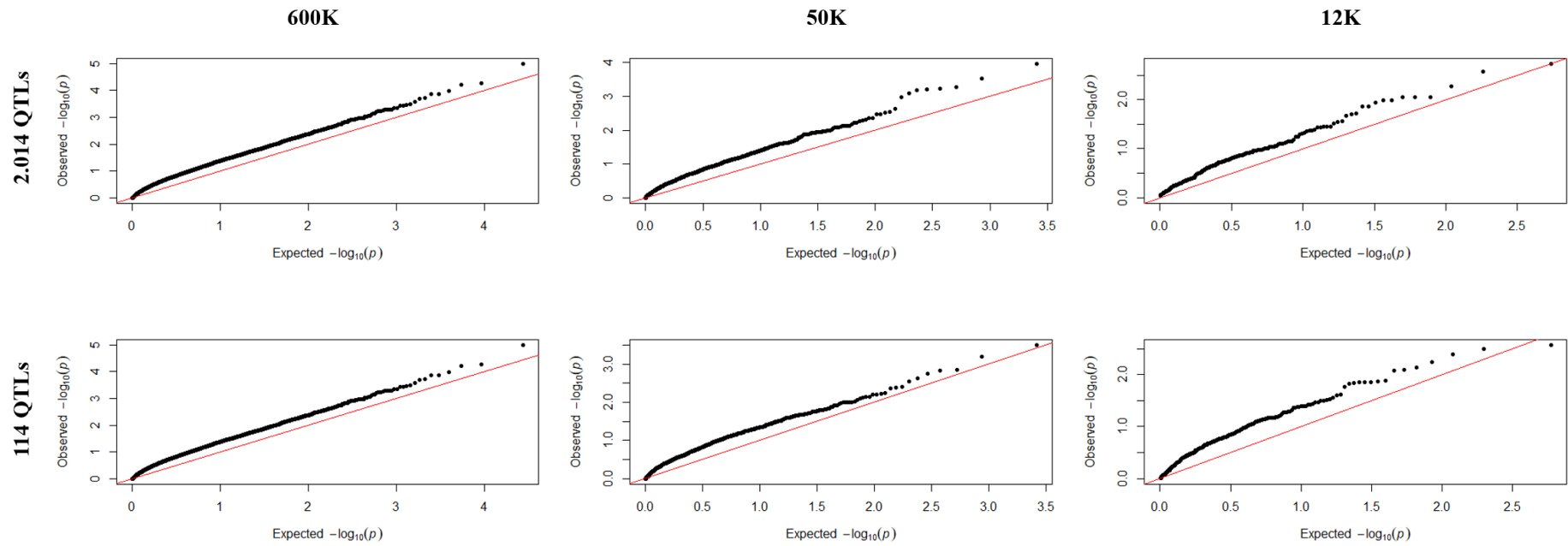


Figura 8. Gráficos Quantil-quantil (Q-Qplot) para o $-\log_{10}(p)$ -valor exato) de dados simulados de ovinos utilizando painéis de alta, média e baixa densidade e duas quantidades de QTLs. Representação da janela do cromossomo 16 referente aos cenários com 5 mil fenótipos e 2.500 informações de genótipos.

Com 2 mil informações de genótipos e fenótipos, no cromossomo dois, foram observados SNPs significativos nos painéis de 600k com vários QTLs próximos às posições dos marcadores. Nos outros painéis, para as duas quantidades de marcadores, não foram observadas associações significativas, com exceção do cenário referente ao painel de baixa densidade e 114 QTLs que apresentou SNPs significativos na posição de 25Mb que não foi captada pelo painel de 50k (Figura 5).

No cromossomo nove, cenários com 2 mil fenótipos e metade dos genótipos (1 mil) foram observados poucos SNPs significativos e apenas pelo FDR. Um SNP foi detectado no cenário com 600k e 50k com 2.014 QTLs foi considerado significativamente associado a pelo menos um QTL simulado. Com 114 QTLs apenas um SNP foi considerado significativo pelo limiar de FDR, porém não como significante associado ao QTL (Figura 6).

Comparando-se os cenários apresentados nas figuras 5 e 6, com todos os genótipos e metade destes, respectivamente, observou-se que apenas nos cenários com os painéis de alta densidade pode-se verificar associações verdadeiras entre SNPs significativos e QTLs, especialmente com 2.014 QTLs. Os SNPs significativos foram obtidos apenas pelo FDR para este cromossomo e, no cenário com 114 QTLs e baixa densidade de SNPs (Figura 5), observou-se SNPs significativos também pelo limiar de *Bonferroni*, sendo a associação entre estes e um QTL nas proximidades dos 25Mb, na região terminal do cromossomo dois, assumida como significativa, o que não foi verificado nos cenários com 1 mil genótipos e mesma quantidade de QTLs (Figura 6).

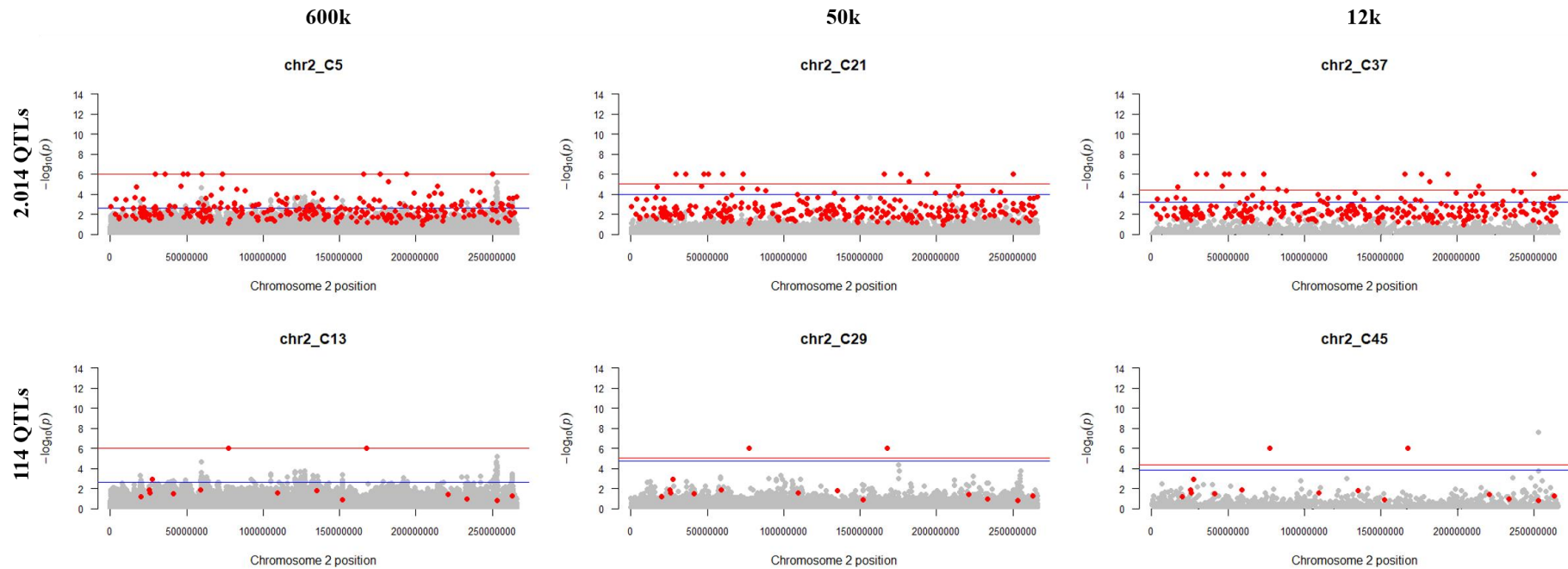


Figura 9. *Manhattan plots* com abordagem do p valor exato para dados simulados de ovinos utilizando painéis de alta, média e baixa densidade e duas quantidades de QTLs. Representação da janela do cromossomo 2 referente aos cenários com 2mil informações de genótipos e fenótipos. Nas linhas estão representados os cenários para 2.014 e 114 QTLs e nas colunas para os painéis de 600k, 50k e 12k, respectivamente. O eixo y representa o $-\log_{10}$ (p-valor exato) e o eixo x mostra as posições dos marcadores em pares de bases (bp). Os círculos cinzas representam os valores estimados dos SNPs e os círculos vermelhos, os QTLs simulados. A linha vermelha corresponde ao limiar de *Bonferroni* e a azul ao FDR.

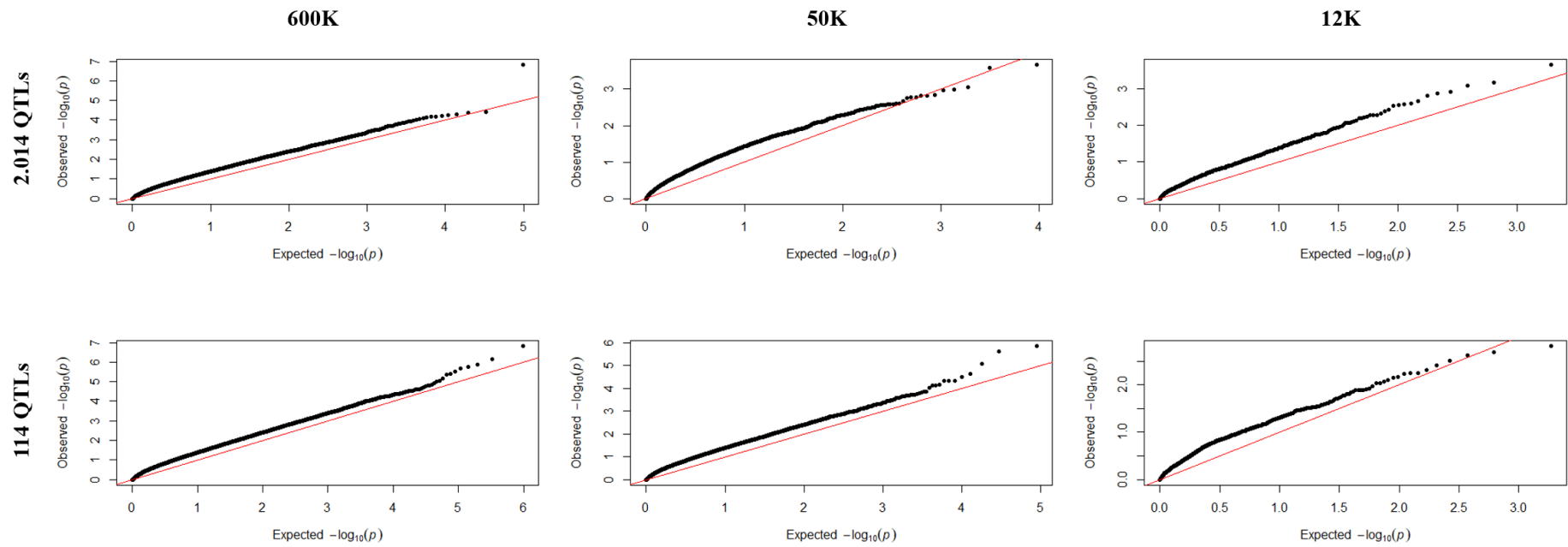


Figura 10. Gráficos Quantil-quantil (Q-Qplot) para o $-\log_{10}(p)$ -valor exato) de dados simulados de ovinos utilizando painéis de alta, média e baixa densidade e duas quantidades de QTLs. Representação da janela do cromossomo 2 referente aos cenários com 2mil informações de genótipos e fenótipos.

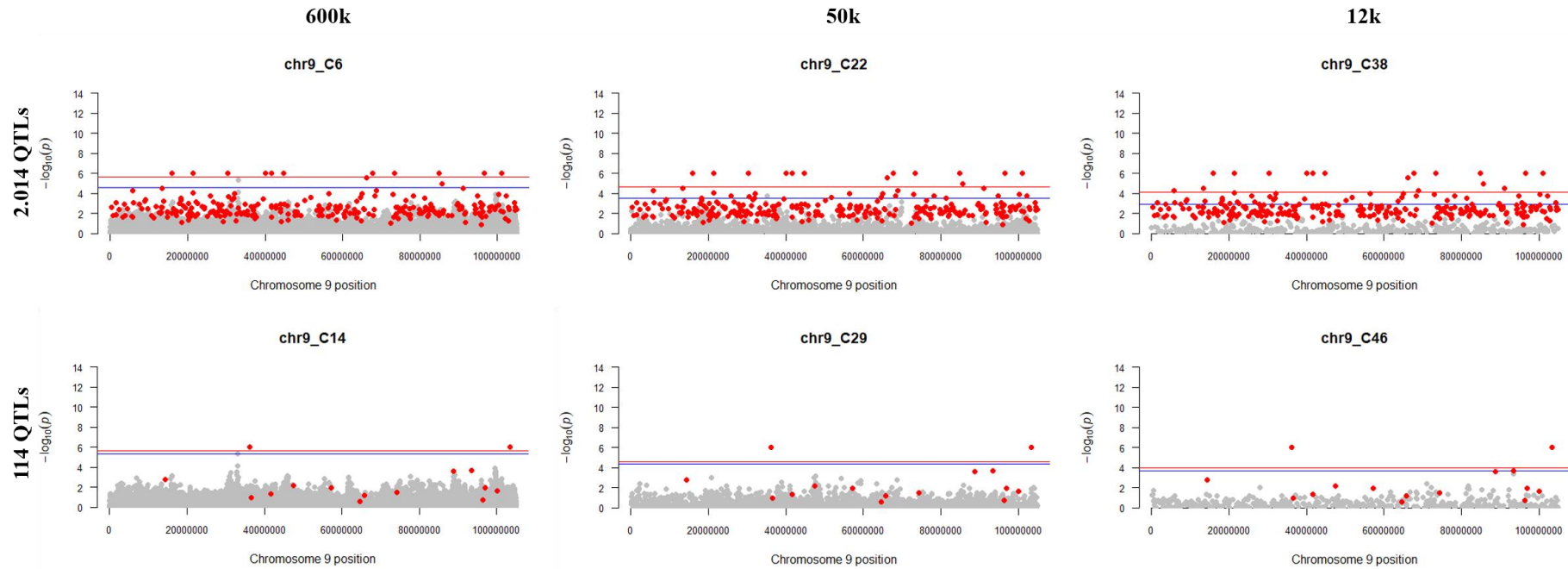


Figura 11. *Manhattan plots* com abordagem do p valor exato para dados simulados de ovinos utilizando painéis de alta, média e baixa densidade e duas quantidades de QTLs. Representação da janela do cromossomo 9 referentes aos cenários com 2mil fenótipos e 1 mil informações de genótipos. Nas colunas estão representados os cenários para 2.014 e 114 QTLs e nas linhas para os painéis de 600k, 50k e 12k, respectivamente. O eixo y representa o $-\log_{10}$ (p-valor exato) e o eixo x mostra as posições dos marcadores em pares de bases (bp). Os círculos cinzas representam os valores estimados dos SNPs e os círculos vermelhos, os QTLs simulados. A linha vermelha corresponde ao limiar de Bonferroni e a azul ao FDR.

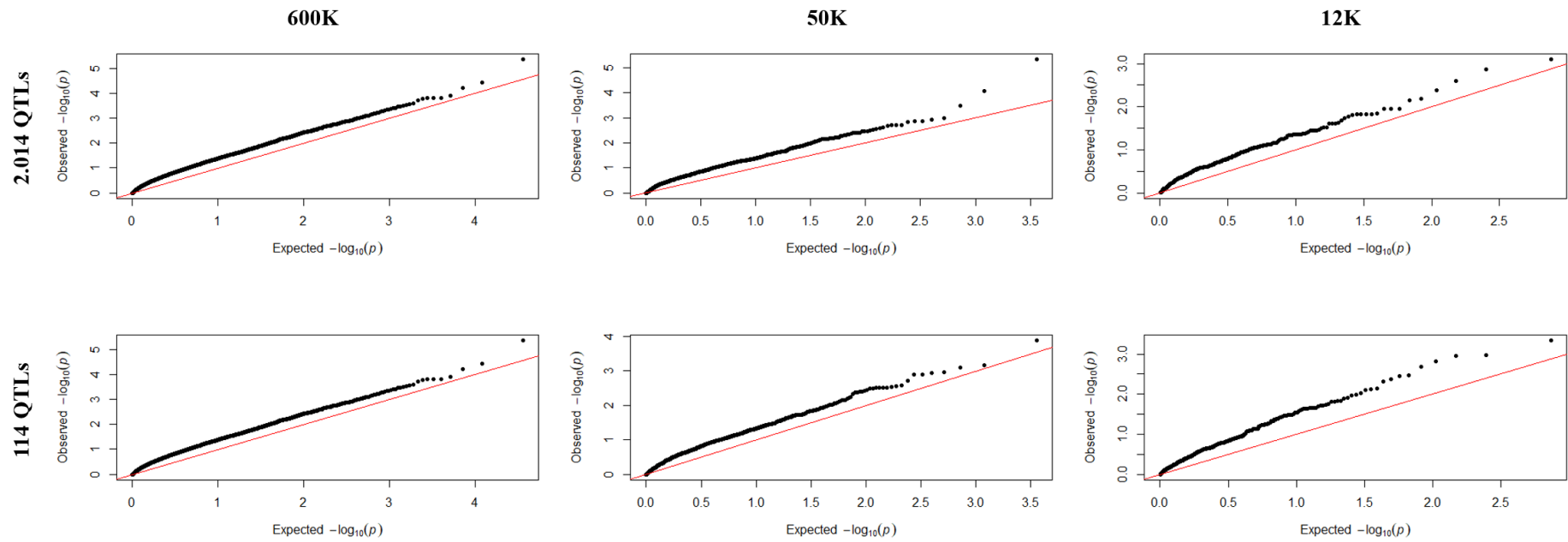


Figura 12. Gráficos Quantil-quantil (Q-Qplot) para o $-\log_{10}(p)$ -valor exato) de dados simulados de ovinos utilizando painéis de alta, média e baixa densidade e duas quantidades de QTLs. Representação da janela do cromossomo 9 referente aos cenários com 2mil fenótipos e 1 mil informações de genótipos.

Na figura 7, com 500 informações de genótipos e fenótipos, foram verificados SNPs significativos, de acordo com o limiar FDR, com 2.014 QTLs apenas com o painel de 600k nas porções inicial, central e terminal do cromossomo um, com pelo menos uma associação verdadeira detectada. No cromossomo um, com 114 QTLs foram observadas SNPs significativos em todos os painéis com redução no número de associações com a diminuição do número de marcadores. Apenas 3 QTLs simulados e de pequeno efeito foram verificados próximos a SNPs significativos, sendo que somente nos painéis de 600k e 50k foram observadas uma associação significativa entre SNP e QTL para esta quantidade de QTLs.

Com menor quantidade de informações de genótipos (250) e 500 de fenótipos, pelo cromossomo 11, foram observadas associações significativas entre SNPs e QTLs nos cenários com 600k e 50k com a maior quantidade de QTLs. Com 114 QTLs, apenas um SNP foi considerado significativo pelo limiar FDR, porém sem associação com um dos QTLs simulados presentes nesse cromossomo (Figura 8).

Assim, nos cenários com menores quantidades de informações de genótipos e fenótipos, verificou-se menor quantidade de SNPs significativos pelos limiares utilizados e menor possibilidade de detecção de associações verdadeiras entre marcadores e QTLs.

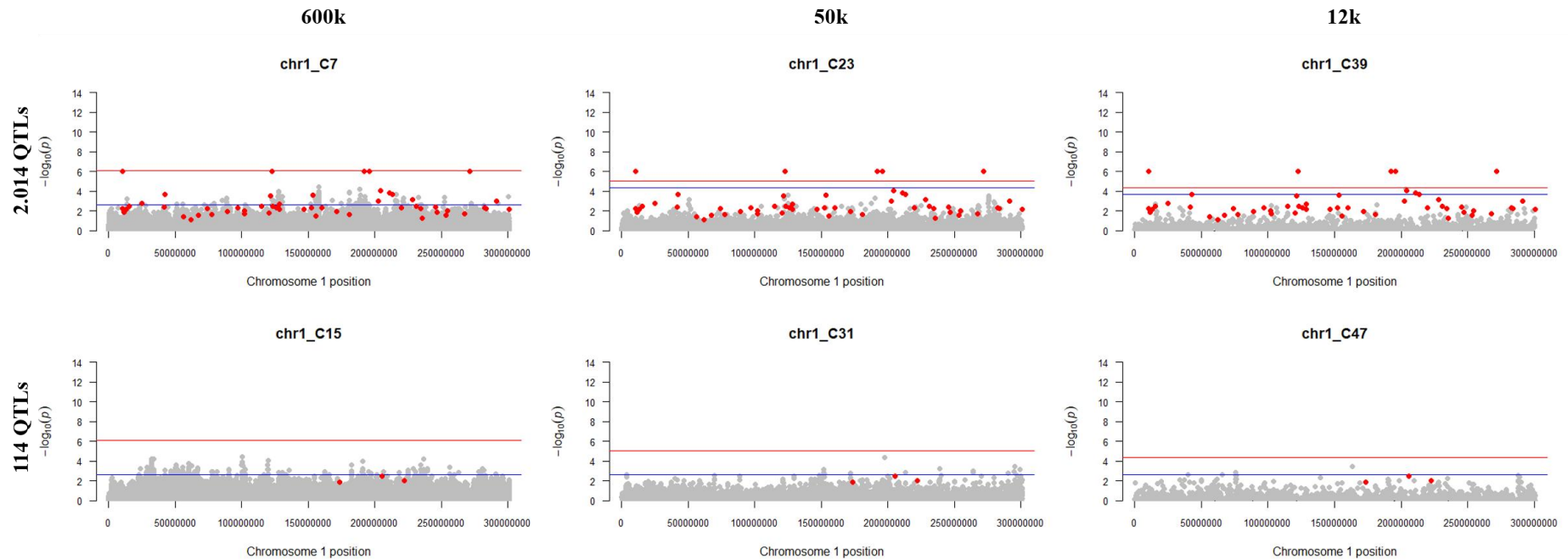


Figura 13. *Manhattan plots* com abordagem do p valor exato para dados simulados de ovinos utilizando painéis de alta, média e baixa densidade e duas quantidades de QTLs. Representação da janela do cromossomo 1 referente aos cenários com 500 informações de genótipos e fenótipos. Nas linhas estão representados os cenários para 2.014 e 114 QTLs e nas colunas para os painéis de 600k, 50k e 12k, respectivamente. O eixo y representa o $-\log_{10}$ (p-valor exato) e o eixo x mostra as posições dos marcadores em pares de bases (bp). Os círculos cinzas representam os valores estimados dos SNPs e os círculos vermelhos, os QTLs simulados. A linha vermelha corresponde ao limiar de *Bonferroni* e a azul ao FDR.

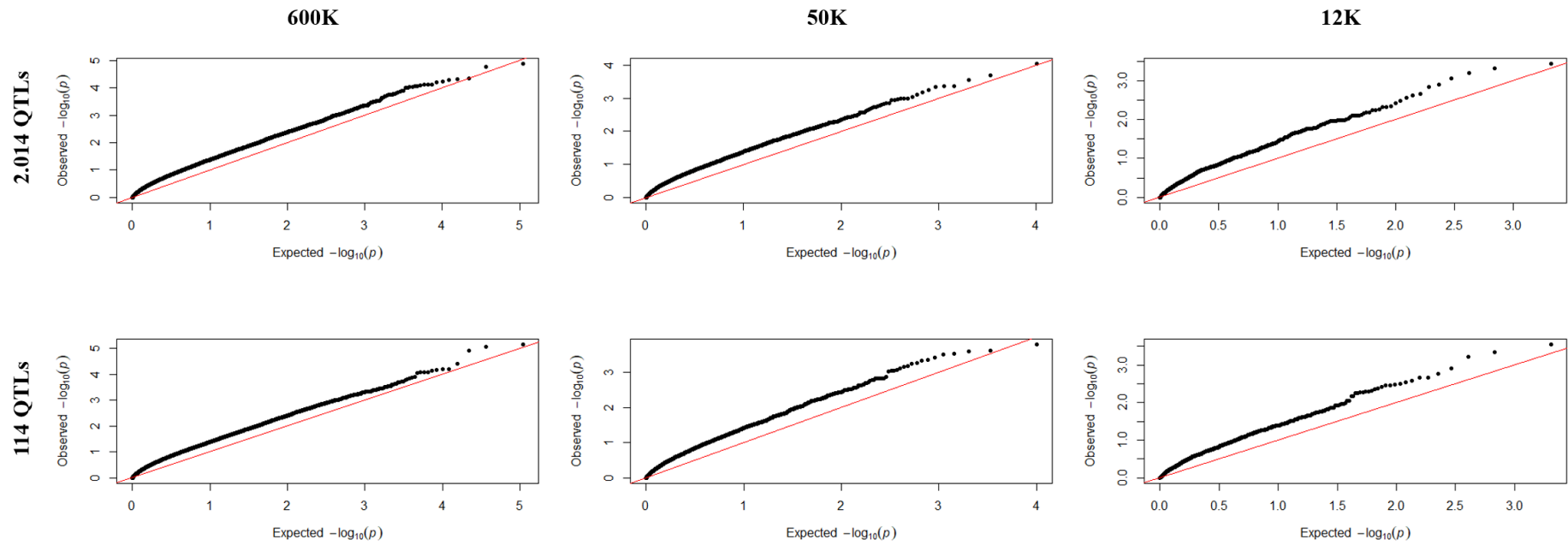


Figura 14. Gráficos Quantil-quantil (Q-Qplot) para o $-\log_{10}(p)$ -valor exato) de dados simulados de ovinos utilizando painéis de alta, média e baixa densidade e duas quantidades de QTLs. Representação da janela do cromossomo 1 referente aos cenários com 500 informações de genótipos e fenótipos.

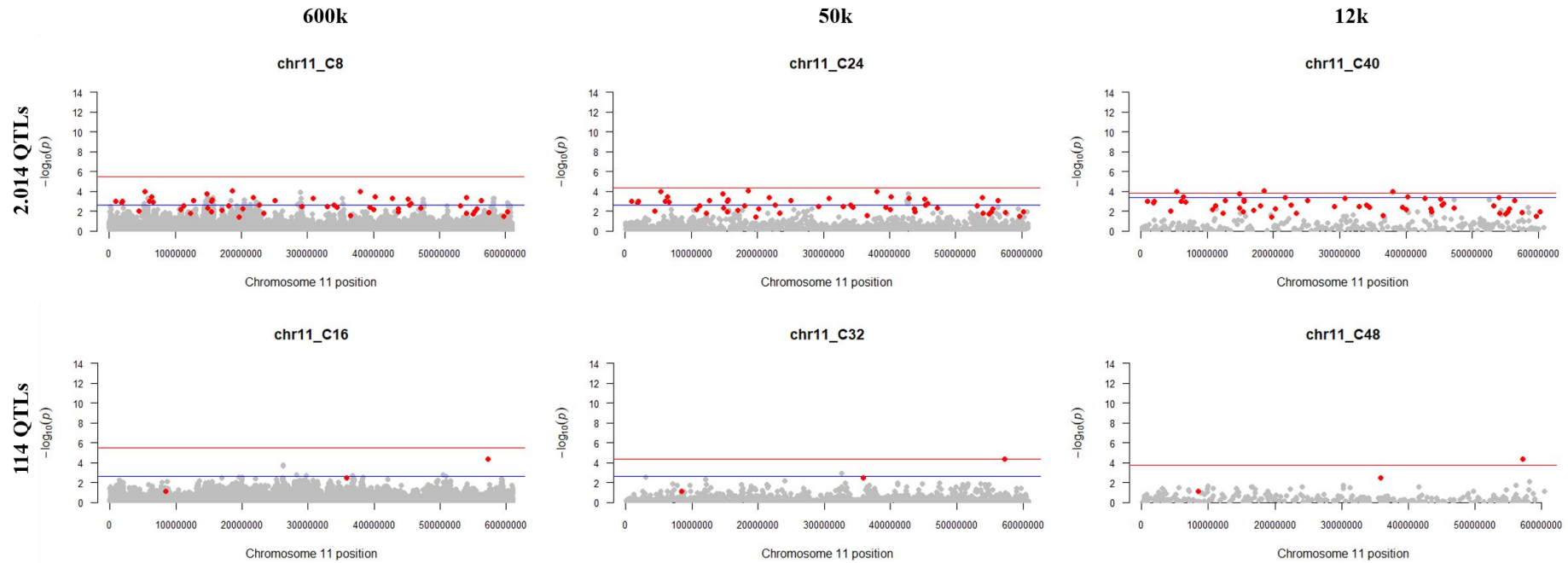


Figura 15. *Manhattan plots* com abordagem do p valor exato para dados simulados de ovinos utilizando painéis de alta, média e baixa densidade e duas quantidades de QTLs. Representação da janela do cromossomo 11 referente aos cenários com 500 fenótipos e 250 informações de genótipos. Nas linhas estão representados os cenários para 2.014 e 114 QTLs e nas colunas para os painéis de 600k, 50k e 12k, respectivamente. O eixo y representa o $-\log_{10}(p)$ (p-valor exato) e o eixo x mostra as posições dos marcadores em pares de bases (bp). Os círculos cinzas representam os valores estimados dos SNPs e os círculos vermelhos, os QTLs simulados. A linha vermelha corresponde ao limiar de *Bonferroni* e a azul ao FDR.

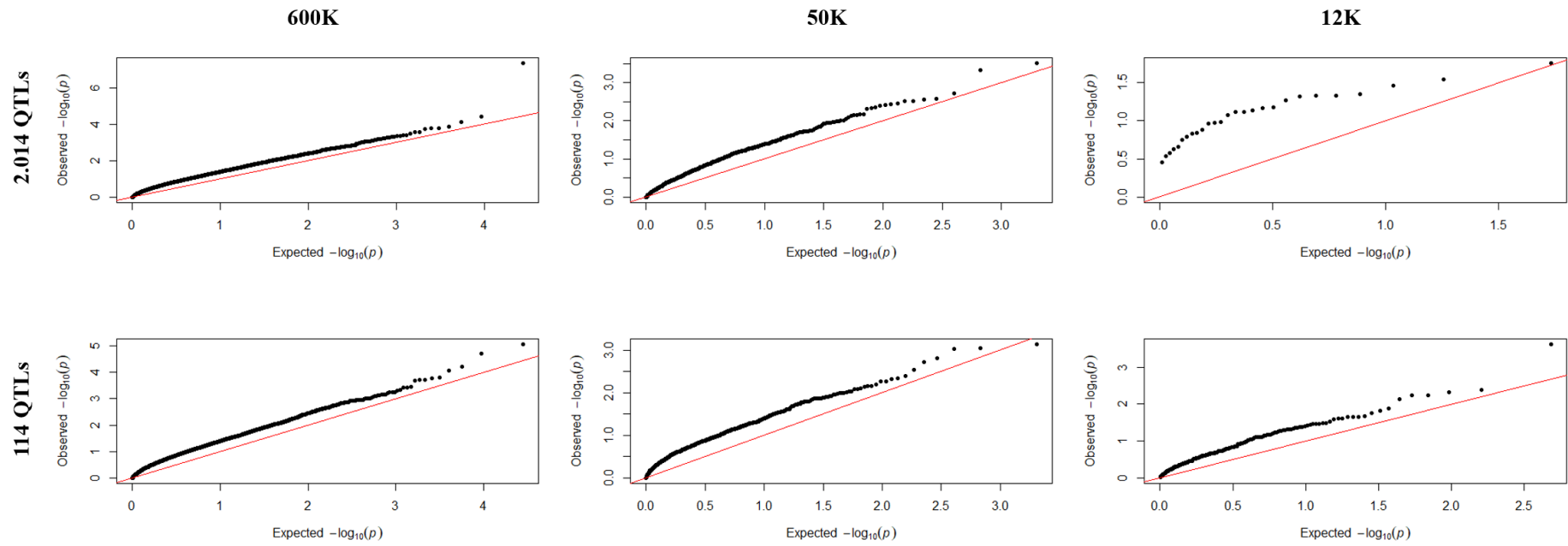


Figura 16. Gráficos Quantil-quantil (Q-Qplot) para o $-\log_{10}(p)$ -valor exato) de dados simulados de ovinos utilizando painéis de alta, média e baixa densidade e duas quantidades de QTLs. Representação da janela do cromossomo 11 referente aos cenários com 500 fenótipos e 250 informações de genótipos.

Verificou-se que quanto maior a quantidade de animais genotipados e SNPs no painel, maior a possibilidade de detecção de associações verdadeiras entre os SNPs significativos e os QTLs simulados. Com 2.014 QTLs, os cenários com o painel de alta densidade proporcionaram maior quantidade de detecções verdadeiras. Com 114 QTLs, geralmente, observou-se pelo menos uma associação verdadeira, com picos maiores nos cenários com maior quantidade de animais.

4. Discussão

As simulações foram realizadas com uma herdabilidade moderada baseada na característica AOL avaliada em ovinos criados nos trópicos, no valor de 0,30 de acordo com estudo de SENA et al. (2020). Neste estudo, as magnitudes das estimativas de herdabilidade foram classificadas como baixa ($< 0,10$), moderada (de 0,11 a 0,40) e alta ($> 0,41$), para fins de comparação (MASSENDER et al., 2019). Assim, as estimativas de herdabilidade obtidas foram consideradas como moderadas, corroborando com a magnitude dos valores inicialmente simulados para a característica utilizada como base. Como mostrado anteriormente, os valores médios de herdabilidade obtidos nas simulações com as duas quantidades de QTLs foram de magnitude moderada e quanto maior a quantidade de SNPs no painel, mais os cenários se aproximaram da herdabilidade simulada. Como exceção, as maiores estimativas de herdabilidade média foram observadas nos cenários com 114 QTLs e menor quantidade de informações de genótipos e fenótipos (500 e 250 animais genotipados), o que pode ser devido a seleção aleatória de animais com diferenças de variabilidade genética.

De acordo com estudo de Pértile et al. (2016), o aumento do coeficiente de herdabilidade leva ao aumento do poder de detecção de QTLs, pois os efeitos estimados dos marcadores nas proximidades dos QTLs simulados são maiores. No mesmo estudo, quando se avaliou uma estrutura com herdabilidade de 0,25, de magnitude moderada, observou-se uma detecção de *loci* semelhante à de cenários com herdabilidade alta, porém com menores efeitos. O aumento da herdabilidade também reduz a ocorrência de falsos positivos (VAN DEN BERG; FRITZ; BOICHARD, 2013).

Maior tamanho da população de estudo e a maior quantidade de QTLs possibilitaram a observação de mais marcadores significativos próximos aos QTLs simulados, tomados como verdadeiros. Kadri et al. (2014) afirmaram que o aumento do

tamanho da população, bem como da quantidade de marcadores disponíveis pode levar ao aumento na precisão da detecção de QTLs. Corroborando com os autores supracitados, Alkaladeh et al. (2019) afirmaram que um maior número de observações e densidade de SNPs em GWAS tende a um maior poder de detecção de associações.

Os gráficos gerados por meio do ssGBLUP pela abordagem do p -valor exato mostraram que houve eficiência na detecção de associações entre SNPs significativos e QTLs simulados. Porém, a quantidade de picos significativos variou dependendo do limiar de significância adotado. Uma tendência entre os cenários foi que a medida que se reduzia a quantidade de informações dos animais, reduzia-se também a possibilidade de detecção de associações próximas às posições dos QTLs. Poucas associações de QTLs de efeito maior com SNPs foram observadas. A abordagem do p -valor exato possibilitou a detecção de pelo menos uma associação significativa entre SNPs e QTLs simulados.

Takeda et al. (2020) associaram a maior quantidade de QTLs que investigaram, no caso 500, com o menor poder de detecção de QTLs por conta da diluição dos efeitos dos QTLs ocasionado pela distribuição gama, sendo a variância explicada por cada QTL com menor participação na variância genética total da característica avaliada. Os autores verificaram que nos cenários em que foram considerados os 300 touros genotipados e informações fenotípicas das progênies (1.500, 4.500 e 9.000) não houve influência da herdabilidade (0,20 e 0,50) e do número de QTLs simulados (10, 50 e 500) no poder de detecção dos QTLs, o que não corroborou com os resultados apresentados nesta pesquisa.

Por fim, pode-se, com esta pesquisa, ter como base a quantidade de informações necessárias (no mínimo, 2 mil animais genotipados) e densidade de SNPs mínimos (painel de média densidade) para a obtenção de associações entre SNPs significativos e QTL simulados baseado em uma característica de magnitude moderada semelhante a AOL em ovinos, com essa estrutura de dados.

5. Conclusões

A abordagem do p -valor exato foi capaz de detectar pelo menos uma associação significativa de marcadores com os QTLs simulados nos cenários de estudo, por meio da metodologia ssGWAS. A maior densidade do painel simulado e quantidade de informações genotípicas e fenotípicas possibilitaram uma melhor capacidade de detecção do método. Essas informações estão de acordo com o esperado, porém sugere-se a adoção

de critérios de comparação entre os cenários, bem como para a verificação de possíveis associações espúrias.

Os resultados apresentados aqui podem servir de base para estudos de associação genômica mais profundos, incluindo cenários com herdabilidades de magnitudes distintas e número de amostras simuladas de acordo com o tamanho efetivo, similar ao de populações de ovinos tropicais reais.

Referências

- AGUILAR, I. et al. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. **Journal of Dairy Science**, v. 93, n. 2, p. 743-752, 2010.
- AGUILAR, I. et al. Frequentist p-values for large-scale-single step genome-wide association, with an application to birth weight in American Angus cattle. **Genetics Selection Evolution**, v. 51, n. 28, 2019.
- AL KALALDEH, M. et al. Detection of genomic regions underlying resistance to gastrointestinal parasites in Australian sheep. **Genet Sel Evol**, v. 51, n. 37, 2019.
- ALVARENGA, A. B. et al. Linkage disequilibrium in Brazilian Santa Inês breed, *Ovis aries*. **Scientific Reports**, v. 8, n. 8851, p. 1-11, 2018.
- BENJAMINI, Y.; HOCHBERG, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. **Journal of the Royal Statistical Society. Series B (Methodological)**, v. 57, n. 1, pp. 289-300, 1995.
- BOLOORMA, S. et al. A genome-wide association study of meat and carcass traits in Australian cattle. **Journal Animal Science**, v. 89, p. 2297-2309, 2011.
- DAETWYLER, H. D. et al. Genomic prediction in animals and plants: simulation of data, validation, reporting, and benchmarking. **Genetics**, v. 193, n. 2, p. 347–365, 2013.
- DEKKERS, J. C. M. Application of Genomics Tools to Animal Breeding. **Current Genomics**, v. 13, p. 207-212, 2012.
- DUNN, O. J. Multiple Comparisons Among Means. **Journal of the American Statistical Association**, v. 56, n. 293, p. 52–64, 1961. doi: <https://doi.org/10.2307/2282330>
- KADRI et al. Comparison of Genome-Wide Association Methods in Analyses of Admixed Populations with Complex Familial Relationships. **PLoS ONE**, v. 9, n. 3, 2014.
- LI, H. et al. An efficient unified model for genome-wide association studies and genomic selection. **Genetics Selection Evolution**, v. 49, n. 64, p. 1-8, 2017.
- MASSENDER et al. A genetic evaluation of growth, ultrasound, and carcass traits at alternative slaughter endpoints in crossbred heavy lambs. **J. Anim. Sci**, v. 97, p. 521-535, 2019.
- MEUWISSEN, T.; HAYES, B. J.; GODDARD M. E. Prediction of total genetic values using genome-wide dense marker maps. **Genetics**, v. 157, p. 1819-1829, 2001.
- MISZTAL, I. et al. **Manual for BLUPF90 family of programs**. Disponível em: <http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all7.pdf>. Acessado em: 15 de Maio de 2019).

NICOLAZZI, E. L. et al. SNPchiMp v.3: integrating and standardizing single nucleotide polymorphism data for livestock species. **BMC Genomics**, v. 16, n. 283, 2015.

PÉRTILE, S. F. N. et al. Seleção e associação genômica ampla para o melhoramento genético animal com uso do método ssGBLUP. **Pesquisa Agropecuária Brasileira**, v.51, n.10, p.1729-736, 2016.

R CORE TEAM (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

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

SENA, L. S. et al. Genetic evaluation of tropical climate-adapted sheep for carcass traits including genomic information. **Small Ruminant Research**, v. 188, 2020.

TAKEDA, M.; UEMOTO, Y.; SATOH, M. Effect of genotyped bulls with different numbers of phenotyped progenies on quantitative trait loci detection and genomic evaluation in a simulated cattle population. **Animal Science Journal**, v. 91, n. 1, 2020.

TURNER, S. D. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. **bioRxiv**, 2014. doi: <https://doi.org/10.1101/005165>

VAN DEN BERG, I.; FRITZ, S.; BOICHARD, D. QTL fine mapping with Bayes C(π): a simulation study. **Genetics Selection Evolution**, v. 45, n.19, 2013.

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

CAPÍTULO II

Seleção genômica ampla em populações simuladas de ovinos

Seleção genômica ampla em populações simuladas de ovinos

BARBOSA, Bruna Lima. **Seleção genômica ampla em populações simuladas de ovinos**. 2020. Tese (Doutorado em Ciência Animal) – Universidade Federal do Piauí, Teresina, Piauí, 2020.

RESUMO: O objetivo deste trabalho foi avaliar a eficiência da seleção genômica em populações simuladas de ovinos, com utilização de diferentes estruturas de treinamento e tamanhos populacionais simulados. As simulações foram realizadas a fim de alcançar a arquitetura genética da característica área de olho de lombo. Dois painéis de SNPs com 11.385 (12k) e 52.326 (50k) marcadores foram utilizados. Diferentes tamanhos de população de referência e validação foram utilizados, totalizando seis cenários diferentes para cada painel. O método GBLUP de único passo (ssGBLUP) foi utilizado para a estimação dos efeitos dos SNPs sobre a característica. A predição da acurácia foi obtida pela correlação de Pearson entre os valores genéticos genômicos estimados (GEBVs) e verdadeiros (TBVs). Coeficientes de regressão e o viés também foram calculados. As estimativas de herdabilidade variaram de 0,17 a 0,38 nos diferentes cenários avaliados. Os maiores valores foram obtidos com uso das estruturas de treinamento com maior quantidade de animais genotipados, enquanto os menores valores foram observados nas populações com menos animais genotipados na estrutura de treinamento. Diferenças significativas foram observadas entre os diferentes tamanhos da população de treinamento, porém, referente aos cenários com o treinamento com menor quantidade de informações de genótipos e fenótipos, não foi verificada diferença significativa entre os resultados obtidos com uso dos dois painéis. Há uma tendência em aumento da acurácia com o acréscimo de animais genotipados na população de treinamento ou referência.

Palavras-chave: acurácia, viés, *Ovis aries*, simulação, ssGBLUP

Genome-wide selection in simulated sheep populations

BARBOSA, Bruna Lima. **Genome-wide selection in simulated population of sheep**. 2020. Thesis (Doctor of Philosophy in Animal Science) – Universidade Federal do Piauí, Teresina, 2020.

ABSTRACT: The objective of this study was to evaluate the efficiency of genomic selection in sheep using different simulated genetic structures and population sizes. Simulations were performed to represent the genetic architecture of the loin eye area (LEA) trait. Two SNP panels with 11,385 (12k) and 52,326 (50k) markers were used. Different sizes of reference and validation populations were used, totaling six different scenarios for each panel. The single-step GBLUP (ssGBLUP) method was used to estimate the effects of SNPs on the phenotype. The prediction accuracy was obtained by Pearson correlation between the estimated genomic breeding values (GEBVs) and true breeding values (TBVs), as well as the regression coefficients and bias of prediction. Heritability estimates ranged from 0.17 to 0.38 in the different scenarios evaluated. The highest values were obtained using the training sets with larger number of genotyped animals, whereas the lowest values were observed in populations containing lower number of genotyped animals in the training set. Significant differences were observed between scenarios with different training population sizes; however, no significant differences were observed among the results of scenarios with lower training population sizes, considering the two SNP panels evaluated. There is a trend of increasing the prediction accuracy with the addition of genotyped animals either in the training or reference populations.

Keywords: accuracy, bias, *Ovis aries*, simulation, ssGBLUP

1. Introdução

A produção pecuária representa uma das principais atividades econômicas praticadas mundialmente. Com o constante aumento da população humana no planeta, há também a tendência de aumento na demanda por proteína animal. Por outro lado, as áreas disponíveis para a prática de agricultura e pecuária têm se mantido constantes ou diminuído (SIMIANER, 2016). Sendo assim, para atender à demanda da população mundial por proteína animal, há a necessidade de investimentos maciços no sentido de aumentar a eficiência da produção pecuária, ou seja, animais que produzam mais rápido e com maior qualidade.

Uma das principais alternativas para aumentar a eficiência da produção animal consiste no melhoramento genético dos rebanhos. Atualmente, entre as metodologias utilizadas no melhoramento genético animal, a seleção genômica (GS, *genomic selection*), idealizada por Meuwissen, Hayes e Goddard (2001), apresenta maior destaque (MEUWISSEN; HAYES; GODDARD, 2016). Para a implementação da GS, basicamente, há a necessidade da formação de uma população de treinamento (animais genotipados e com informação fenotípica) e uma população de validação (candidatos à seleção que possuem apenas informação de genótipo). De modo geral, quanto maiores a densidade dos painéis de marcadores moleculares e o tamanho amostral, maiores serão as acurácias das estimativas dos efeitos de marcadores e da predição de mérito genético dos animais avaliados (MEUWISSEN; HAYES; GODDARD, 2013).

Painéis de marcadores com maior densidade possibilitarão aumento da acurácia de predição devido à obtenção de maiores níveis de desequilíbrio de ligação (LD) entre *loci* de características quantitativas (QTL) e marcadores (TOOSI; FERNANDO; DEKKERS, 2010). Além disso, um maior tamanho amostral resultará em aumento nas acurácias devido à diminuição do viés e da variância dos efeitos dos marcadores, e devido à possibilidade de aumentar a extensão do relacionamento genético entre indivíduos das populações de treinamento e validação (DE LOS CAMPOS et al., 2013).

Na realidade dos sistemas produtivos, normalmente, quanto maior a quantidade de animais no rebanho, mais elevados são os custos para a manutenção da atividade pecuária. Além disso, painéis de marcadores moleculares com maior densidade custam mais caro. Estes são os principais fatores que oneram a implementação eficiente da seleção genômica em programas de melhoramento genético animal. As cadeias produtivas que têm investido mais intensamente na adoção de GS são aquelas que apresentam melhores estruturas, como é o caso da bovinocultura (de leite e corte),

suinocultura e avicultura (MEUWISSEN; HAYES; GODDARD, 2016). Em atividades como ovinocultura e caprinocultura, a GS ainda apresenta aplicação bem menos expressiva e têm sido observadas menores acurácias, principalmente, em comparação a bovinos (RUPP et al., 2016). Esta situação é mais perceptível em países em desenvolvimento, onde há os menores investimentos na produção de pequenos ruminantes e em seleção genômica (MRODE et al., 2018).

Os métodos que possibilitaram inicialmente a utilização de informação genômica em esquemas de melhoramento genético animal foram desenvolvidos a partir de dados simulados (MEUWISSEN; HAYES; GODDARD, 2001). Como alternativa para a utilização de dados reais na seleção genômica, vários pesquisadores têm utilizado a simulação de dados para avaliarem a viabilidade da implementação de GS em diferentes espécies. O desenvolvimento de estudos com este tipo de dados requer menores investimentos de tempo e dinheiro, em relação a dados reais, e possibilita aos pesquisadores avaliarem aspectos que não seriam possíveis com o uso de bancos de dados reais (DAETWYLER et al., 2013).

Os estudos genômicos por meio de simulações têm se tornado cada vez mais importantes para os avanços no melhoramento animal de várias espécies. Devido às limitações referentes a GWAS e GWS, especialmente pelo alto custo para obtenção e incorporação de informações genômicas oriundas de chips de SNPs em pesquisas e programas de melhoramento genético, informações advindas de simulações são bastante apreciadas como uma alternativa viável para estudos com características complexas e de mensuração tardia, possibilitando a comparação de métodos genômicos e sua adequação a estruturas populacionais, detecção de QTLs, entre outros (HICKEY; GORJANC, 2012).

Uma ferramenta interessante para contornar algumas limitações do uso de dados reais, para verificar a viabilidade de aplicação da genômica, consiste no uso de análises de dados simulados, pois as mesmas demandam apenas recursos computacionais e definições adequadas de métodos estatísticos (DAETWYLER et al., 2013). Portanto, o objetivo deste trabalho foi avaliar a eficiência da seleção genômica em ovinos com utilização de diferentes estruturas de treinamento e estratégias de genotipagem no genoma simulado de ovinos.

2. Material e métodos

2.1. Simulações

O presente estudo não necessitou de aprovação da Comissão de Ética no Uso de Animais (CEUA/UFPI), da Universidade Federal do Piauí devido ao fato dos dados serem gerados computacionalmente através de simulações.

2.1.1 Estrutura da população

As simulações foram realizadas com uso do *software* QMSim versão 1.10 (SARGOLZAEI; SCHENKEL, 2009). As populações simuladas foram obtidas em três passos: O primeiro consistiu na simulação da população histórica, composta por 1.500 animais, em iguais proporções entre os sexos, quantidade esta que se manteve constante até a geração 1.000. Após as 20 gerações seguintes, o efetivo populacional decaiu para 600 animais, também em iguais proporções entre os sexos. Esta etapa foi realizada com o intuito de gerar o desequilíbrio de ligação na população histórica a partir do efeito de mutação-deriva.

Para o segundo passo, que consistiu na formação da população fundadora, foram selecionados os 600 indivíduos da última geração da população histórica, com igual proporção entre machos e fêmeas (300 animais de cada sexo). Estes indivíduos foram acasalados segundo os critérios: 3 progênies por fêmea por geração, 8 gerações de acasalamento aleatório com união aleatória de gametas, crescimento exponencial das fêmeas e sem descarte de animais. A terceira etapa (processo de expansão da população) foi realizada a partir de animais da população fundadora, sendo 400 machos e 10.000 fêmeas da última geração da população anterior. Foram consideradas duas progênies por fêmea e 8 gerações de acasalamento aleatório. A taxa de reposição de machos e fêmeas foi de 60% e 20%, respectivamente. Os animais foram selecionados e descartados com base no fenótipo. Ao final desse processo, 170.010 animais foram obtidos (Figura 1), dos quais 60.000 indivíduos provenientes das três últimas gerações foram selecionados.

2.1.2 Genoma

Um painel contendo 576.595 SNPs foi simulado, através de informações obtidas no *SNPchiMp* versão 3 (NICOLAZZI et al. 2015) e, a partir dos arquivos gerados, foram formados dois painéis de densidades distintas, um de baixa (12k) e um de média densidade (50k), contendo 11.385 e 52.326 marcadores, respectivamente, os quais foram utilizados como base para as análises desta pesquisa. As informações sobre número de SNPs por cromossomo foram baseadas nos trabalhos de Bolormaa et al. (2015), para o

painel de baixa densidade e Alvarenga et al. (2018) para o painel de média densidade. Uma variância fenotípica de 5,61 e herdabilidade de 0,30 (SENA, 2020) foram considerados como parâmetros globais na simulação. A herdabilidade dos QTLs utilizada foi de 0,18 (60% da herdabilidade de AOL) para assumir a pressuposição de que os efeitos dos QTLs têm contribuição expressiva na variância fenotípica da característica.

O genoma foi constituído de 26 cromossomos autossômicos distribuídos de forma equidistante com um tamanho total de 2.656 centiMorgans, baseado em informações do genoma ovino contidas na plataforma NCBI (<https://www.ncbi.nlm.nih.gov/genome/?term=Ovis+aries>), *Assembly Oar_rambouillet* versão 1.0. Foi considerada a presença de 2.014 QTLs, obtidos a partir de informações da plataforma *SheepQTLdb*, (<https://www.animalgenome.org/cgi-bin/QTLdb/OA/index>), onde constavam 12 QTLs reais referentes a AOL, distribuídos aleatoriamente nos cromossomos com efeito de distribuição gama e parâmetro Shape 0,4. A taxa de mutação recorrente para marcadores e QTLs foi de 1×10^{-9} . As taxas de genótipos faltantes e de erros de genotipagem foram de 1×10^{-2} e 5×10^{-3} , respectivamente.

2.1.3 Cenários de estudo

Os animais obtidos nas três últimas gerações de acasalamento aleatório das simulações foram utilizados nas análises posteriores. De acordo com as informações genotípicas, fenotípicas e de pedigree, os animais foram agrupados em 12 cenários, sendo 6 relacionados a cada um dos painéis utilizados e considerando estimativas de herdabilidade moderada para a característica simulada.

A partir de 15 mil animais selecionados aleatoriamente das três últimas gerações pelo *chip* simulado de 600k (para que os SNPs nos painéis estejam nas mesmas posições nos cromossomos), selecionou-se primeiramente os marcadores que formaram o painel de média densidade e, a partir deste, o painel de baixa densidade, tomando como base o arquivo de mapa e a posição de cada marcador no cromossomo em que estão localizados, sendo selecionados aleatoriamente. Posteriormente, 15 mil animais dos painéis de 12k e 50k funcionaram como base para a formação dos cenários com animais que continham todas as informações de genótipo, fenótipo e pedigree. Cenários adicionais foram gerados a partir dos animais de cada cenário subsequente, sendo constituídos de metade dos genótipos dos ovinos e todas as informações de fenótipo e pedigree.

Para cada painel, tamanhos populacionais de ovinos, respectivamente, 500, 2.000 e 5.000 animais, foram agrupados com todos os animais genotipados e com todas as

informações de fenótipo e pedigree, formando os cenários de estudo. Os animais selecionados na sexta e sétima geração formaram as populações de treinamento ou referência, já os indivíduos referentes à oitava geração constituíram a população de validação. O pedigree analisado consistiu de todos os animais das três últimas gerações (60 mil animais). Os cenários formados estão apresentados na Tabela 1.

Tabela 1. Cenários utilizados para seleção genômica com dados simulados em ovinos

Cenários	50k		Cenários	12k	
	Pop. Treinamento	Pop. Validação		Pop. Treinamento	Pop. validação
C1	3.750 ^t	1.250 ^g	C7	3.750 ^t	1.250 ^g
C2	3.750 ^m	1.250 ^g	C8	3.750 ^m	1.250 ^g
C3	1.500 ^t	500 ^g	C9	1.500 ^t	500 ^g
C4	1.500 ^m	500 ^g	C10	1.500 ^m	500 ^g
C5	375 ^t	125 ^g	C11	375 ^t	125 ^g
C6	375 ^m	125 ^g	C12	375 ^m	125 ^g

^t fenótipos e genótipos; ^g todos os animais genotipados; ^m todos os fenótipos e metade dos genótipos do cenário anterior.

2.2. Modelo

Foi utilizado o método GBLUP de passo único (*single-step* GBLUP ou ssGBLUP), proposto com base nos trabalhos de Legarra et al. (2009), Misztal et al. (2009), Aguilar et al. (2010) e Christensen e Lund (2010) para realização predição dos GEBVs. Essa metodologia baseia-se no uso da matriz de parentesco **H**, resultante da combinação das matrizes **A** (parentesco de pedigree) e matriz **G** (parentesco genômico). A inversa da matriz **H** foi calculada de acordo com Aguilar et al. (2010), como:

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

em que \mathbf{A}^{-1} é a inversa da matriz de parentesco com base em informações de pedigree, \mathbf{G}^{-1} é a inversa da matriz de parentesco genômico e \mathbf{A}_{22}^{-1} a inversa da matriz de parentesco baseada em pedigree apenas para os animais genotipados.

No modelo unicaracterístico, foram considerados: como efeitos fixos, o sexo dos animais, com dois níveis, e geração, com três níveis; e como efeito aleatório, os animais e o resíduo, com níveis variando de acordo com o cenário. Para as estimativas de componentes de (co) variância e valores genéticos, foi utilizado o seguinte modelo:

$$y = \mathbf{X}\beta + \mathbf{Z}\alpha + \mathbf{e}$$

em que: y é o vetor de observações para a característica AOL; β é o vetor de efeitos fixos; α é o vetor de efeitos genéticos aditivos, X é a matriz de incidência dos efeitos fixos; Z é a matriz de incidência dos efeitos genéticos aditivos; e é o vetor de efeitos residuais, de modo que $e \sim N(0, I\sigma_e^2)$.

2.3 Análises estatísticas

Os bancos de dados gerados das simulações foram editados com auxílio do software R versão 3.6.1 (R CORE TEAM, 2019) para utilização nos programas da família BLUPf90 (MISZTAL et al., 2019). O controle de qualidade dos genótipos foi feito com base nos critérios *Call rate* para marcadores (<95%), *call rate* por animal (<0,90), Frequência do alelo menos frequente (MAF <0,01) e desvios de equilíbrio de *Hardy-Weinberg* ($p < 10^{-15}$), utilizando o software PreGSf90 (MISZTAL et al., 2019). Nenhuma informação foi excluída com a utilização destes critérios de controle de qualidade.

O software Airemlf90 (MISZTAL et al., 2019) foi utilizado, com as informações dos marcadores, fenótipos e pedigree, para obter os componentes de variância na população, no intuito de verificar se as estimativas de herdabilidade estavam próximas do desejado (aproximadamente 0,30) e para obter os GEBVs (valores genômicos preditos) utilizando o programa Blupf90 (MISZTAL et al., 2019).

O programa Blupf90 (MISZTAL et al., 2019) foi utilizado para obter a predição dos GEBVs, através do ssGBLUP, na população de treinamento e a correlação destes com os TBVs e o programa Predictf90 (MISZTAL et al., 2019) utiliza-se dos GEBVs gerados no Blupf90 para obter os efeitos dos marcadores estimados.

2.3.1 Comparação dos cenários

A comparação das populações com diferentes quantidades de animais e painéis de baixa e média densidade para as predições foram realizadas no programa R (R CORE TEAM, 2019). Primeiramente, a predição da acurácia foi obtida conforme estabelecido pela correlação de Pearson (acurácia da predição) entre os GEBVs gerados e TBVs na população de validação. A avaliação do grau de inflação e deflação (viés) das estimativas foi obtido através da diferença do coeficiente de regressão estimado (GEBV) do valor unitário.

Testes para verificação da distribuição dos dados foram realizados através do Programa R para verificar a provável distribuição dos efeitos nos cenários. Um modelo linear misto com dois fatores, densidade do painel com dois níveis (50k e 12k) e tamanho

populacional com seis níveis (5.000, 2.000 e 500) e três tamanhos com a metade dos genótipos da população de treinamento dessas quantidades), obtidos por meio de combinações das estratégias de genotipagem com o tamanho das populações de treinamento; a interação entre os dois fatores e a herdabilidade como efeito aleatório, como mostrado a seguir:

$$y_{ijl} = \mu + D_i + T_j + (DT_{ij}) + H + e_{ijl}$$

y_{ijl} : Correlações (acurácias da predição) para 10 repetições;

μ : Média geral das observações;

D_i : densidade do painel, com $i = 1$ ou 2 ;

T_j : tamanho populacional, com $j = 1$ a 6 ;

DT_{ij} : Interação entre densidade do painel e tamanho populacional;

H : Herdabilidade;

e_{ijl} : erro associado as repetições l .

Por fim, foi realizado o teste *SNK* (*Student-Newman-Keuls*) para comparar as médias das estimativas das 10 repetições simuladas a fim de demonstrar se houve diferenças significativas entre os cenários analisados.

3. Resultados

3.1 Parâmetros genéticos

As médias dos componentes de variância e herdabilidade estão apresentadas na Tabela 2. A variância genética aditiva média variou de 1,08 (C2) a 2,14 (C6) nos cenários referentes ao painel de densidade média e de 0,96 (C9) a 1,34 (C11) para as estruturas de treinamento com o painel de baixa densidade. Já as estimativas de variância residual para cenários com uso do painel de 50k variaram de 1,21 (C4) a 4,47 (C2), e para os cenários referentes ao painel de 12k, estas estimativas variaram de 4,24 (C11) a 4,62 (C9). A herdabilidade média para as 10 repetições foi de moderada e alta magnitude, variando de 0,17 (C9) a 0,38 (C6).

Tabela 2. A média das estimativas de componentes de variância, herdabilidade e erro padrão (entre parênteses) obtidas com uso do método ssGBLUP com 2.014 QTLs para os cenários com dois painéis de densidades moderada e baixa e populações com tamanhos e número de animais genotipados distintos

Painel	Cenários	σ_a^2 (SE)	σ_e^2 (SE)	h^2
50k	C1	1,10(0,16)	4,45(0,16)	0,20
	C2	1,08(0,14)	4,47(0,14)	0,19
	C3	1,20(0,25)	4,26(0,25)	0,22
	C4	1,21(0,30)	1,21(0,30)	0,22
	C5	1,24(0,87)	4,34(0,88)	0,22
	C6	2,14(1,13)	3,45(1,10)	0,38
12k	C7	0,98(0,11)	4,59(0,12)	0,18
	C8	1,04(0,14)	4,53(0,15)	0,19
	C9	0,96(0,23)	4,62(0,24)	0,17
	C10	1,06(0,29)	4,55(0,30)	0,19
	C11	1,34(0,80)	4,24(0,80)	0,24
	C12	1,31(1,02)	4,26(1,02)	0,23

σ_a^2 : variância genética aditiva; σ_e^2 : variância residual; h^2 : herdabilidade; SE: erro padrão. Valores referentes a média de 10 repetições dos parâmetros. (C1 e C7) 3.750 fenótipos e genótipos (pop de treinamento) e 1.250 genótipos (pop. de validação); (C2 e C8) 3.750 fenótipos e metade dos genótipos (1.875) de C1 (pop. treinamento) e 1.250 genótipos (pop.validação); (C3 e C9) 1.500 informações de fenótipo e genótipos (pop.treinamento) e 500 genótipos (pop.validação); (C4 e C10) 1.500 fenótipos e metade (750) dos genótipos de C3 (pop. treinamento) e 500 genótipos (pop.validação); (C5 e C11) 375 fenótipos e genótipos (pop.treinamento) e 125 genótipos (pop.validação); (C6 e C12) 375 fenótipos e metade dos genótipos (187) de C5 e 125 genótipos (pop. validação).

3.2 Acurácia da predição genômica

Na Tabela 3, são apresentadas as médias das correlações entre GEBV e TBV e o viés para 10 repetições das análises com o método ssGBLUP para os cenários em estudo. As acurácias para os cenários simulados com painéis de 50k e 12k variaram de 0,14 (C5) a 0,34 (C1) e de 0,10 (C12) a 0,32 (C7), respectivamente. Os maiores valores foram obtidos com uso das estruturas de treinamento com maior quantidade de animais genotipados, enquanto os menores valores foram observados nas populações com menos animais genotipados na estrutura de treinamento.

Os valores de viés médio detectados pelas análises com o procedimento ssGBLUP variaram de -0,464 a 0,023 e de -0,125 a 0,070 para os painéis de 50k e 12k, respectivamente (Tabela 3). Os cenários com 2.000 animais distribuídos nas populações de treinamento e validação (C3, C4, C9 e C10) apresentaram os menores vieses dentre as estruturas de seleção genômica avaliadas. Por outro lado, os maiores vieses foram observados nos cenários C5, C6, C7 e C8.

Tabela 3. Médias das Correlações e viés dos cenários com painéis de 12k e 50k marcadores SNPs simulados*

Painel	Cenários	Corr (y,yhat)	Viés
50k	1	0,34	-0,063
	2	0,29	-0,054
	3	0,24	0,023
	4	0,21	-0,006
	5	0,14	-0,373
	6	0,16	-0,464
12k	7	0,32	-0,106
	8	0,27	-0,125
	9	0,22	0,0397
	10	0,21	0,0705
	11	0,13	-0,092
	12	0,10	-0,104

*Correlações entre TBV e GEBV e viés como média de 10 repetições das análises de cada cenário. (1) 3.750 fenótipos e genótipos (pop de treinamento) e 1.250 genótipos (pop. de validação); (2) 3.750 fenótipos e metade dos genótipos (1.875) de C1 (pop. treinamento) e 1.250 genótipos (pop.validação); (3) 1.500 informações de fenótipo e genótipos (pop.treinamento) e 500 genótipos (pop.validação); (4) 1.500 fenótipos e metade (750) dos genótipos de C3 (pop. treinamento) e 500 genótipos (pop.validação); (5) 375 fenótipos e genótipos (pop.treinamento) e 125 genótipos (pop.validação); (6) 375 fenótipos e metade dos genótipos (187) de C5 e 125 genótipos (pop. validação).

3.3 Comparações entre estruturas populacionais

Os testes para verificar a possível distribuição teórica dos dados indicaram que os resultados para as estruturas populacionais assumiram distribuição normal. Com isso, adotou-se um modelo linear misto com dois fatores de variação, ou seja, a densidade do painel com dois níveis (50k e 12k) e o a combinação de tamanho de treinamento e estratégias de genotipagem com seis níveis (5.000, 2.000 e 500 e a metade dos animais de treinamento para cada um destes), a interação entre os fatores, como efeito aleatório, e a herdabilidade. A análise de variância evidenciou que apenas os tamanhos populacionais apresentaram diferenças significativas ($p < 0,05$) pelo teste SNK (Tabela 4).

Tabela 4. Análise de variância para os efeitos de densidade dos painéis, tamanho populacional de treinamento e para a interação entre ambos

	GL	SQm	QM	F	P valor (0,05)
Densidade do painel	1	0,00504	0,005040	2,2035	0,1405108
Tamanho de treinamento	5	0,60723	0,303613	132,7431	0,000000*
Densidade x tamanho treinamento	5	0,00683	0,003417	1,4940	0,2289152
Total	11	0,6191	0,31207		

GL, graus de liberdade; SQm, soma de quadrados médios; QM, quadrado médio; F, estatística F de Fisher; *, diferença significativa ($p < 0,05$) pelo teste F.

Observou-se, pelo teste de comparação de média SNK, que não houve diferenças significativas na acurácia preditiva ao se utilizar os painéis de SNPs com 50k e 12k ($p < 0,05$). Diferenças significativas foram observadas entre os diferentes tamanhos da população de treinamento, porém, referente aos cenários com o treinamento com menor quantidade de informações de genótipos e fenótipos, não foi verificada diferença significativa entre os resultados obtidos com uso dos dois painéis (Tabela 5).

Tabela 5. Acurácia da predição (Corr) e desvios padrão (SD) para cenários com tamanhos populacionais com diferentes quantidades de animais genotipados na população de treinamento.

	C1		C2		C3		C4		C5		C6							
	Corr	SD		Corr	SD	Corr	SD	Corr	SD	Corr	SD	Corr	SD					
50k	0,340	0,038	a	0,302	0,038	abc	0,260	0,051	cd	0,219	0,049	de	0,108	0,071	F	0,107	0,068	f
	C7		C8		C9		C10		C11		C12							
12k	0,315	0,045	ab	0,274	0,048	bc	0,224	0,047	de	0,207	0,063	e	0,129	0,049	F	0,096	0,036	f

Letras diferentes na mesma linha ou coluna indicam diferenças significativas ($p < 0,05$) pelo teste de SNK (*Student-Newman-Keuls*).

4. Discussão

O aumento observado nas estimativas de herdabilidade em cenários com maior quantidade de SNPs simulados pode ser justificado pelo aumento nos valores dos componentes de variância aditiva e diminuição da variância residual nos cenários com uso do chip de 50k. Neste estudo, as estimativas de herdabilidades foram classificadas como baixa ($< 0,15$), moderada (de $0,15$ a $0,30$) e alta ($> 0,30$), para fins de comparação (MASSENDER et al., 2019). Assim, a magnitude das estimativas médias de herdabilidade corrobora com os valores inicialmente simulados para a característica utilizada como base ($0,30$), com exceção do C6 ($h^2 = 0,38$), que apresentou herdabilidade de magnitude alta.

A utilização do painel com maior densidade de SNPs simulados, em comparação aos cenários utilizando o painel de 12k, resultou em maiores estimativas de herdabilidade de magnitude moderada. Com a utilização de painéis de marcadores com maior densidade, é esperado que ocorra aumento na quantidade de variância genética explicada pelos marcadores, devido ao aumento do desequilíbrio de ligação entre marcadores e QTLs (JENSEN; SU; MADSEN, 2012).

Além do aumento nas estimativas de componentes de variância genética, a quantidade de informações de genótipos está diretamente relacionada com os valores de acurácia. Isto justifica as maiores acurácias nos cenários com uso de estruturas de treinamento com maior quantidade de animais genotipados. As estimativas de acurácia obtidas foram influenciadas pelo fator tamanho populacional e não pela densidade de marcadores adotada. De acordo com Lee; Clark; Van Der Werf (2017), o tamanho da população de treinamento é um dos fatores que afetam as estimativas de predição genômica. Maiores acurácias são geralmente esperadas quando há maior quantidade de animais genotipados. Além disso, com a utilização do método *single-step*, as estimativas de componentes de variância e acurácia são favorecidas em situações em que nem todos os animais são genotipados, pois este método possibilita a utilização combinada de informações de animais genotipados e não-genotipados (VEERKAMP et al., 2011).

Os valores da acurácia de predição foram considerados de baixa e moderada magnitudes, adotando-se como baixa ($< 0,30$), moderada (entre $0,30$ e $0,70$) e alta ($> 0,70$) (MASSENDER et al., 2019), sendo similares entre os cenários em que foram utilizados tamanhos de estruturas de treinamento semelhantes com uso dos dois painéis de SNPs avaliados. Em predições genômicas, com a utilização de dados simulados, têm-se a vantagem de se poder verificar as correlações (acurácia de predição) entre os GEBVs e

TBV's e obter os coeficientes de regressão e o viés estimado e esperado pelas análises dos dados, assim pode-se ter uma noção da precisão das estimativas. As estimativas de correlação entre GEBV e TBV são fundamentais para se ter uma boa resposta à seleção (MACEDO et al., 2020).

Os valores de acurácia obtidos neste trabalho foram menores do que os relatados por Gowane et al. (2019), que também observaram aumento de acurácia com a inclusão de informações genômicas nos cenários avaliados. Estes autores obtiveram acurácias de 0,59 e 0,68 para cenários com 500 e 2.000 animais genotipados na população de referência e chegaram ao valor de 0,79 com 9.550 animais contendo todas as informações genotípicas, com desvios de aproximadamente 0,01, respectivamente. Adicionalmente, os mesmos autores verificaram que houve um aumento na precisão das estimativas conforme aumentaram a quantidade de QTLs, devido à redução dos efeitos destes quando em maior quantidade, devido à distribuição gama adotada nas simulações.

No presente estudo, os cenários com maiores estimativas de herdabilidade e menores acurácias apresentaram os maiores vieses de predição. O viés representa a diferença entre os valores genéticos estimados (\hat{u}) e os valores genéticos verdadeiros (u) (MACEDO et al., 2020). Valores de viés mais próximos de 1 indicam que os GEBVs estimados são condizentes com os valores genéticos verdadeiros (TBVs), porém as estimativas podem ser menores que 1 (superestimadas) ou maiores que 1 (subestimadas). Assim, neste estudo, as estimativas foram superestimadas e subestimadas, pois os valores de viés obtidos foram bem abaixo e acima de 1, respectivamente.

De acordo com Macedo et al. (2020), o viés pode super ou subestimar a tendência genética e levar a decisões equivocadas por melhoristas ou produtores. Além disso, as inclinações dos coeficientes de regressão dos TBVs sobre os GEBVs abaixo ou acima de 1 podem ter como consequência a sub ou superestimação do mérito genético dos animais. Os maiores desvios em relação ao valor 1 (valor de correlação esperado entre TBV e GEBV) podem ser devidos a problemas no modelo de seleção (DAETWYLER et al., 2013). Por exemplo, quando a seleção ocorre na população de referência, as estimativas dos valores genéticos genômicos, em populações apenas com animais genotipados, por meio do método BLUP genômico (GBLUP) se tornam passíveis de viés (GOWANE e al., 2019).

Há uma tendência em aumento da acurácia com o acréscimo de animais genotipados na população de treinamento ou referência. Além disso, os parâmetros para obtenção de predições acuradas dependem da extensão do DL entre marcadores e QTLs,

distribuição dos efeitos dos QTLs, métodos estatísticos, entre outros (HAYES et al., 2009). Em trabalhos futuros, recomenda-se simular populações com menores variações entre as estimativas de herdabilidade, pois este parâmetro genético influencia os resultados finais das predições genômicas.

5. Conclusões

A utilização de um painel de marcadores de média densidade proporcionou maiores estimativas de variância genética e herdabilidade em relação ao uso de um painel de baixa densidade, na maioria dos cenários avaliados. No entanto, a acurácia de predição foi mais influenciada pela quantidade de animais genotipados na população de referência do que pela densidade dos painéis. Do mesmo modo, estimativas menos viesadas foram obtidas em cenários com maior número de animais genotipados na população de referência. Há uma tendência em aumento da acurácia com o acréscimo de animais genotipados na população de treinamento ou referência.

Referências

- AGUILAR, I. et al. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. **Journal of Dairy Science**, v. 93, n. 2, p. 743-752, 2010.
- ALVARENGA, A. B. et al. Linkage disequilibrium in Brazilian Santa Inês breed, *Ovis aries*. **Scientific Reports**, v. 8, n. 8851, p. 1-11, 2018.
- BOLOORMA, S. et al. A genome-wide association study of meat and carcass traits in Australian cattle. *Journal Animal Science*, v. 89, p. 2297-2309, 2011.
- CHRISTENSEN, O. F.; LUND, M. S. Genomic prediction when some animals are not genotyped. **Genetics Selection Evolution**, v. 42, n. 2, p. 1-8, 2010.
- DAETWYLER, H. D. et al. Genomic prediction in animals and plants: simulation of data, validation, reporting, and benchmarking. **Genetics**, v. 193, n. 2, p. 347-365, 2013.
- DE LOS CAMPOS, G. et al. Whole-genome regression and prediction methods applied to plant and animal breeding. **Genetics**, v. 193, p. 327-345, 2013.
- GOWANE, G. R. et al. Effect of selection and selective genotyping for creation of reference on bias and accuracy of genomic prediction. **Journal of Animal Breeding and Genetics**, v. 00, p. 1-18, 2019.
- HAYES, B. J. et al. Invited review: Genomic selection in dairy cattle: progress and challenges. **J Dairy Sci**, v.92, p. 433-443, 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. **G3 Genes, Genomics, Genetics**, v. 2, 2012.
- JENSEN, J.; SU, G.; MADSEN, P. Partitioning additive genetic variance into genomic and remaining polygenic components for complex traits in dairy cattle. **BMC Genetics**, v. 13, n. 44, 2012.
- LEE, S. H.; CLARK, S.; VAN DER WERF, J. H. J. Estimation of genomic prediction accuracy from reference populations with varying degrees of relationship. **PLoS ONE**, 12, e0189775, 2017.
- LEGARRA, A.; AGUILAR, I.; MISZTAL, I. A relationship matrix including full pedigree and genomic information. **Journal of Dairy Science**, v. 92, p. 4656-4663, 2009.
- MACEDO, F. L. et al. Bias and accuracy of dairy sheep evaluations using BLUP and SSGBLUP with metafounders and unknown parent groups. **Genetics Selection Evolution**, v. 52, n. 47, 2020.
- MASSENDER et al. A genetic evaluation of growth, ultrasound, and carcass traits at alternative slaughter endpoints in crossbred heavy lambs. **J. Anim. Sci**, v. 97, p. 521-535, 2019.

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

MEUWISSEN, T.; HAYES, B. J.; GODDARD, M. E. Genomic selection: A paradigm shift in animal breeding. **Animal Frontiers**, v.6, n.1, p. 6-14, 2016.

MEUWISSEN, T.; HAYES, B.; GODDARD, M. Accelerating improvement of livestock with genomic selection. **Annual Review of Animal Biosciences**, v. 1, p. 221-37, 2013.

MISZTAL, I. et al. **Manual for BLUPF90 family of programs**. Disponível em: <http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all7.pdf>. Acessado em: 15 de Maio de 2019).

MISZTAL, I.; LEGARRA, A.; AGUILAR, I. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. **Journal of Dairy Science**, v. 92, n. 9, p. 4648–4655, 2009.

MRODE, R. et al. Invited review: Genomic selection for small ruminants in developed countries: how applicable for the rest of the world? **Animal**, v. 12, n. 7, p. 1333-1340, 2018.

NICOLAZZI, E. L. et al. SNPchiMp v.3: integrating and standardizing single nucleotide polymorphism data for livestock species. **BMC Genomics**, v. 16, n. 283, 2015.

R CORE TEAM (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

RUPP, R. et al. Genomic application in sheep and goat breeding. **Animal Frontiers**, v. 6, n. 1, p. 39-44, 2016.

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

SENA, L. S. et al. Genetic evaluation of tropical climate-adapted sheep for carcass traits including genomic information. **Small Ruminant Research**, v. 188, n. 106120, 2020.

SIMIANER, H. Genomic and other revolutions – why some technologies are quickly adopted and others are not. **Animal Frontiers**, v. 6, n. 1, p. 53-58, 2016.

TOOSI, A.; FERNANDO, R.L.; DEKKERS, J.C. Genomic selection in admixed and crossbred populations. **Journal of Animal Science**, v. 88, n. 1, p. 32-46, 2010.

VEERKAMP, R. F. et al. Genomic and pedigree-based genetic parameters for scarcely recorded traits when some animals are genotyped. **Journal of Dairy Science**, v. 94, n. 8, p. 4189-4197, 2011.

CONSIDERAÇÕES FINAIS

A pesquisa apresentou aspectos importantes no âmbito de estudos de associação e seleção genômica com o uso do método ssGBLUP com dados simulados do genoma ovino. Observou-se a importância da quantidade de informações disponíveis sobre as populações em estudo para a obtenção resultados mais acurados.

O ssGWAS pela abordagem do p -valor exato foi eficiente na detecção de associações verdadeiras entre SNPs significativos e QTLs associados a uma característica de moderada magnitude, em populações simuladas de ovinos. Porém, a adoção de critérios bem definidos de comparação dos cenários, a verificação de possíveis associações espúrias, a utilização de dados simulados com herdabilidades de alta e baixa magnitudes, com menores variações entre essas estimativas, bem como a definição de populações com base no tamanho efetivo e o estudo do decaimento de desequilíbrio de ligação, devem ser realizados em trabalhos futuros.

O aumento da acurácia foi verificado com o acréscimo de animais genotipados na população de treinamento. Em ambos os capítulos, verificou-se que a maior quantidade de animais genotipados nos cenários propiciou melhor detecção de associações e predições genômicas. Estes resultados sugerem que o número mínimo de animais com informações genóticas e fenóticas seria de 2.000 e um painel de média densidade geram informações suficientes para programas de melhoramento genético de ovinos, no Brasil.

ANEXOS

Cartão de parâmetros para simulação no *software QMSim* baseado no genoma ovino com painel de alta densidade e 2.014 QTLs

```

/*****
** Global parameters **
*****/
title = "Ovinos600k_2mil QTLs";
nrep = 10;           //Number of replicates
h2 = 0.30;          //Heritability
qtlh2 = 0.18; //QTL heritability
phvar = 5.61; //Phenotypic variance
nthread = 1;
/*****
** Historical population **
*****/
begin_hp;
hg_size = 1500 [0] 1500 [1000] 600 [1020]; //Size of the historical generations
//nmfhg = 750; //Number of males in the first generation
nmlhg = 300; //Number of males in the last generation
end_hp;
/*****
** Populations **
*****/
begin_pop = "EG";
begin_founder;
  male [n = 300, pop = "hp"];
  female [n = 300, pop = "hp"];
end_founder;
ng = 8;
ls = 3;
md = rnd_ug;
sr = 0.00 1.00;
dr = 0.00 1.00/e;
begin_popoutput;
stat;
end_popoutput;
end_pop;

begin_pop = "POP1";
begin_founder;
  male [n = 400, pop = "EG", gen = 8];
  female [n = 10000, pop = "EG", gen = 8];
end_founder;
ng = 8;
ls = 2;
md = rnd;
sr = 0.60 0.00;
dr = 0.20 0.00;
sd = phen;

```

```

cd = phen;
ebv_est = blup;
begin_popoutput;
stat;
ld /maft 0.05 /gen 6 7 8;
data;
genotype /snp_code /gen 6 7 8;
allele_freq;
end_popoutput;
end_pop;

/*****
** Genome **
*****/
begin_genome;
  begin_chr = 1; // CHR 1 //number of chromosomes with the same parameter
  chrln = 301; //Chromosome length cM
  nmloci = 64360; //Number of markers
  mpos = even; //Marker positions
  nma = all 2; //Number of marker alleles
  maf = eql; //Marker allele frequencies
  nqloci = 53; //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 ~ Gamma
  end_chr;
//
begin_chr = 1; // CHR 2 //number of chromosomes with the same parameter
  chrln = 266; //Chromosome length cM
  nmloci = 57659; //Number of markers
  mpos = even; //Marker positions
  nma = all 2; //Number of marker alleles
  maf = eql; //Marker allele frequencies
  nqloci = 265; //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 ~ Gamma
  end_chr;
//
begin_chr = 1; // CHR 3 //number of chromosomes with the same parameter
  chrln = 241; //Chromosome length cM
  nmloci = 53250; //Number of markers
  mpos = even; //Marker positions
  nma = all 2; //Number of marker alleles
  maf = eql; //Marker allele frequencies
  nqloci = 53; //Number of QTL
  qpos = rnd;
  nqa = all 2; //Number of QTL alleles

```



```

qaf = eql;           //QTL allele frequencies
qae = rndg 0.4;     //QTL allele effects ~ Gamma
end_chr;
//
begin_chr = 1; // CHR 4 //number of chromosomes with the same parameter
chrln = 130;       //Chromosome length cM
nmloci = 27607;   //Number of markers
mpos = even;      //Marker positions
nma = all 2;      //Number of marker alleles
maf = eql;        //Marker allele frequencies
nqloci = 53;      //Number of QTL
qpos = rnd;
nqa = all 2;      //Number of QTL alleles
qaf = eql;        //QTL allele frequencies
qae = rndg 0.4;   //QTL allele effects ~ Gamma
end_chr;
//
begin_chr = 1; // CHR 5 //number of chromosomes with the same parameter
chrln = 118;      //Chromosome length cM
nmloci = 25429;   //Number of markers
mpos = even;      //Marker positions
nma = all 2;      //Number of marker alleles
maf = eql;        //Marker allele frequencies
nqloci = 53;      //Number of QTL
qpos = rnd;
nqa = all 2;      //Number of QTL alleles
qaf = eql;        //QTL allele frequencies
qae = rndg 0.4;   //QTL allele effects ~ Gamma
end_chr;
//
begin_chr = 1; // CHR 6 //number of chromosomes with the same parameter
chrln = 130;      //Chromosome length cM
nmloci = 26661;   //Number of markers
mpos = even;      //Marker positions
nma = all 2;      //Number of marker alleles
maf = eql;        //Marker allele frequencies
nqloci = 106;     //Number of QTL
qpos = rnd;
nqa = all 2;      //Number of QTL alleles
qaf = eql;        //QTL allele frequencies
qae = rndg 0.4;   //QTL allele effects ~ Gamma
end_chr;
//
begin_chr = 1; // CHR 7 //number of chromosomes with the same parameter
chrln = 108;      //Chromosome length cM
nmloci = 23561;   //Number of markers
mpos = even;      //Marker positions
nma = all 2;      //Number of marker alleles
maf = eql;        //Marker allele frequencies
nqloci = 159;     //Number of QTL

```

```

qpos = rnd;
nqa = all 2;           //Number of QTL alleles
qaf = eql;           //QTL allele frequencies
qae = rndg 0.4;      //QTL allele effects ~ Gamma
end_chr;
//
begin_chr = 1; // CHR 8 //number of chromosomes with the same parameter
chrln = 99;         //Chromosome length cM
nmloci = 20920;     //Number of markers
mpos = even;       //Marker positions
nma = all 2;       //Number of marker alleles
maf = eql;         //Marker allele frequencies
nqloci = 53;       //Number of QTL
qpos = rnd;
nqa = all 2;       //Number of QTL alleles
qaf = eql;         //QTL allele frequencies
qae = rndg 0.4;   //QTL allele effects ~ Gamma
end_chr;
//
begin_chr = 1; // CHR 9 //number of chromosomes with the same parameter
chrln = 105;       //Chromosome length cM
nmloci = 21568;    //Number of markers
mpos = even;       //Marker positions
nma = all 2;       //Number of marker alleles
maf = eql;         //Marker allele frequencies
nqloci = 265;     //Number of QTL
qpos = rnd;
nqa = all 2;       //Number of QTL alleles
qaf = eql;         //QTL allele frequencies
qae = rndg 0.4;   //QTL allele effects ~ Gamma
end_chr;
//
begin_chr = 1; // CHR 10 //number of chromosomes with the same parameter
chrln = 97;        //Chromosome length cM
nmloci = 19678;    //Number of markers
mpos = even;       //Marker positions
nma = all 2;       //Number of marker alleles
maf = eql;         //Marker allele frequencies
nqloci = 53;       //Number of QTL
qpos = rnd;
nqa = all 2;       //Number of QTL alleles
qaf = eql;         //QTL allele frequencies
qae = rndg 0.4;   //QTL allele effects ~ Gamma
end_chr;
//
begin_chr = 1; // CHR 11 //number of chromosomes with the same parameter
chrln = 61;        //Chromosome length cM
nmloci = 15913;    //Number of markers
mpos = even;       //Marker positions
nma = all 2;       //Number of marker alleles

```

```

maf = eql; //Marker allele frequencies
nqloci = 53; //Number of QTL
qpos = rnd;
nqa = all 2; //Number of QTL alleles
qaf = eql; //QTL allele frequencies
qae = rndg 0.4; //QTL allele effects ~ Gamma
end_chr;
//
begin_chr = 1; // CHR 12 //number of chromosomes with the same parameter
chrln = 84; //Chromosome length cM
nmloci = 18937; //Number of markers
mpos = even; //Marker positions
nma = all 2; //Number of marker alleles
maf = eql; //Marker allele frequencies
nqloci = 53; //Number of QTL
qpos = rnd;
nqa = all 2; //Number of QTL alleles
qaf = eql; //QTL allele frequencies
qae = rndg 0.4; //QTL allele effects ~ Gamma
end_chr;
//
begin_chr = 1; // CHR 13 //number of chromosomes with the same parameter
chrln = 87; //Chromosome length cM
nmloci = 19785; //Number of markers
mpos = even; //Marker positions
nma = all 2; //Number of marker alleles
maf = eql; //Marker allele frequencies
nqloci = 53; //Number of QTL
qpos = rnd;
nqa = all 2; //Number of QTL alleles
qaf = eql; //QTL allele frequencies
qae = rndg 0.4; //QTL allele effects ~ Gamma
end_chr;
//
begin_chr = 1; // CHR 14 //number of chromosomes with the same parameter
chrln = 71; //Chromosome length cM
nmloci = 15991; //Number of markers
mpos = even; //Marker positions
nma = all 2; //Number of marker alleles
maf = eql; //Marker allele frequencies
nqloci = 53; //Number of QTL
qpos = rnd;
nqa = all 2; //Number of QTL alleles
qaf = eql; //QTL allele frequencies
qae = rndg 0.4; //QTL allele effects ~ Gamma
end_chr;
//
begin_chr = 1; // CHR 15 //number of chromosomes with the same parameter
chrln = 90; //Chromosome length cM
nmloci = 19255; //Number of markers

```

```

mpos = even;           //Marker positions
nma = all 2;           //Number of marker alleles
maf = eql;             //Marker allele frequencies
nqloci = 53;           //Number of QTL
qpos = rnd;
nqa = all 2;           //Number of QTL alleles
qaf = eql;             //QTL allele frequencies
qae = rndg 0.4;        //QTL allele effects ~ Gamma
end_chr;
//
begin_chr = 1; // CHR 16 //number of chromosomes with the same parameter
chrln = 78;           //Chromosome length cM
nmloci = 16499;        //Number of markers
mpos = even;           //Marker positions
nma = all 2;           //Number of marker alleles
maf = eql;             //Marker allele frequencies
nqloci = 53;           //Number of QTL
qpos = rnd;
nqa = all 2;           //Number of QTL alleles
qaf = eql;             //QTL allele frequencies
qae = rndg 0.4;        //QTL allele effects ~ Gamma
end_chr;
//
begin_chr = 1; // CHR 17 //number of chromosomes with the same parameter
chrln = 82;           //Chromosome length cM
nmloci = 16671;        //Number of markers
mpos = even;           //Marker positions
nma = all 2;           //Number of marker alleles
maf = eql;             //Marker allele frequencies
nqloci = 53;           //Number of QTL
qpos = rnd;
nqa = all 2;           //Number of QTL alleles
qaf = eql;             //QTL allele frequencies
qae = rndg 0.4;        //QTL allele effects ~ Gamma
end_chr;
//
begin_chr = 1; // CHR 18 //number of chromosomes with the same parameter
chrln = 71;           //Chromosome length cM
nmloci = 16129;        //Number of markers
mpos = even;           //Marker positions
nma = all 2;           //Number of marker alleles
maf = eql;             //Marker allele frequencies
nqloci = 106;          //Number of QTL
qpos = rnd;
nqa = all 2;           //Number of QTL alleles
qaf = eql;             //QTL allele frequencies
qae = rndg 0.4;        //QTL allele effects ~ Gamma
end_chr;
//
begin_chr = 1; // CHR 19 //number of chromosomes with the same parameter

```

```

chrln = 63;           //Chromosome length cM
nmloci = 14540;      //Number of markers
mpos = even;        //Marker positions
nma = all 2;        //Number of marker alleles
maf = eql;          //Marker allele frequencies
nqloci = 53;        //Number of QTL
qpos = rnd;
nqa = all 2;        //Number of QTL alleles
qaf = eql;          //QTL allele frequencies
qae = rndg 0.4;     //QTL allele effects ~ Gamma
end_chr;
//
begin_chr = 1; // CHR 20 //number of chromosomes with the same parameter
chrln = 56;         //Chromosome length cM
nmloci = 12583;     //Number of markers
mpos = even;       //Marker positions
nma = all 2;       //Number of marker alleles
maf = eql;         //Marker allele frequencies
nqloci = 53;       //Number of QTL
qpos = rnd;
nqa = all 2;       //Number of QTL alleles
qaf = eql;         //QTL allele frequencies
qae = rndg 0.4;   //QTL allele effects ~ Gamma
end_chr;
//
begin_chr = 1; // CHR 21 //number of chromosomes with the same parameter
chrln = 53;         //Chromosome length cM
nmloci = 11896;    //Number of markers
mpos = even;       //Marker positions
nma = all 2;       //Number of marker alleles
maf = eql;         //Marker allele frequencies
nqloci = 53;       //Number of QTL
qpos = rnd;
nqa = all 2;       //Number of QTL alleles
qaf = eql;         //QTL allele frequencies
qae = rndg 0.4;   //QTL allele effects ~ Gamma
end_chr;
//
begin_chr = 1; // CHR 22 //number of chromosomes with the same parameter
chrln = 56;         //Chromosome length cM
nmloci = 12097;    //Number of markers
mpos = even;       //Marker positions
nma = all 2;       //Number of marker alleles
maf = eql;         //Marker allele frequencies
nqloci = 53;       //Number of QTL
qpos = rnd;
nqa = all 2;       //Number of QTL alleles
qaf = eql;         //QTL allele frequencies
qae = rndg 0.4;   //QTL allele effects ~ Gamma

```

```

end_chr;
//
begin_chr = 1; // CHR 23 //number of chromosomes with the same parameter
chrln = 68; //Chromosome length cM
nmloci = 14178; //Number of markers
mpos = even; //Marker positions
nma = all 2; //Number of marker alleles
maf = eql; //Marker allele frequencies
nqloci = 53; //Number of QTL
qpos = rnd;
nqa = all 2; //Number of QTL alleles
qaf = eql; //QTL allele frequencies
qae = rndg 0.4; //QTL allele effects ~ Gamma
end_chr;
//
begin_chr = 1; // CHR 24 //number of chromosomes with the same parameter
chrln = 43; //Chromosome length cM
nmloci = 10458; //Number of markers
mpos = even; //Marker positions
nma = all 2; //Number of marker alleles
maf = eql; //Marker allele frequencies
nqloci = 53; //Number of QTL
qpos = rnd;
nqa = all 2; //Number of QTL alleles
qaf = eql; //QTL allele frequencies
qae = rndg 0.4; //QTL allele effects ~ Gamma
end_chr;
//
begin_chr = 1; // CHR 25 //number of chromosomes with the same parameter
chrln = 48; //Chromosome length cM
nmloci = 10824; //Number of markers
mpos = even; //Marker positions
nma = all 2; //Number of marker alleles
maf = eql; //Marker allele frequencies
nqloci = 53; //Number of QTL
qpos = rnd;
nqa = all 2; //Number of QTL alleles
qaf = eql; //QTL allele frequencies
qae = rndg 0.4; //QTL allele effects ~ Gamma
end_chr;
//
begin_chr = 1; // CHR 26 //number of chromosomes with the same parameter
chrln = 49; //Chromosome length cM
nmloci = 10146; //Number of markers
mpos = even; //Marker positions
nma = all 2; //Number of marker alleles
maf = eql; //
nqloci = 53; //Number of QTL
qpos = rnd;
nqa = all 2; //Number of QTL alleles

```

```
qaf = eq1;           //QTL allele frequencies
qae = rndg 0.4;      //QTL allele effects ~ Gamma
end_chr;

rmmg = 1e-2; //rate of missing marker genotypes
rmge = 5e-3; //rate of marker genotyping error
mmutr = 1e-9 /recurrent; //Marker mutation rate
qmutr = 1e-9; //QTL mutation rate
end_genome;

/*****
** Output options **
*****/
begin_output;
  linkage_map;
  allele_effect;
end_output;
```

Cartão de parâmetros para simulação no *software QMSim* baseado no genoma ovino com painel de alta densidade e 114 QTLs

```

/*****
** Global parameters **
*****/
title = "Ovinos 600k 114QTLs";
nrep = 10; //Number of replicates
h2 = 0.30; //Heritability
qtlh2 = 0.18; //QTL heritability
phvar = 5.61; //Phenotypic variance
nthread = 1;
/*****
** Historical population **
*****/
begin_hp;
hg_size = 1500 [0] 1500 [1000] 600 [1020]; //Size of the historical generations
//nmfhg = 750; //Number of males in the first generation
nmlhg = 300; //Number of males in the last generation
end_hp;
/*****
** Populations **
*****/
begin_pop = "EG";
begin_founder;
male [n = 300, pop = "hp"];
female [n = 300, pop = "hp"];
end_founder;
ng = 8;
ls = 3;
md = rnd_ug;
sr = 0.00 1.00;
dr = 0.00 1.00/e;
begin_popoutput;
stat;
end_popoutput;
end_pop;

begin_pop = "POP1";
begin_founder;
male [n = 400, pop = "EG", gen = 8];
female [n = 10000, pop = "EG", gen = 8];
end_founder;
ng = 8;
ls = 2;
md = rnd;
sr = 0.60 0.00;
dr = 0.20 0.00;
sd = phen;
cd = phen;
ebv_est = blup;

```



```

begin_popoutput;
stat;
ld /maft 0.05 /gen 6 7 8;
data;
genotype /snp_code /gen 6 7 8;
allele_freq;
end_popoutput;
end_pop;

/*****
** Genome **
*****/
begin_genome;
  begin_chr = 1; // CHR 1 //number of chromosomes with the same parameter
  chrLen = 301; //Chromosome length cM
  nmloci = 64360; //Number of markers
  mpos = even; //Marker positions
  nma = all 2; //Number of marker alleles
  maf = eql; //Marker allele frequencies
  nqloci = 3; //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 ~ Gamma
  end_chr;
//
  begin_chr = 1; // CHR 2 //number of chromosomes with the same parameter
  chrLen = 266; //Chromosome length cM
  nmloci = 57659; //Number of markers
  mpos = even; //Marker positions
  nma = all 2; //Number of marker alleles
  maf = eql; //Marker allele frequencies
  nqloci = 15; //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 ~ Gamma
  end_chr;
//
  begin_chr = 1; // CHR 3 //number of chromosomes with the same parameter
  chrLen = 241; //Chromosome length cM
  nmloci = 53250; //Number of markers
  mpos = even; //Marker positions
  nma = all 2; //Number of marker alleles
  maf = eql; //Marker allele frequencies
  nqloci = 3; //Number of QTL
  qpos = rnd;
  nqa = all 2; //Number of QTL alleles
  qaf = eql; //QTL allele frequencies
  qae = rndg 0.4; //QTL allele effects ~ Gamma

```

```

end_chr;
//
begin_chr = 1; // CHR 4 //number of chromosomes with the same parameter
chrln = 130; //Chromosome length cM
nmloci = 27607; //Number of markers
mpos = even; //Marker positions
nma = all 2; //Number of marker alleles
maf = eql; //Marker allele frequencies
nqloci = 3; //Number of QTL
qpos = rnd;
nqa = all 2; //Number of QTL alleles
qaf = eql; //QTL allele frequencies
qae = rndg 0.4; //QTL allele effects ~ Gamma
end_chr;
//
begin_chr = 1; // CHR 5 //number of chromosomes with the same parameter
chrln = 118; //Chromosome length cM
nmloci = 25429; //Number of markers
mpos = even; //Marker positions
nma = all 2; //Number of marker alleles
maf = eql; //Marker allele frequencies
nqloci = 3; //Number of QTL
qpos = rnd;
nqa = all 2; //Number of QTL alleles
qaf = eql; //QTL allele frequencies
qae = rndg 0.4; //QTL allele effects ~ Gamma
end_chr;
//
begin_chr = 1; // CHR 6 //number of chromosomes with the same parameter
chrln = 130; //Chromosome length cM
nmloci = 26661; //Number of markers
mpos = even; //Marker positions
nma = all 2; //Number of marker alleles
maf = eql; //Marker allele frequencies
nqloci = 6; //Number of QTL
qpos = rnd;
nqa = all 2; //Number of QTL alleles
qaf = eql; //QTL allele frequencies
qae = rndg 0.4; //QTL allele effects ~ Gamma
end_chr;
//
begin_chr = 1; // CHR 7 //number of chromosomes with the same parameter
chrln = 108; //Chromosome length cM
nmloci = 23561; //Number of markers
mpos = even; //Marker positions
nma = all 2; //Number of marker alleles
maf = eql; //Marker allele frequencies
nqloci = 9; //Number of QTL
qpos = rnd;
nqa = all 2; //Number of QTL alleles

```

```

qaf = eql;           //QTL allele frequencies
qae = rndg 0.4;     //QTL allele effects ~ Gamma
end_chr;
//
begin_chr = 1; // CHR 8 //number of chromosomes with the same parameter
chrln = 99;        //Chromosome length cM
nmloci = 20920;    //Number of markers
mpos = even;      //Marker positions
nma = all 2;       //Number of marker alleles
maf = eql;        //Marker allele frequencies
nqloci = 3;       //Number of QTL
qpos = rnd;
nqa = all 2;      //Number of QTL alleles
qaf = eql;       //QTL allele frequencies
qae = rndg 0.4;  //QTL allele effects ~ Gamma
end_chr;
//
begin_chr = 1; // CHR 9 //number of chromosomes with the same parameter
chrln = 105;      //Chromosome length cM
nmloci = 21568;   //Number of markers
mpos = even;     //Marker positions
nma = all 2;     //Number of marker alleles
maf = eql;      //Marker allele frequencies
nqloci = 15;    //Number of QTL
qpos = rnd;
nqa = all 2;    //Number of QTL alleles
qaf = eql;     //QTL allele frequencies
qae = rndg 0.4; //QTL allele effects ~ Gamma
end_chr;
//
begin_chr = 1; // CHR 10 //number of chromosomes with the same parameter
chrln = 97;    //Chromosome length cM
nmloci = 19678; //Number of markers
mpos = even;  //Marker positions
nma = all 2;  //Number of marker alleles
maf = eql;   //Marker allele frequencies
nqloci = 3;  //Number of QTL
qpos = rnd;
nqa = all 2; //Number of QTL alleles
qaf = eql;  //QTL allele frequencies
qae = rndg 0.4; //QTL allele effects ~ Gamma
end_chr;
//
begin_chr = 1; // CHR 11 //number of chromosomes with the same parameter
chrln = 61;    //Chromosome length cM
nmloci = 15913; //Number of markers
mpos = even;  //Marker positions
nma = all 2;  //Number of marker alleles
maf = eql;   //Marker allele frequencies
nqloci = 3;  //Number of QTL

```

```

qpos = rnd;
nqa = all 2;           //Number of QTL alleles
qaf = eql;           //QTL allele frequencies
qae = rndg 0.4;      //QTL allele effects ~ Gamma
end_chr;
//
begin_chr = 1; // CHR 12 //number of chromosomes with the same parameter
chrln = 84;          //Chromosome length cM
nmloci = 18937;      //Number of markers
mpos = even;        //Marker positions
nma = all 2;         //Number of marker alleles
maf = eql;          //Marker allele frequencies
nqloci = 3;         //Number of QTL
qpos = rnd;
nqa = all 2;         //Number of QTL alleles
qaf = eql;          //QTL allele frequencies
qae = rndg 0.4;     //QTL allele effects ~ Gamma
end_chr;
//
begin_chr = 1; // CHR 13 //number of chromosomes with the same parameter
chrln = 87;          //Chromosome length cM
nmloci = 19785;      //Number of markers
mpos = even;        //Marker positions
nma = all 2;         //Number of marker alleles
maf = eql;          //Marker allele frequencies
nqloci = 3;         //Number of QTL
qpos = rnd;
nqa = all 2;         //Number of QTL alleles
qaf = eql;          //QTL allele frequencies
qae = rndg 0.4;     //QTL allele effects ~ Gamma
end_chr;
//
begin_chr = 1; // CHR 14 //number of chromosomes with the same parameter
chrln = 71;          //Chromosome length cM
nmloci = 15991;      //Number of markers
mpos = even;        //Marker positions
nma = all 2;         //Number of marker alleles
maf = eql;          //Marker allele frequencies
nqloci = 3;         //Number of QTL
qpos = rnd;
nqa = all 2;         //Number of QTL alleles
qaf = eql;          //QTL allele frequencies
qae = rndg 0.4;     //QTL allele effects ~ Gamma
end_chr;
//
begin_chr = 1; // CHR 15 //number of chromosomes with the same parameter
chrln = 90;          //Chromosome length cM
nmloci = 19255;      //Number of markers
mpos = even;        //Marker positions
nma = all 2;         //Number of marker alleles

```

```

maf = eql; //Marker allele frequencies
nqloci = 3; //Number of QTL
qpos = rnd;
nqa = all 2; //Number of QTL alleles
qaf = eql; //QTL allele frequencies
qae = rndg 0.4; //QTL allele effects ~ Gamma
end_chr;
//
begin_chr = 1; // CHR 16 //number of chromosomes with the same parameter
chrln = 78; //Chromosome length cM
nmloci = 16499; //Number of markers
mpos = even; //Marker positions
nma = all 2; //Number of marker alleles
maf = eql; //Marker allele frequencies
nqloci = 3; //Number of QTL
qpos = rnd;
nqa = all 2; //Number of QTL alleles
qaf = eql; //QTL allele frequencies
qae = rndg 0.4; //QTL allele effects ~ Gamma
end_chr;
//
begin_chr = 1; // CHR 17 //number of chromosomes with the same parameter
chrln = 82; //Chromosome length cM
nmloci = 16671; //Number of markers
mpos = even; //Marker positions
nma = all 2; //Number of marker alleles
maf = eql; //Marker allele frequencies
nqloci = 3; //Number of QTL
qpos = rnd;
nqa = all 2; //Number of QTL alleles
qaf = eql; //QTL allele frequencies
qae = rndg 0.4; //QTL allele effects ~ Gamma
end_chr;
//
begin_chr = 1; // CHR 18 //number of chromosomes with the same parameter
chrln = 71; //Chromosome length cM
nmloci = 16129; //Number of markers
mpos = even; //Marker positions
nma = all 2; //Number of marker alleles
maf = eql; //Marker allele frequencies
nqloci = 6; //Number of QTL
qpos = rnd;
nqa = all 2; //Number of QTL alleles
qaf = eql; //QTL allele frequencies
qae = rndg 0.4; //QTL allele effects ~ Gamma
end_chr;
//
begin_chr = 1; // CHR 19 //number of chromosomes with the same parameter
chrln = 63; //Chromosome length cM
nmloci = 14540; //Number of markers

```

```

mpos = even;           //Marker positions
nma = all 2;           //Number of marker alleles
maf = eql;             //Marker allele frequencies
nqloci = 3;           //Number of QTL
qpos = rnd;
nqa = all 2;           //Number of QTL alleles
qaf = eql;             //QTL allele frequencies
qae = rndg 0.4;       //QTL allele effects ~ Gamma
end_chr;
//
begin_chr = 1; // CHR 20 //number of chromosomes with the same parameter
chrln = 56;         //Chromosome length cM
nmloci = 12583;     //Number of markers
mpos = even;       //Marker positions
nma = all 2;       //Number of marker alleles
maf = eql;         //Marker allele frequencies
nqloci = 3;        //Number of QTL
qpos = rnd;
nqa = all 2;       //Number of QTL alleles
qaf = eql;         //QTL allele frequencies
qae = rndg 0.4;   //QTL allele effects ~ Gamma
end_chr;
//
begin_chr = 1; // CHR 21 //number of chromosomes with the same parameter
chrln = 53;     //Chromosome length cM
nmloci = 11896; //Number of markers
mpos = even;   //Marker positions
nma = all 2;   //Number of marker alleles
maf = eql;     //Marker allele frequencies
nqloci = 3;    //Number of QTL
qpos = rnd;
nqa = all 2;   //Number of QTL alleles
qaf = eql;     //QTL allele frequencies
qae = rndg 0.4; //QTL allele effects ~ Gamma
end_chr;
//
begin_chr = 1; // CHR 22 //number of chromosomes with the same parameter
chrln = 56;     //Chromosome length cM
nmloci = 12097; //Number of markers
mpos = even;   //Marker positions
nma = all 2;   //Number of marker alleles
maf = eql;     //Marker allele frequencies
nqloci = 3;    //Number of QTL
qpos = rnd;
nqa = all 2;   //Number of QTL alleles
qaf = eql;     //QTL allele frequencies
qae = rndg 0.4; //QTL allele effects ~ Gamma
end_chr;
//

```

```

begin_chr = 1; // CHR 23 //number of chromosomes with the same parameter
chrln = 68; //Chromosome length cM
nmloci = 14178; //Number of markers
mpos = even; //Marker positions
nma = all 2; //Number of marker alleles
maf = eql; //Marker allele frequencies
nqloci = 3; //Number of QTL
qpos = rnd;
nqa = all 2; //Number of QTL alleles
qaf = eql; //QTL allele frequencies
qae = rndg 0.4; //QTL allele effects ~ Gamma
end_chr;
//
begin_chr = 1; // CHR 24 //number of chromosomes with the same parameter
chrln = 43; //Chromosome length cM
nmloci = 10458; //Number of markers
mpos = even; //Marker positions
nma = all 2; //Number of marker alleles
maf = eql; //Marker allele frequencies
nqloci = 3; //Number of QTL
qpos = rnd;
nqa = all 2; //Number of QTL alleles
qaf = eql; //QTL allele frequencies
qae = rndg 0.4; //QTL allele effects ~ Gamma
end_chr;
//
begin_chr = 1; // CHR 25 //number of chromosomes with the same parameter
chrln = 48; //Chromosome length cM
nmloci = 10824; //Number of markers
mpos = even; //Marker positions
nma = all 2; //Number of marker alleles
maf = eql; //Marker allele frequencies
nqloci = 3; //Number of QTL
qpos = rnd;
nqa = all 2; //Number of QTL alleles
qaf = eql; //QTL allele frequencies
qae = rndg 0.4; //QTL allele effects ~ Gamma
end_chr;
//
begin_chr = 1; // CHR 26 //number of chromosomes with the same parameter
chrln = 49; //Chromosome length cM
nmloci = 10146; //Number of markers
mpos = even; //Marker positions
nma = all 2; //Number of marker alleles
maf = eql; //
nqloci = 3; //Number of QTL
qpos = rnd;
nqa = all 2; //Number of QTL alleles
qaf = eql; //QTL allele frequencies
qae = rndg 0.4; //QTL allele effects ~ Gamma

```

```
end_chr;

rmmg = 1e-2; //rate of missing marker genotypes
rmge = 5e-3; //rate of marker genotyping error
mmutr = 1e-9 /recurrent; //Marker mutation rate
qmutr = 1e-9; //QTL mutation rate
end_genome;

/*****
** Output options **
*****/
begin_output;
  linkage_map;
  allele_effect;
end_output;
```