

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

Modelos genômicos não aditivos na avaliação de vacas leiteiras

Fabício Pilonetto

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

Piracicaba
2021

Fabrcio Pilonetto
Zootecnista

Modelos genômicos não aditivos na avaliação de vacas leiteiras
versão revisada de acordo com a resolução CoPGr 6018 de 2011

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

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

Piracicaba
2021

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

Pilonetto, Fabrício

Modelos genômicos não aditivos na avaliação de vacas leiteiras /
Fabrício Pilonetto. - - versão revisada de acordo com a resolução CoPGr
6018 de 2011. - - Piracicaba, 2021.

93 p.

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

1. Genômica. 2. GEBV. 3. Heterose. 4. Heterozigosidade. 5. Simulação
I. Título.

DEDICATÓRIA

Dedico essa tese aos meus maiores apoiadores: mãe, pai e irmã, vocês tem todo o meu amor.

AGRADECIMENTOS

Primeiramente, agradeço as pessoas que sempre estiveram ao meu lado, em todas as minhas conquistas, quedas e que, mesmo distantes geograficamente, lutaram ao meu lado e, por vezes, deixaram os seus sonhos para viverem os meus. Mãe e Pai, vocês são o que tenho de mais valioso nesta vida. Angelita e Valdir Pilonetto, não há palavras suficientes para agradecê-los.

A minha irmã, Ingrid Pilonetto, que fez de mim uma pessoa mais responsável, atenciosa e dedicada. Mana, você é a parte mais linda de mim, eu te amo imensamente e sou grato por estarmos tão unidos.

Aos meus incríveis amigos e amigas, Bruna Petry (Bru), Mayara Salvian (May), Izally C. Gervásio (Iza), Giovanni C. Ladeira (Djova), Sofia A. e Silva (Sofi) e Felipe E. Braguim (Biscate). Obrigado pelas conversas, apoio, sorrisos e por todos os momentos compartilhados nestes últimos anos. Que nunca nos falte amor, persistência, trabalho, sucesso e é claro, churrasco e cerveja (rsrs).

Ao meu orientador Gerson Barreto Mourão, por toda lição, conhecimento, conselhos, conversas e pela paciência. Obrigado por me acolher em seu grupo de pesquisa (GEMA) e em sua família.

Aos colegas de pós-graduação, estagiários e ex-integrantes do grupo GEMA. Obrigado pelo tempo e conhecimento compartilhado: Luiz, Letícia Sartori (Le), Brayan, Paola, Eula Carrara, Juliana Petrini (Ju), Gregori Rovadoscki (Greg), Fátima A. Bogdanski, Isabella (Isa), Flaviane e Adhonis (Doni).

À Dra. Aline Zampar, minha nobre amiga e uma das minhas principais influências como profissional e ser humano. Tenho muito orgulho da nossa amizade. Obrigado por fazer parte da minha vida e deste momento.

A comunidade LGBTQI+ que lutou e ainda luta pelo direito de expressar sua verdadeira essência, pela tolerância e respeito! Obrigado por representar esta causa de forma tão corajosa. Jamais permita que invalidam sua competência, caráter, trabalho e a sua vida!

Por fim e não menos importante, desejo expressar minha gratidão à Universidade de São Paulo, USP – ESALQ, pela oportunidade e privilégio de estudar nesta renomada Instituição. Além disso, agradeço à FAPESP (Processo nº 2019/03373-9), CAPES e CNPq pelo apoio financeiro.

*"Livin' in a world where no one's innocent,
Oh, but at least we try,
Gotta live my truth, not keep it bottled in,
So, I don't lose my mind"*

"Rain on me", Lady Gaga & Ariana Grande.

SUMÁRIO

RESUMO.....	8
ABSTRACT.....	9
1. INTRODUÇÃO.....	11
REFERÊNCIAS.....	13
2. REVISÃO BIBLIOGRÁFICA.....	15
2.1 Avaliação genômica na produção animal.....	15
2.2 Efeito genético aditivo e não-aditivo.....	16
2.3 Heterozigose.....	17
REFERÊNCIAS.....	19
3. COMPONENTES DE VARIÂNCIA, ACURÁCIA E VIÉS DE PREDIÇÃO ESTIMADOS POR MODELO GENÔMICO ADITIVO E ADITIVO-DOMINANTE EM POPULAÇÕES SIMULADAS E REAIS DE VACAS DA RAÇA HOLANDESA.....	23
RESUMO.....	23
ABSTRACT.....	23
3.1. INTRODUÇÃO.....	24
3.2. MATERIAL E MÉTODOS.....	25
3.2.1. Simulação da estrutura populacional – Etapa I.....	25
3.2.2. Simulação dos efeitos genéticos aditivos e de desvios de dominância – Etapa II.....	27
3.2.3. Dados simulados.....	33
3.2.4. Dados reais.....	35
3.2.5. Análises genéticas.....	37
3.2.6. Avaliação da acurácia e viés de predição.....	38
3.3. RESULTADOS.....	40
3.3.1. Componentes de variância.....	40
3.3.2. Acurácia e viés de predição.....	42
3.4. DISCUSSÃO.....	45
3.4.1. Componentes de variância.....	45
3.4.2. Acurácia e viés de predição.....	48
3.5. CONCLUSÃO.....	49
REFERÊNCIAS.....	50

4. ESTIMATIVA DA HETEROZIGOSIDADE MOLECULAR COMO RECURSO PARA GERAR HETEROSE GENÔMICA EM POPULAÇÕES DE RAÇA PURA	55
RESUMO	57
ABSTRACT	58
4.1. INTRODUÇÃO.....	56
4.2. MATERIAL E MÉTODOS.....	57
4.2.1. Simulação da estrutura populacional e do genoma	57
4.2.2. Simulação dos desvios de dominância	59
4.2.3. Cálculo da heterozigose molecular e seus fatores	60
4.2.4. Análise de associação ampla do genoma (GWAS) adaptado para heterozigosidade	61
4.2.5. Delineamento da análise de GWAS adaptado.....	64
4.3. RESULTADOS	66
4.3.1. GWAS adaptado	66
4.3.2. Genes adjacentes a região do SNP significativo	76
4.3.3. Heterozigose genômica e seus fatores – Simulação	78
4.3.4. Exploração da heterozigosidade no genoma real	83
4.4. DISCUSSÃO	85
4.5. CONCLUSÃO.....	87
REFERÊNCIAS	88
5. CONSIDERAÇÕES FINAIS	93

RESUMO

Modelos genômicos não aditivos na avaliação de vacas leiteiras

Na pecuária leiteira, o interesse em selecionar animais com alto potencial genético para produção e qualidade do leite, pode causar um decréscimo na variabilidade genotípica, devido à intensa pressão de seleção. Neste contexto, as avaliações genéticas têm considerado apenas os efeitos genéticos aditivos, não sendo incluídos efeitos de dominância nos modelos de predição, por exemplo. O objetivo deste trabalho foi verificar a contribuição dos efeitos genéticos não-aditivos nos modelos de predição genômica e inferir a heterozigose molecular e a heterose genômica. Foram testadas a habilidade e o viés na predição dos valores genéticos genômicos (GEBVs) por um modelo puramente aditivo e outro aditivo-dominante. Estes foram comparados e as estimativas de variância computadas e avaliadas de acordo com o modelo proposto. Uma segunda fase compreendeu estimar a heterozigosidade e a heterose genômica, e realizar uma adaptação da associação ampla do genoma (GWAS) para capturar as variâncias não-aditivas, identificar regiões do genoma com possível vantagem heterozigótica e de marcadores moleculares associados ao fenótipo. Em ambas as fases foram utilizados dados simulados e registros fenotípicos reais de bovinos leiteiros da raça Holandesa. As proporções da variância de dominância em relação ao fenótipo foram de 2,9 a 5,2% em dados simulados, e de 0,08 a 3,8% para as características produção e qualidade do leite. A maior acurácia foi de 0,79 e 0,36 no modelo aditivo e aditivo-dominante simulado, respectivamente. Para os dados reais, as estimativas foram semelhantes entre os modelos. O melhor desempenho do modelo foi para gordura e ácido graxo, com maior vantagem para ácido graxo quando incluído os efeitos de dominância na avaliação. Em média, para o cenário de menor herdabilidade ($h^2 = 0,10$), a taxa de heterozigosidade foi de 35,6%, e níveis próximos foram encontrados para os demais cenários (35,4 e 35,7%). O cenário de maior herdabilidade apresentou os maiores valores médios de heterose genômica (17,0%), comparado aos demais (3,02 a 11,0%). Os coeficientes de regressão revelaram que a heterozigosidade tem impacto positivo sobre a heterose genômica para a maioria das populações simuladas. A heterozigosidade avaliada de forma empírica e por meio de dados reais revelou resultados otimistas quanto à possibilidade de selecionar os animais de raça pura, para gerar maior e melhor heterose genômica. Incluir os efeitos de dominância no modelo de GWAS adaptado pode contribuir para a identificação de marcadores associados ao fenótipo, de forma semelhante ou superior ao modelo aditivo adaptado. Conclui-se que incluir os efeitos de dominância deve beneficiar uma melhor estimativas dos componentes de variância genética, principalmente para as características de menor herdabilidade e altamente influenciadas pelo ambiente. Além disso, sugere-se que a heterose genômica em populações de raça pura pode ser acessada se considerado os fatores de heterozigosidade molecular e suas associações com o fenótipo de interesse.

Palavras-chave: Genômica, GEBV, Heterose, Heterozigosidade, Simulação

ABSTRACT

Non-additive genomic models for the evaluation of dairy cows

In dairy farming, the interest in selecting animals with high genetic potential for milk production and quality may cause a decrease in phenotypic and genotypic variability, due to intense selective pressure. Furthermore, genetic evaluations have been limited to additive genetic effects, not including dominance effects in genomic prediction models, for example. The aim of this study was to verify the contribution of non-additive genetic effects in genomic prediction models and to infer molecular heterozygosity and genomic heterosis. Two main lines of research were followed. Initially, the ability and bias in predicting genomic breeding values by a purely additive model and an additive-dominant model were tested. The models were compared for ability and bias to genomic estimated breeding values (GEBVs) and the estimates of variance computed and evaluated according to the proposed model. The second phase involved inferring genomic heterozygosity and heterosis, and performing an adaptation of genome-wide association (GWAS) to capture non-additive variances, identify regions of the genome with possible heterozygous advantage, and molecular markers associated with the phenotype. In both phases, simulated data and real phenotypic records of Holstein dairy cattle were used. The proportions of dominance variance in relation to phenotype were 2.9 to 5.2% in simulated data, and from 0.08 to 3.8% for the milk yield and quality traits. The highest accuracy was 0.79 and 0.36 in the simulated additive and additive-dominant model. The estimates were similar across models for the real data. Higher variance estimates were found for fat and fatty acid, with a greater advantage for fatty acid when dominance effects were included in the assessment. On average, for the lower heritability scenario, heterozygosity was 35.6%, and similar levels were found for the other scenarios (35.4 and 35.7%). The high heritability scenario presented the highest mean values of genomic heterosis (17.0%), compared to the others (3.02 to 11.0%). Regression coefficients revealed that heterozygosity has a positive impact on genomic heterosis for most simulated populations. The heterozygosity evaluated empirically and through real data revealed optimistic results regarding the possibility of selecting purebred animals to genomic heterosis. Including dominance effects in the adapted GWAS model can contribute to the identification of markers associated with the phenotype in a similar or superior way to the adapted additive model. We conclude that accounting for dominance effects in genomic prediction models should benefit better estimates of genetic variance components, especially for traits with lower heritability and highly influenced by the environment. Furthermore, it is suggested that genomic heterosis in purebred populations can be accessed if molecular heterozygosity factors and their associations with the phenotype of interest are considered.

Keywords: Genomic, GEBV, Heterosis, Heterozygosity, Simulation

1. INTRODUÇÃO

O desenvolvimento tecnológico na pecuária é inevitável. Com o advento da genômica, os estudos teóricos voltados a arquitetura genética de características de interesse econômico, desenvolveram-se significativamente em prol do aumento da eficiência produtiva dos animais, por meio da identificação de regiões genômicas e de genes associados a expressão do fenótipo (Hayes, 2013; Steri et al., 2018; Bonnafous et al., 2018).

Em populações altamente especializadas, como em bovinos leiteiros, a intensa seleção para poucas características condiciona o decréscimo no desempenho produtivo, devido à redução da variabilidade genética e formação de longos trechos genômicos em homozigose (Hayes, 2013; Steri et al., 2018; Rebelato et al., 2018). Contudo, atualmente é possível explorar fontes de variação genética e efeitos de complexa avaliação estatística, que podem ser utilizadas nas avaliações genéticas para melhorar acurácia de predição dos valores genéticos (Tsairidou et al., 2018). Visto isso, os efeitos genéticos não-aditivos (e.g. desvios de dominância) são importantes para os modelos de avaliação genética e podem ser obtidos via informação genômica com maior precisão, comparado às abordagens tradicionais que consideram apenas o pedigree como referência.

Adicionalmente, a teoria de heterose em animais cruzados pode ser explorada via heterozigose genômica, na qual as informações provenientes do genoma serão úteis para a estimativa dos efeitos de interação gênica, como os desvios de dominância (Alillo et al., 2017; Akanno et al., 2017; Tsairidou et al., 2018). Desta forma, a seleção dos animais baseado na informação de heterozigose pode ser uma abordagem promissora para promover progresso genético e variabilidade fenotípica em rebanhos de raça pura. Além disso, tal informação não é encontrada nas avaliações genéticas de touros comerciais, a qual poderá contribuir para aumentar os ganhos econômicos.

A utilização de softwares e programas de simulação tem sido amplamente utilizados nos estudos populacionais e de melhoramento genético, em diversas espécies animais (Xiang et al. 2018; Mao et al., 2020; Abdollahi-Arpanahi et al., 2020; Perez et al., 2019; Seno et al., 2018). Optar por utilizar métodos de simulação estocástica, auxiliam na compreensão do mecanismo de ação de características complexas, pouco herdáveis, influenciadas por diversos genes e, principalmente, por fatores não genéticos, associados ao ambiente de criação dos animais. Além disso, essa abordagem contribui para o estudo de diferentes modelos e métodos estatísticos na avaliação genética (Bradford et al., 2018; Wellmann & Bennewitz, 2012). Dessa forma, o direcionamento analítico em uma população real qualquer pode mostrar resultados mais

satisfatórios quanto a escolha do método de avaliação mais adequado para determinado cenário de criação animal.

Devido a utilização de marcadores moleculares como base de informação do genoma, tornou-se possível explorar diferentes fontes de variação genética e incorporá-las nos modelos tradicionais de avaliação. Recentemente, Abdollahi-Arpanahi et al. (2020) publicaram um trabalho que testou o método de aprendizagem profunda (deep learning), utilizando algoritmos e redes neurais (*Convolutional Neural Network* – CNN) e métodos paramétricos de avaliação genética (Genomic Best Linear Unbiased Prediction - GBLUP e Bayesiano), comumente utilizados na predição dos valores genéticos genômicos (Genomic Estimated Breeding Value - GEBV). Neste estudo, os autores utilizaram dados reais e simulados, e curiosamente descobriram que ao incluir efeitos não-genéticos, como dominância e epistasia, os métodos tradicionais (GBLUP, por exemplo) não foram tão eficientes nas predições, comparado aos métodos de aprendizagem, para características complexas como a fertilidade ($h^2 = 0,30$). Este achado não exime ou reduz a eficiência no uso de abordagens e procedimentos tradicionais de avaliação, mas sim indica a necessidade em ajustar os modelos para que comportem diferentes fontes de variação genética, que comprovadamente, influenciam nas predições genéticas e na variabilidade fenotípica de determinada população (Akanno et al., 2017; Tsairidou et al., 2018).

Mesmo com a oportunidade e interesse dos pesquisadores em estudar a complexidade de características importantes para a produção animal, ainda é um desafio transformar as informações genômicas em conhecimento aplicado na prática (Abdollahi-Arpanahi et al., 2020). As metodologias consolidadas para a avaliação genética obtêm predições com base no relacionamento aditivo entre os indivíduos e, conseqüentemente, baseiam-se nesses resultados para selecionar os animais para reprodução (Tsairidou et al., 2018). No entanto, torna-se incerto qual a verdadeira contribuição dos efeitos genéticos não-aditivos para a predição do valor genético dos animais.

Os desvios de dominância têm sido amplamente utilizados nos modelos de predição genética. Tais efeitos por muito tempo foram desconsiderados nos estudos genéticos, devido à dificuldade em estimá-los. Embora os efeitos não-aditivos, em contraste com os efeitos aditivos, possam contribuir pouco para a variância genética (Jiang et al., 2017), estudos sugerem que a dominância, por exemplo, pode contribuir de 3,8% a 7,1% na variância genética para produção de leite e 1,2% para características reprodutivas em bovinos leiteiros (Alillo et al., 2016), de 5 a 11% para número de leitões nascidos (Xiang et al., 2016) e de 7,3% a 19,2% para peso e composição corporal em ovinos de raça pura e cruzados, respectivamente (Moghaddar & van der Werf, 2017).

Dessa forma, interessado em explorar os benefícios da inclusão de fontes genéticas não convencionais na avaliação dos animais, o objetivo deste estudo foi comparar a habilidade de predição do modelo genético aditivo tradicional versus modelo genético não-aditivo. Além disso, com o intuito de identificar os indivíduos com potencial para gerar maior e melhor heterose, também foi alvo de estudo estimar a heterozigose molecular individual de uma população de animais de raça pura e identificar as regiões com vantagem heterozigótica para características de produção e qualidade do leite.

REFERÊNCIAS

- Abdollahi-Arpanahi, R.; Gianola, D.; Peñagaricano, F. Deep learning versus parametric and ensemble methods for genomic prediction of complex phenotypes. **Genetics Selection Evolution**, v.52, n.12, 2020.
- Akanno, E. C.; Chen, I.; Abo-Ismael, M. K.; Crowley, J. J.; Wang, Z.; Li, C.; Basarab, J. A.; MacNeil, M. D.; Plastow, G. Genomic prediction of breed composition and heterosis effects in Angus, Charolais, and Hereford crosses using 50K genotypes. **Canadian Journal Animal Science**, v.97, p.431-438, 2017.
- Alillo, H.; Pryce, J. E.; González-Recio, O.; Cocks, B. G.; Hayes, B. J. Accounting for dominance to improve genomic evaluations of dairy cows for fertility and milk production traits. **Genetics Selection Evolution**, v.48, n.8, p.1-11, 2016.
- Bonnafous, F.; Fievet, G.; Blanchet, N.; Carrère, S.; Gouz, J.; Legrand, L.; Marage, G.; Bret-Mestries, E.; Munos, S.; Pouilly, N.; Vincourt, P.; Langlade, N.; Mangin, B. Comparison of GWAS models to identify non-additive genetic control of flowering time in sunflower hybrids. **Theoretical and Applied Genetics**, v.131, p.319-332, 2018.
- Bradford, H. L.; Masuda, Y.; VanRaden, P. M.; Legarra, A.; Misztal, I. Modeling missing pedigree in single-step genomic BLUP. **Journal of Dairy Science**, v.102, p.2336-2346, 2019.
- Hayes, B. J. Overview of statistical methods for genome-wide association studies (GWAS). **Methods in Molecular Biology**, v.1019, p.146-169, 2013.
- Jiang, J.; Shen, B.; O'Connell, R.; VanRaden, P. M.; Cole, J. B.; Ma, L. Dissection of additive, dominance, and imprinting effects for production and reproduction traits in Holstein cattle. **BMC Genetics**, v.18, n.425, 2017.
- Mao, X.; Sahana, G.; Johansson, M. A.; Liu, A.; Ismael, A.; Lovendahl, P.; De Koning, D.-J.; Guldbbrandtsen, B. Genome-wide association mapping for dominance effects in female fertility using real and simulated data from Danish Holstein cattle. **Nature**, v.10, n.2953, 2020.
- Perez, B. C.; Balieiro, J. C. C.; Carneiro, R.; Tirelo, F.; Oliveira Junior, G. A.; Dementshuk, J. M.; Eler, J. P.; Ferraz, J. B. S.; Ventura, R. V. Accounting for population structure in selective cow genotyping strategies. **Journal of Animal Breeding and Genetics**, v.136, p.23-39, 2018.

- Rebelato, A. B.; Caetano, A. R. Runs of homozygosity for autozygosity estimation and genomic analysis in production animals. **Pesquisa Agropecuária Brasileira**, v.53, n.9, p.975-984, 2018.
- Seno, L. de O.; Guidolin, D. G. F.; Aspilcueta-Borquis, R.; Nascimento, G. B.; Silva, T. B. R. da; Oliveira, H. N. de; Munari, D. P. Genomic selection in dairy cattle simulated populations. **Journal of Dairy Research**, v.85, 0.125-132, 2018.
- Steri, R.; Moioli, B.; Catillo, G.; Galli, A.; Buttazzoni, L. Genome-wide association study for longevity in the Holstein cattle population. *Animal*, v.3, p.1-8, 2018.
- Tsairidou, S.; Allen, A. R.; Pong-Wong, R.; McBride, S. H.; Wright, D. M.; Matika, O.; Pooley, C. M.; McDowell, S. W. J.; Glass E. J.; Skuce, R. A.; Bishop, S. C.; Woolliams, J. A. An analysis of effects of heterozygosity in dairy cattle for bovine tuberculosis resistance. **Animal Genetics**, v.49, p.103-109, 2018.
- Wellmann, R.; Bennewitz, J. Bayesian models with dominance effects for genomic evaluation of quantitative traits. **Genetics Research Cambridge**, v.94, p.21-37, 2012.
- Xiang, T.; Christensen, O. F.; Vitezica, Z. G.; Legarra, A. Genomic evaluation by including dominance effects and inbreeding depression for purebred and crossbred performance with an application in pigs. **Genetics Selection Evolution**, v.48, n.92, 2016.
- Xiang, T.; Christensen, F. O.; Vitezica, G. Z.; Legarra, A. Genomic model with correlation between additive and dominance effects. **Genetics**, v.209, p.711-723, 2018.

2. REVISÃO BIBLIOGRÁFICA

2.1 Avaliação genômica na produção animal

Dentre as ferramentas aplicadas na avaliação genômica dos animais, a seleção genômica (GS, Genomic Selection) e estudo de associação genômica ampla (GWAS, Genome-Wide Association Study) são as principais aplicações na predição de valores genéticos genômicos (GEBV), e para a identificação e associação de marcadores moleculares com o fenótipo de interesse. Ambas são complementares, tendo em vista que é necessário haver um banco de dados com informações genômicas, captadas via painel de marcadores moleculares, principalmente do tipo SNP (Single Nucleotide Polymorphism), na qual os efeitos são estimados e utilizados na predição dos GEBVs para a seleção dos melhores indivíduos (Meuwissen et al., 2001; VanRaden et al., 2008; Resende et al., 2008).

Adicionalmente, GS e GWAS proporcionaram aos pesquisadores investigar e quantificar as variâncias genéticas e fenotípicas dos caracteres quantitativos (Petrini et al., 2016). Além disso, especificamente por meio dos GWAS, diversos marcadores genéticos, QTLs e genes foram identificados em populações comerciais, como em um estudo recente que identificou 117 SNPs e 10 novos QTLs associados à resistência à mastite em vacas da raça Holandesa (Kurz et al., 2019). Também em Iung et al., (2018), um estudo em GWAS com dados de uma população de bovinos da raça Holandesa no Brasil, encontraram 46 regiões genômicas em 10 autossomos *Bos taurus*, associados à produção de leite e ácidos graxos. Neste estudo, ainda foram identificados genes candidatos que estão envolvidos no reparo da glândula mamária (COL18A1), resposta imune (LTFC19), homeostase da glicose (SLC37A1), síntese de ácidos graxos insaturados (LTBP1) e transporte de açúcar (SLC37A1 e MFSD4A).

Outro estudo com vacas leiteiras, Steri et al. (2018), identificaram regiões do genoma e genes candidatos que podem ser utilizadas para aumentar a longevidade. De acordo com estes autores, as características de difícil mensuração e de baixa herdabilidade, como a fertilidade, historicamente não recebiam foco nos programas de melhoramento genético pela dificuldade em analisá-las, no entanto, com a genômica, tornou-se possível estudar a complexidade destas características a nível molecular, ao ponto de, atualmente, serem utilizadas como estratégia de seleção dos animais. As investigações desta natureza são importantes para potencializar a seleção genômica dos animais, pois a identificação de regiões do genoma de interesse pode otimizar a construção de matrizes de SNP específicas para as características analisadas (Iung et al., 2018).

Nesse contexto, especialmente para bovinos leiteiros, os estudos com GWAS apresentam alguns pontos especiais que diferem de outras espécies. Um deles é a redução da

independência entre os estudos de GWAS, devido ao uso intensivo de touros comercialmente importantes, que por sua vez aumenta o relacionamento entre as populações leiteiras. Além disso, a utilização intensiva de um pequeno número de touros, contribuiu para o alto nível de desequilíbrio de ligação (LD) e reduzido tamanho efetivo da população, levando ao maior poder estatístico nas análises de associação dos SNPs com o fenótipo (Hayes et al., 2003). Estas especificidades possibilitaram um grande avanço na pecuária leiteira, com uma rápida evolução das metodologias genômicas e rendendo inúmeros trabalhos realizados com vistas na melhoria da produção de leite (Owens et al., 2014; Jiang et al., 2017; Nicolini et al., 2018; Wang et al., 2019).

2.2 Efeito genético aditivo e não-aditivo

Os efeitos não-aditivos têm uma importante relevância nos sistemas de produção animal, principalmente para os programas de cruzamento, nos quais os efeitos da heterose são, em grande parte, responsáveis pelo desempenho superior dos animais cruzados comparado a animais de raça pura (Wei & Van Der Werf, 1994; Xiang et al., 2016). Espera-se que a inclusão desses efeitos nos modelos de avaliação genética contribua para o aumento da resposta à seleção, principalmente para as características complexas, que apresentam baixa magnitude de herdabilidade (Varona & Misztal, 1999; Toro & Varona, 2010).

Com a genômica, tem sido possível estimar os efeitos de desvio de dominância dos marcadores que, associados às informações dos efeitos aditivos, permitem predizer o mérito genético dos animais, melhorar a confiabilidade na predição dos valores genômicos e direcionamento dos acasalamentos (Toro & Verona, 2010; Amuzu-Aweh et al., 2013; Alillo et al., 2017; Akanno et al., 2017). Adicionalmente, dentre os efeitos não-aditivos, os desvios de dominância são os mais explorados nos estudos genômicos. Estes podem contribuir de 5 a 7% para a variância genética total, compondo um fator explorável importante em estudos genômicos (Sun et al., 2014). Contudo, o principal argumento que justifica a exclusão dos efeitos não-aditivos nos modelos de avaliação é o fato destes não serem diretamente herdáveis (diferentemente do efeito genético aditivo), além da complexidade computacional para a estimação e baixa acurácia das estimativas dos componentes de variância associados a estes efeitos (Toro & Varona, 2010; Aliloo et al., 2017).

Entretanto, estes efeitos aliados ao GWAS podem potencializar o estudo das variantes genéticas que sofrem efeito não-aditivo (Nicolini et al., 2018). Além disso, os modelos de GWAS são construídos a partir de modelos genéticos aditivo-dominante, que auxiliam na identificação de regiões genômicas que contenham genes reguladores de características complexas (Bonnafous et al., 2018). Em Jiang et al. (2017) foi identificado um sinal de dominância no cromossomo 23 de

bovinos raça Holandesa que estava associado à produção de leite. Neste mesmo estudo, próximo a este sinal foram identificados dois SNPs na forma de dominância associados à produção de leite, sendo eles Hapmap48809-BTA-55698 e BovineHD2300004730. Este último é *upstream* do gene RUNX2, que foi descrito como responsável pela diferenciação das células epiteliais no desenvolvimento da glândula mamária, que em humanos pode estar associado ao desenvolvimento do câncer de mama (Owens et al., 2014). Para Jiang et al., (2017), os efeitos de dominância contribuem significativamente para o estudo e predições genômicas para as características de produção.

Recentemente, foram identificadas quatro regiões genômicas marcadas por efeito de dominância, que abrigava seis genes diretamente ligados ao desenvolvimento testicular, manutenção da linhagem germinativa masculina e maturação espermática em touros da raça Holandesa (Nicolini et al., 2018). É importante observar que os estudos genômicos concentram grande parte da importância nos efeitos genéticos aditivos, contudo, os efeitos não-aditivos podem ser incluídos nos modelos de GWAS, com importante relevância nas predições genéticas para as diversas características economicamente importantes.

2.3 Heterozigose

A heterose ou vigor híbrido é caracterizada pela superioridade na performance fenotípica da progênie em relação ao desempenho médio de seus pais, sendo um evento genético importante para criações comerciais de frangos, suínos e bovinos de corte, por exemplo (Shen et al., 2014; Amuzu-Aweh et al., 2013). A ocorrência de heterose é geralmente devido ao efeito genético de dominância (Xiang et al., 2016), calculada pela diferença proporcional quadrática da frequência alélica, isto é, é a diferença entre a média do desempenho do F1 e da população parental, medida pelas frequências alélicas de p e q ($H_{F_1} = M_{F_1} - M_{\bar{P}}$) (Falconer e Mackay, 1996).

O efeito de dominância ocorre dependente da sua direção, isto é, ao considerar um locus bialélico, se o genótipo heterozigoto for superior ou inferior aos homozigotos, pressupõe-se que há sobredominância, por outro lado, se o heterozigoto for intermediário, diz-se que a uma dominância parcial (Tsairidou et al., 2018). Dado que a heterose é dependente destes efeitos, é importante que estas variáveis sejam mantidas na população para manter diversidade genética (Nietlisbach et al., 2016). Além disso, a fixação de QTLs com efeito aditivo favorável pode depender de tais efeitos, principalmente da sobredominância (Tsairidou et al., 2018). Soresen et al. (2008), em um amplo estudo de revisão com cruzamentos entre raças de bovinos leiteiros,

admitem que a velocidade de aprimoramento das características desejáveis deverá ser afetada em decorrência da seleção dos animais de raças puras.

Isto é consistente quando realizado, por exemplo, o cruzamento de bovinos da raça Jersey, com alto potencial genético para sólidos no leite, e de bovinos da raça Holandesa, conhecida por ser uma das raças leiteiras com maior volume de produção de leite. Soresen et al. (2008) relataram aumento de produtividade leiteira de até 10% em animais cruzados, com grau de heterose entre 1,5 a 8,4% para produção de leite, gordura e proteína. Kargo et al. (2021) também relataram acréscimos consideráveis na produtividade de animais leiteiros cruzados, com aumento de 4,0 a 4,4% nos níveis produtivos de leite e sólidos.

Estas são informações valiosas quanto ao progresso genético em populações cruzadas, contudo, Kargo et al. (2021) cita que o padrão racial nos cruzamentos nem sempre será determinante em relação a heterose, principalmente para características de baixa herdabilidade e com maior influência ambiental. Isto concorda com Clasen et al. (2017), que relataram padrão racial, definido como os efeitos aditivos estimados, não significativo para a taxa de heterose em características de herdabilidade entre 0,02 e 0,09, como a longevidade. No entanto, isso não concorda com o estudo de Saborío-Montero et al. (2018), na qual a estimativa de regressão da incidência de febre do leite ou hipocalcemia (herdabilidades entre 0,02 e 0,07) em função da raça foi significativa, e entre a característica com a heterose, não foi relatado coeficientes significativos. Estes autores dão suporte para a hipótese de que o grupo de fatores avaliados precisa ser informativo quanto ao fenótipo avaliado.

Recentemente, Akanno et al. (2018), utilizaram ferramentas genômicas para prever a heterose em bovinos de corte cruzados e encontraram evidências da contribuição da heterose na seleção dos animais, pois ao incluir esse efeito nos modelos genômicos, houve melhora na acurácia de predição dos valores genéticos em até 20%. Além disso, encontraram associações positivas entre efeitos genéticos não-aditivos com as características de crescimento, destacando a importância da heterose na inclusão em análises genômicas para explicar a variância fenotípica destas características. Da mesma forma, Su et al. (2012), ao comparar diferentes modelos genômicos, relataram uma melhora na capacidade preditiva das estimativas genômicas ao utilizar modelos com os efeitos de dominância.

Adicionalmente, alguns estudos com GWAS foram desenvolvidos para explorar a heterozigose a nível molecular. Recentemente, um estudo com animais da raça Holandesa, objetivou detectar variações genéticas não-aditivas associadas à resistência à tuberculose bovina. Neste estudo, os autores adaptaram o modelo padrão de GWAs para identificar as variações genéticas não-aditivas que, no modelo convencional, não eram detectadas, desse modo,

encontraram uma região genômica candidata em que a heterozigose do loco estava relacionada ao aumento da susceptibilidade a esta doença, ou seja, os animais heterozigotos têm maior probabilidade de ficarem doentes (Tsairidou et al., 2018).

Os estudos desenvolvidos com o objetivo de explorar a utilização dos efeitos genéticos em modelos de avaliação genômica, são promissores para a compreensão das variações fenotípicas. O objetivo dos programas de melhoramento é selecionar os animais com potencial genético favorável, para que desta forma os alelos benéficos para as características selecionadas sejam fixados na população (Cole e VanRaden, 2011). Isto posto, é prudente que a diversidade genética seja mantida na população, para que o efeito desfavorável de alelos e genes não cause decréscimo no desempenho produtivo dos animais. Assim, o conhecimento de regiões com alta heterozigose pode contribuir na identificação de genes candidatos que possibilitem a manutenção da diversidade genética, além de melhorar a performance das populações de gado leiteiro (Williams et al., 2015).

REFERÊNCIAS

- Abdollahi-Arpanahi, R.; Gianola, D.; Peñagaricano, F. Deep learning versus parametric and ensemble methods for genomic prediction of complex phenotypes. **Genetics Selection Evolution**, v.52, n.12, 2020.
- Akanno, E. C.; Chen, I.; Abo-Ismael, M. K.; Crowley, J. J.; Wang, Z.; Li, C.; Basarab, J. A.; MacNeil, M. D.; Plastow, G. Genomic prediction of breed composition and heterosis effects in Angus, Charolais, and Hereford crosses using 50K genotypes. **Canadian Journal Animal Science**, v.97, p.431-438, 2017.
- Alillo, H.; Pryce, J. E.; González-Recio, O.; Cocks, B. G.; Hayes, B. J. Accounting for dominance to improve genomic evaluations of dairy cows for fertility and milk production traits. **Genetics Selection Evolution**, v.48, n.8, p.1-11, 2016.
- Amuzu-aweh, E. N.; Kinghorn, B. P.; Vereijken, A.; Visscher, J.; Am van Arendonk, J.; Bovenhuis, H. Prediction of heterosis using genome-wide SNP-marker data: application to egg production traits in White Leghorn crosses. **Heredity**, v.11, p.530-538, 2013.
- Bonafous, F.; Fievet, G.; Blanchet, N.; Carrère, S.; Gouz, J.; Legrand, L.; Marage, G.; Bret-Mestries, E.; Munos, S.; Pouilly, N.; Vincourt, P.; Langlade, N.; Mangin, B. Comparison of GWAS models to identify non-additive genetic control of flowering time in sunflower hybrids. **Theoretical and Applied Genetics**, v.131, p.319-332, 2018.
- Bradford, H. L.; Masuda, Y.; VanRaden, P. M.; Legarra, A.; Misztal, I. Modeling missing pedigree in single-step genomic BLUP. **Journal of Dairy Science**, v.102, p.2336-2346, 2019.

- Clasen, J. B.; Norberg, E.; Madsen, P.; Pederson, J.; Kargo, M. Estimation of genetic parameters and heterosis for longevity in crossbred Danish dairy cattle. **Journal of Dairy Science**, v.100, n.8, p.6337-6342, 2017.
- Cole, J. B. & Vanraden, P. M. Use of haplotypes to estimate Mendelian sampling effects and selection limits. **Journal Animal Breeding and Genetics**, v.128, p.446-455, 2011.
- Hayes, B. J. Overview of statistical methods for genome-wide association studies (GWAS). **Methods in Molecular Biology**, v.1019, p.146-169, 2013.
- Lung, L. H. S.; Petrini, J.; Ramírez-Díaz, J.; Salvian, M.; Ravadoscki, G. A.; Pilonett, F.; Dauria, B. D.; Machado, P. F.; Coutinho, L. L.; Wiggans, G. R.; Mourão, G. B. Genome-wide association study for milk production traits in a Brazilian Holstein population. **Journal of Dairy Science**, v.102, n.6, p.5305-5314, 2019.
- Jiang, J.; Shen, B.; O'Connell, R.; VanRaden, P. M.; Cole, J. B.; Ma, L. Dissection of additive, dominance, and imprinting effects for production and reproduction traits in Holstein cattle. **BMC Genetics**, v.18, n.425, 2017.
- Kargo, M.; Clasen, J. B.; Nielsen, H. M.; Byskov, K.; Norberg, E. *Short communication*: Heterosis and breed effects for milk production and udder health traits in crosses between Danish Holstein, Danish Red, and Danish Jersey. **Journal of Dairy Science**, v.104, n.4, p.678-682, 2021.
- Kurz, J. P.; Yang, Z.; Weiss, R. B.; Wilson, D. J.; Rood, K. A.; Liu, G. E.; Wang, Z. A genome-wide association study for mastitis resistance in phenotypically well-characterized Holstein dairy cattle using a selective genotyping approach. **Immunogenetics**, v.71, n.1, p.35-47, 2019.
- Mao, X.; Sahana, G.; Johansson, M. A.; Liu, A.; Ismael, A.; Lovendahl, P.; De Koning, D.-J.; Guldbbrandtsen, B. Genome-wide association mapping for dominance effects in female fertility using real and simulated data from Danish Holstein cattle. **Nature**, v.10, n.2953, 2020.
- Meuwissen, T. H. E.; Hayes, B. J.; Goddard, M. E. Prediction of total genetic value using genome-wide dense marker maps. **Genetics**, v.157, p.1819-1829, 2001.
- Moghaddar, N.; Van der Werf, J. H. J. Genomic estimation of additive and dominance effects and impact of accounting for dominance on accuracy of genomic evaluation in sheep populations. **Journal Animal Breeding and Genetics**, v.134, p.453-462, 2017.
- Nicolini, P.; Amorín, R.; Han, Y.; Peñagaricano, F. Whole-genome scan reveals significant non-additive effects for sire conception rate in Holstein cattle. **BMC Genetics**, v.19, n.14, 2018.
- Nietlisbach, P.; Keller, L. F.; Postma, E. Genetic variance components and heritability of multiallelic heterozygosity under inbreeding. **Heredity**, v.116, p.1-11, 2016.
- Oakey, H.; Cullis, B.; Thompson, R.; Comadran, J.; Halpin, C.; Waugh, R. Genomic Selection in Multi-environment Crop Trials. **Genomics Selection**, v.6, p.1313-1327, 2016.
- Owens, T. W.; Rogers, R. L.; Best, S.; Ledger, A.; Mooney, A. A.; Ferguson, A.; Shore, P.; Swarbrick, A.; Ormandy, C.; Simpson, P. T.; Carroll, J. S.; Visvader, J.; Naylor, M. Runx2 is a novel regulator of mammary epithelial cell fate in development and breast cancer. **Cancer Research**, v.74, n.18, p.5277-5286.

- Perez, B. C.; Balieiro, J. C. C.; Carvalheiro, R.; Tirelo, F.; Oliveira Junior, G. A.; Dementshuk, J. M.; Eler, J. P.; Ferraz, J. B. S.; Ventura, R. V. Accounting for population structure in selective cow genotyping strategies. **Journal of Animal Breeding and Genetics**, v.136, p.23-39, 2018.
- Petrini, J.; Iung, L. H. S.; Rodriguez, M. A. P.; Salvian, M.; Pértille, F.; Ravadoscki, G. A.; Cassoli, L. D.; Coutinho, L. L.; Machado, P. F.; Wiggans, G. R.; Mourão, G. B. Genetic parameters for milk fatty acids, milk yield and quality traits of a Holstein cattle population reared under tropical conditions. **Animal Breeding and Genetics**, v.133, n.5, p.384-395, 2016.
- Rebelato, A. B.; Caetano, A. R. Runs of homozygosity for autozygosity estimation and genomic analysis in production animals. **Pesquisa Agropecuária Brasileira**, v.53, n.9, p.975-984, 2018.
- Resende, M. D. V.; Lopes, P. S.; Silva, R. L.; Pires, I. E. Seleção genômica ampla (GWAS) e maximização da eficiência do melhoramento genético. **Pesquisa Florestal Brasileira**, v.56, p.63-77, 2008.
- Saborío-Montero, A.; Vargas-Leitón, B.; Romero-Zúñiga, J. J.; Camacho-Sandoval, J. Additive genetic and heterosis effects for milk fever in a population of Jersey, Holstein x Jersey, and Holstein cattle under grazing conditions. **Journal of Dairy Science**, v.101, n.10, p.9128-9134, 2018.
- Seno, L. de O.; Guidolin, D. G. F.; Aspilcueta-Borquis, R.; Nascimento, G. B.; Silva, T. B. R. da; Oliveira, H. N. de; Munari, D. P. Genomic selection in dairy cattle simulated populations. **Journal of Dairy Research**, v.85, p.125-132, 2018.
- Shen, G.; Zhan, W.; Chen, H.; Xing, Y. Dominance and epistasis are the main contributors to heterosis for plant height in rice. **Plant Science**, v.11, n.18, 2014.
- Steri, R.; Moioli, B.; Catillo, G.; Galli, A.; Buttazzoni, L. Genome-wide association study for longevity in the Holstein cattle population. *Animal*, v.3, p.1-8, 2018.
- Su, G.; Christensen, O. F.; Ostensen, T.; Henryon, M.; Lund, M. S. Estimating additive and non-additive genetic variances and predicting genetic merits using genome-wide dense single nucleotide polymorphism markers. **PLoS ONE**, v.7, n.9, 2012.
- Sun, C.; VanRaden, P. M.; Cole, J. B.; O'Connell, J. R. Improvement of prediction ability for genomic selection of dairy cattle by including dominance effects. **PLoS ONE**, v.9, n.8, 2014.
- Tsairidou, S.; Allen, A. R.; Pong-Wong, R.; McBride, S. H.; Wright, D. M.; Matika, O.; Pooley, C. M.; McDowell, S. W. J.; Glass, E. J.; Skuce, R. A.; Bishop, S. C.; Woolliams, J. A. An analysis of effects of heterozygosity in dairy cattle for bovine tuberculosis resistance. **Animal Genetics**, v.49, p.103-109, 2018.
- Toro, M. A. & Verona, L. A note on mate allocation for dominance handling in genomic selection. **Genetics Selection Evolution**, v.42, n.33, 2010.
- VanRaden, P. M. Efficient Methods to Compute Genomic Predictions. **Journal of Animal Science**, v.91, p.4414-4423, 2008.
- Varona, L. & Misztal, I. Prediction of parental dominance combination for planned mating, methodology and simulation results. **Journal of Dairy Science**, v.82, p.2186-2181, 1999.
- Wang, W. Y. S.; Barratt, B. J.; Clayton, D. G.; Todd, J. A. Genome-wide association studies: theoretical and practical concerns. **Genetics**, v.6, p.109-118, 2005.

- Wei, M. & Van der Werf, J. H. J. Maximizing genetic response in crossbreds using both purebred and crossbred information. **Animal Production**, v.59, p.401-413, 1994.
- Wellmann, R.; Bennewitz, J. Bayesian models with dominance effects for genomic evaluation of quantitative traits. **Genetics Research Cambridge**, v.94, p.21-37, 2012.
- Williams, J. L. et al. Inbreeding and purging at the genomic Level: the Chillingham cattle reveal extensive, non-random SNP heterozygosity. **Animal Genetics**, v. 47, p.19-27, 2015.
- Xiang, T.; Christensen, O. F.; Vitezica, Z. G.; Legarra, A. Genomic evaluation by including dominance effects and inbreeding depression for purebred and crossbred performance with an application in pigs. **Genetics Selection Evolution**, v.48, n.92, 2016.
- Xiang, T.; Christensen, F. O.; Vitezica, G. Z.; Legarra, A. Genomic model with correlation between additive and dominance effects. **Genetics**, v.209, p.711-723, 2018.

3. COMPONENTES DE VARIÂNCIA, ACURÁCIA E VIÉS DE PREDIÇÃO ESTIMADOS POR MODELO GENÔMICO ADITIVO E ADITIVO-DOMINANTE EM POPULAÇÕES SIMULADAS E REAIS DE VACAS DA RAÇA HOLANDESA

RESUMO

O objetivo das avaliações genéticas é obter as melhores previsões dos valores genéticos dos animais para a seleção. Os métodos tradicionais consideram apenas os efeitos genéticos aditivos, como principal ajuste do fenótipo. Contudo, sabe-se que a variabilidade fenotípica de uma população vai além de efeitos aditivos. Neste estudo, o objetivo foi verificar a capacidade de previsão de valores genéticos ao incluir os efeitos de dominância nos modelos genômicos. Inicialmente, um estudo de simulação estocástica foi utilizado para testar o modelo puramente aditivo e outro aditivo-dominante em três cenários, simulados com magnitudes de herdabilidade de 0,10, 0,30 e 0,50. Após a fase de simulação, um conjunto de informações de produção e qualidade do leite foi avaliado, utilizando as mesmas premissas do estudo de simulação. Foram estimados os componentes de variância e herdabilidade para cada cenário e para os registros de produção de leite, concentração de proteína, gordura, caseína e conteúdo de ácido graxo poli-insaturado, de uma população de vacas da raça holandesa. As estimativas de proporção da variância de dominância em relação ao fenótipo foram de 2,9 a 5,2% em dados simulados, e de 0,08 a 3,8% para as características produção e qualidade do leite. A maior acurácia foi de 0,79 e 0,36 no modelo aditivo e aditivo-dominante simulado, respectivamente. Para os dados reais, as estimativas foram semelhantes entre os modelos, sendo entre 0,21 e 0,38. No estudo de simulação, as previsões foram viesadas para o cenário de baixa a moderada herdabilidade e para em todos os cenários para o modelo aditivo-dominante. Para as características de produção, o melhor desempenho do modelo foi para gordura e ácido graxo, com vantagem para ácido graxo quando incluído os efeitos de dominância na avaliação. De modo geral, não houveram grandes avanços e contribuição dos efeitos de dominância na estimativa dos componentes de variância e na habilidade de previsão dos valores genéticos, exceto para o conteúdo de ácido graxo poli-insaturado.

Palavras-chave: Ácido graxo. Caseína. Efeito genético. GBLUP. Produção de leite.

ABSTRACT

The purpose of genetic evaluations is to obtain the best predictions of the genetic values of the animals candidates for selection. Traditional methods consider only additive genetic effects as the main phenotype adjustment. However, we know that the phenotypic variability of a population is beyond the additive effects. In this study, the objective was to verify the ability predict to predict the genetic values, accounting dominance effects in genomic models. First, a stochastic simulation study was used to test the purely additive and additive-dominant model in three scenarios simulated with heritability magnitudes of 0.10, 0.30 and 0.50. After the simulation, a set of milk production and quality traits was evaluated using the same assumptions of

the simulation study. The components of variance and heritability were estimated for each scenario and for records of milk production, protein percentage, fat percentage, casein percentage and polyunsaturated fatty acid content of a population of Holstein cows. Estimates of the proportion of dominance variation relative to phenotypic variation were 2.9 to 5.2% in simulated data, and from 0.08 to 3.8% for the milk yield and quality traits. The highest accuracy was 0.79 and 0.36 in the simulated additive and additive-dominant model. For real data, estimates were similar between models, being between 0.21 and 0.38. In the simulation study, predictions were biased for the low to moderate heritability scenario and for all scenarios for the additive-dominant model. The best performance of the model was for fat and polyunsaturated fatty acid, with advantage in prediction when the dominance effects were included. In general, there were no great advances and contribution of dominance effects in the estimations and in the predicted ability genomics values, except for the polyunsaturated fatty acid content.

Keywords: Fatty acid. Casein. Genetic effect, GBLUP. Milk production.

3.1. INTRODUÇÃO

Os modelos genômicos para avaliação de características quantitativas, normalmente são ajustadas para estimar os efeitos aditivos de cada marcador molecular, por meio de uma matriz de relacionamento genômico (Meuwissen et al., 2001; VanRaden et al., 2008; de los Campos et al., 2009; Goddard et al., 2009). Tal efeito é também chamado de efeito de substituição alélica, que representa a mudança média do valor genotípico, quando um alelo A1 é aleatoriamente substituído por A2 em uma população de indivíduos qualquer (Falconer e Mackay 1996). Por esse motivo, a seleção dos animais ocorre devido a ação genética aditiva dos genes ou sobre os efeitos de substituição, pois os alelos, não os genótipos, serão transmitidos a progênie.

O interesse teórico e prático no estudo dos efeitos não-aditivos foi renovado com o advento da genômica (Vitezica et al., 2013; Akanno et al., 2017; Tsairidou et al., 2018; Nicolini et al., 2018). No entanto, ainda é um desafio incorporá-lo nas avaliações genômicas, devido à complexidade matemática atribuída aos modelos de predição e a falta de pedigrees altamente informativos (Misztal et al., 1998; Mrode e Thompson, 2005; Vitezica et al., 2013; Varona et al., 2018).

Não é claro o impacto do efeito de dominância na capacidade preditiva dos modelos de avaliação genética. Em diferentes espécies, incluir a dominância na predição dos valores genéticos não apresentou nenhuma vantagem preditiva, comparado ao modelo aditivo (Garcia-Baccino et al., 2019; Ertl et al., 2014; Xiang et al., 2016; Moghaddar e van der Werf, 2017). Porém, outros estudos relatam vantagens em acurácia de predição de até 17% e proporção da variância

fenotípica devido ao efeito de dominância de até 9,5% (Toro e Varona, 2010; Wellman e Benewitz, 2012; Su et al., 2012; Garcia-Bacino et al., 2020).

Dessa forma, o objetivo desse estudo foi quantificar e incorporar os efeitos de desvio de dominância em modelos avaliação genômica e avaliar a capacidade de estimação dos componentes de variância, sob as premissas de simulação de dados estocástica e em dados reais, para bovinos leiteiros da raça Holandesa.

3.2. MATERIAL E MÉTODOS

3.2.1. Simulação da estrutura populacional – Etapa I

A estrutura de população e as informações genômicas foram simulados utilizando o software QMSim (Figura 1) (Sargolzaei e Schenkel, 2009). Três cenários genéticos foram simulados para imprimir fenótipos com magnitude de herdabilidade (h^2) de 0,10, 0,30 e 0,50, que compreendem, basicamente, características de produção e qualidade do leite. Foi simulada uma população de gado leiteiro relativamente pequena, semelhante a trabalhos anteriores (Perez et al., 2018; Compton et al., 2017 e Plieshke et al., 2016), considerando 20 repetições.

Em seguida, foi criada uma população histórica composta por 2.000 animais não relacionados e com proporção sexual equilibrada (50% de machos e 50% de fêmeas). Esses animais foram acasalados aleatoriamente por 1.000 gerações e um gargalo genético foi introduzido a partir da geração 1.001, reduzindo a população histórica de 2.000 para 200 animais por 20 gerações. O gargalo genético é utilizado para criar um cenário que exista desequilíbrio de ligação (LD) consistente na população simulada (Bradford et al., 2019).

A partir dessa população, uma segunda foi gerada e chamada de expansão, na qual os fundadores eram 100 machos e 100 fêmeas, amostrados aleatoriamente da geração 1.020 da população histórica. Essa população em expansão foi acasalada aleatoriamente por mais 60 gerações, que deu origem a 10.000 filhos, sendo 150 machos e 9.850 fêmeas na última geração (geração 1.080). A partir da população de expansão, foi criada uma população denominada recente, na qual foram selecionados aleatoriamente 50 machos e 1.500 fêmeas da geração 1.080, sendo que os acasalamentos foram administrados em 10 gerações.

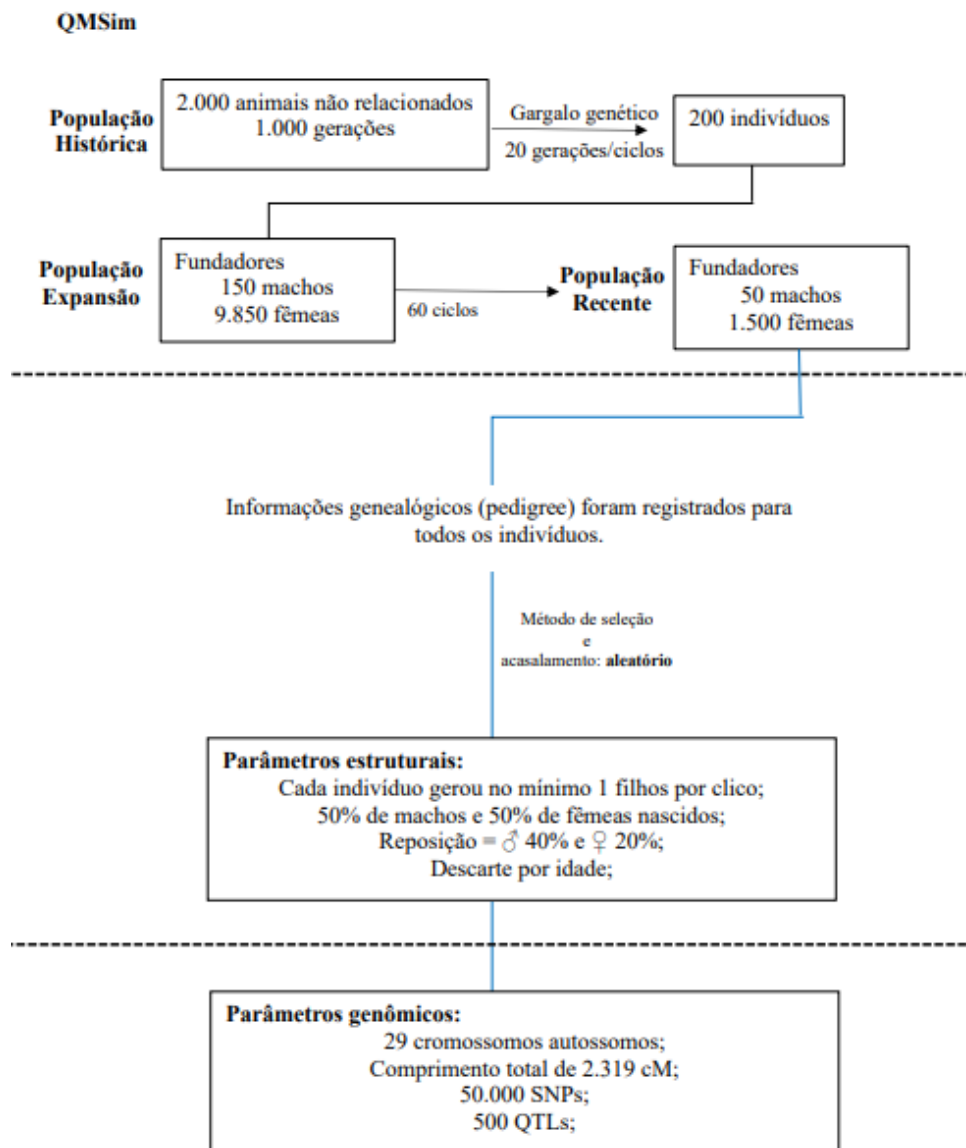


Figura 1. Esquema ilustrativo do processo de simulação.

Em cada geração, eram formados aproximadamente 764 machos e 891 fêmeas, que ao final totalizaram 16.500 animais em 10 gerações. Foi utilizado uma taxa de descarte de 40% de machos e 20% fêmeas, para que dessa forma, a cada geração os progenitores fossem substituídos ao ponto de que todos pudessem ter a oportunidade de deixar uma progênie. Assim, a cada geração, 20 reprodutores e 300 matrizes eram substituídos e seus registros genealógicos anotados. Uma seleção aleatória dos animais foi adotada como parâmetro, no intuito de trabalhar com uma população que não sofreu nenhum tipo de seleção artificial, proporcionando a mesma oportunidade para gerar descendentes.

Simulação do genoma

O genoma foi simulado para 29 pares de cromossomos autossômicos (Figura 1), com comprimento total de 2.319 cM, imitando o genoma bovino. As características foram controladas por 500 QTLs bialélicos (Quantitative Trait Locus), que foram espaçados de forma aleatória por todo o genoma. Além disso, os efeitos de QTL foram simulados semelhante a Bradford et al. (2019), na qual obedeceram a uma distribuição gama de parâmetro 0,4 para os três cenários genéticos.

Adicionalmente, havia 50.200 marcadores moleculares do tipo SNP (Single Nucleotide Polymorphism) bialélicos distribuídos equidistantes por todo o genoma. Para SNP e QTL, a frequência alélica foi de 0,50 para iniciar a população histórica e $2,5 \times 10^5$ de taxa de mutação, sendo a posição de ambos distribuída aleatoriamente ao longo do genoma.

3.2.2. Simulação dos efeitos genéticos aditivos e de desvios de dominância – Etapa II

O software QMSim é limitado para condições em que as bases estocásticas de simulação derivam apenas de efeitos genéticos aditivos. Dessa forma, é necessário que os efeitos não-aditivos sejam obtidos externamente, em ambientes que permitam o desenvolvimento integrado e execução de comandos e funções. Nesta etapa, o software R versão 4.1.0 (R Core Team, 2020) foi utilizado como ambiente principal de simulação. Adicionalmente, para tornar a simulação padronizada, apenas a estrutura de população e o genoma obtida pelo QMSim foram utilizados nesta etapa, isto é, os efeitos aditivos, valor genético verdadeiro e os fenótipos foram obtidos concomitante à simulação dos efeitos de dominância.

Os parâmetros populacionais permaneceram os mesmos, sendo eles 16.500 indivíduos gerados em 10 gerações de acasalamento sob seleção aleatória e três características quantitativas de diferentes magnitudes de herdabilidade, definidas como o cenário de estudo. Foi gerado um total de 60 arquivos genômicos, advindos de 20 repetições para cada cenário genético simulado: $h_1^2 = 0,10$; $h_2^2 = 0,30$; $h_3^2 = 0,50$. Os genótipos foram recodificados como sendo 0 para homocigoto $A1A1$, 1 para heterocigoto $A1A2$ e $A2A1$, e 2 para homocigoto $A2A2$.

Um total de 500 QTLs foram simulados e distribuídos aleatoriamente ao longo do genoma, sendo que cada indivíduo foi genotipado para 50.200 marcadores genéticos do tipo SNP. Considerando que um QTL representa uma região genômica candidata, respeitosamente optou-se por utilizar o conceito de QTN (*Quantitative Trait Nucleotide*) devido a aproximação de um gene com o fenótipo simulado (Peltier et al., 2019; Abdollahi-Arpanahi et al., 2020). A

atribuição desse conceito é mais apropriada para as etapas simuladas, pois é possível identificar as regiões sob efeito causal de genes.

A estatística descritiva foi obtida pelo somatório dos valores das 20 repetições e extraído os valores de média, mínimo, máximo, desvio-padrão e a variância, para os fenótipos simulados. Além disso, com o conjunto de dados médios, foram construídos gráficos de dispersão e distribuição para os efeitos de QTN em função da frequência, e para os valores genéticos preditos, um gráfico de densidade foi adicionado. Abaixo segue a distribuição dos efeitos aditivos e de dominância dos QTNs, para o cenário I (Figura 2), II (Figura 3) e III (Figura 4).

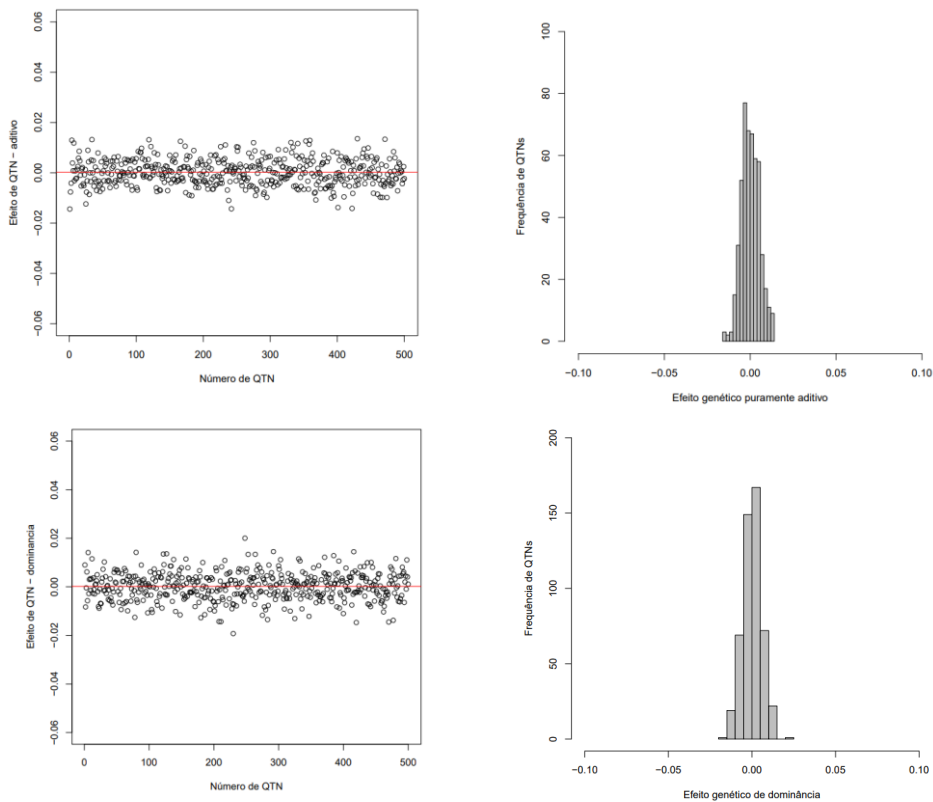


Figura 2. Gráfico de dispersão (à direita) e distribuição (à esquerda) dos efeitos médios de 500 QTNs simulados para efeito genética puramente aditiva (acima) e de dominância (abaixo) no cenário II (herdabilidade de 0,10). A linha vermelha representa a média aritmética.

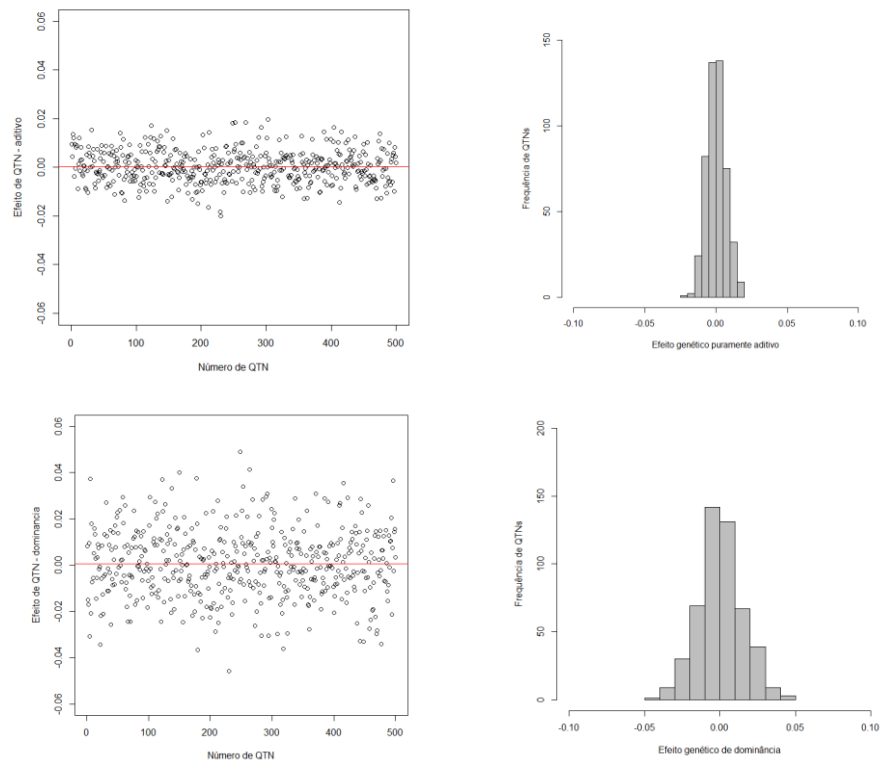


Figura 3. Gráfico de dispersão (à direita) e distribuição (à esquerda) dos efeitos médios de 500 QTNs simulados para efeito genética puramente aditiva (acima) e de dominância (abaixo) no cenário II (herdabilidade de 0,30). A linha vermelha representa a média aritmética.

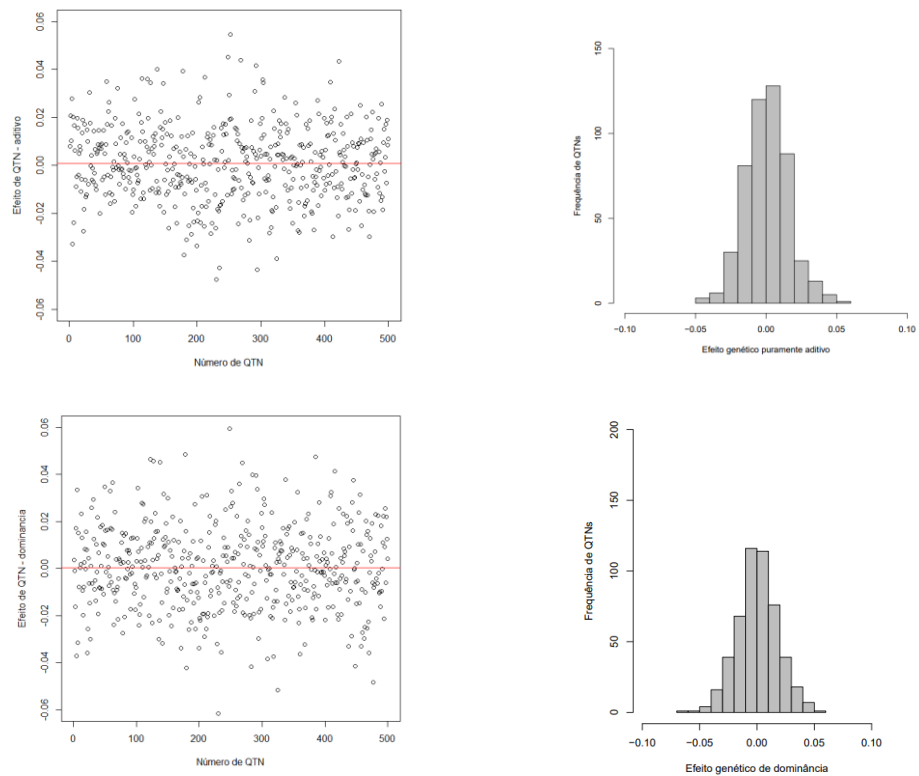


Figura 4. Gráfico de dispersão (à direita) e distribuição (à esquerda) dos efeitos médios de 500 QTNs simulados para efeito genética puramente aditiva (abaixo) e de dominância (acima) no cenário II (herdabilidade de 0,50). A linha vermelha representa a média aritmética.

Ação genética puramente aditiva

Para dimensionar os efeitos dos QTNs, a variação dos QTNs (k) foi calculada em função da sua frequência e , para atingir a herdabilidade pré-definida (h^2), a dimensão dos efeitos de QTN foram escalonados, dividindo h^2 em k (h^2/k). Dessa forma, os efeitos de substituição alélica ou ação genética aditiva (α), foram obtidos por meio de uma distribuição normal padrão $\alpha \sim N(\mu = 0, \sigma = h^2/k)$. Para obter o efeito puramente aditivo, o efeito de dominância (d) foi definido como sendo 0. Cada efeito foi calculado como sendo:

$$\alpha = a + d(q - p)$$

onde α é o efeito de substituição alélica; a e d são os efeitos aditivos e de dominância, respectivamente; p é a frequência alélica com $q = p - 1$.

Adicionalmente, foram estimados os valores genéticos aditivos, definidos aqui como valor genético verdadeiro (*True Breeding Value* – TBV), realizando a multiplicação dos códigos dos genótipos pelos efeitos de substituição dos QTNs.

Os efeitos aditivos e não-aditivos, foram distribuídos de forma que, aproximadamente 50% deles apresentasse sinal positivo e 50% negativo. Na figura 5 são apresentadas as distribuições dos TBVs para efeito aditivo e não-aditivo, para os três cenários. Os pontos representam indivíduos com TBV mais distantes da média e em azul a densidade de indivíduos com TBV próximo da média (linha vermelha).

Os valores fenotípicos (Tabela 1) foram obtidos para cada indivíduo (y_i). Para que cada i receba-se um valor de fenótipo, foi adicionado um valor residual normalmente distribuído $e_i \sim N(0, \sigma_e^2)$, sobre a soma de um QTN (valor genético verdadeiro):

$$y_i = \sum_{k=1}^m X_{ik} \alpha_k + e_i$$

onde X_{ik} ($i = 1, \dots, n$; $k = 1, \dots, m$) é um componente de incidência da matriz de marcadores genéticos para os efeitos aditivos (α_k) e e_i é o valor residual aleatório, sendo σ_e^2 a variância residual.

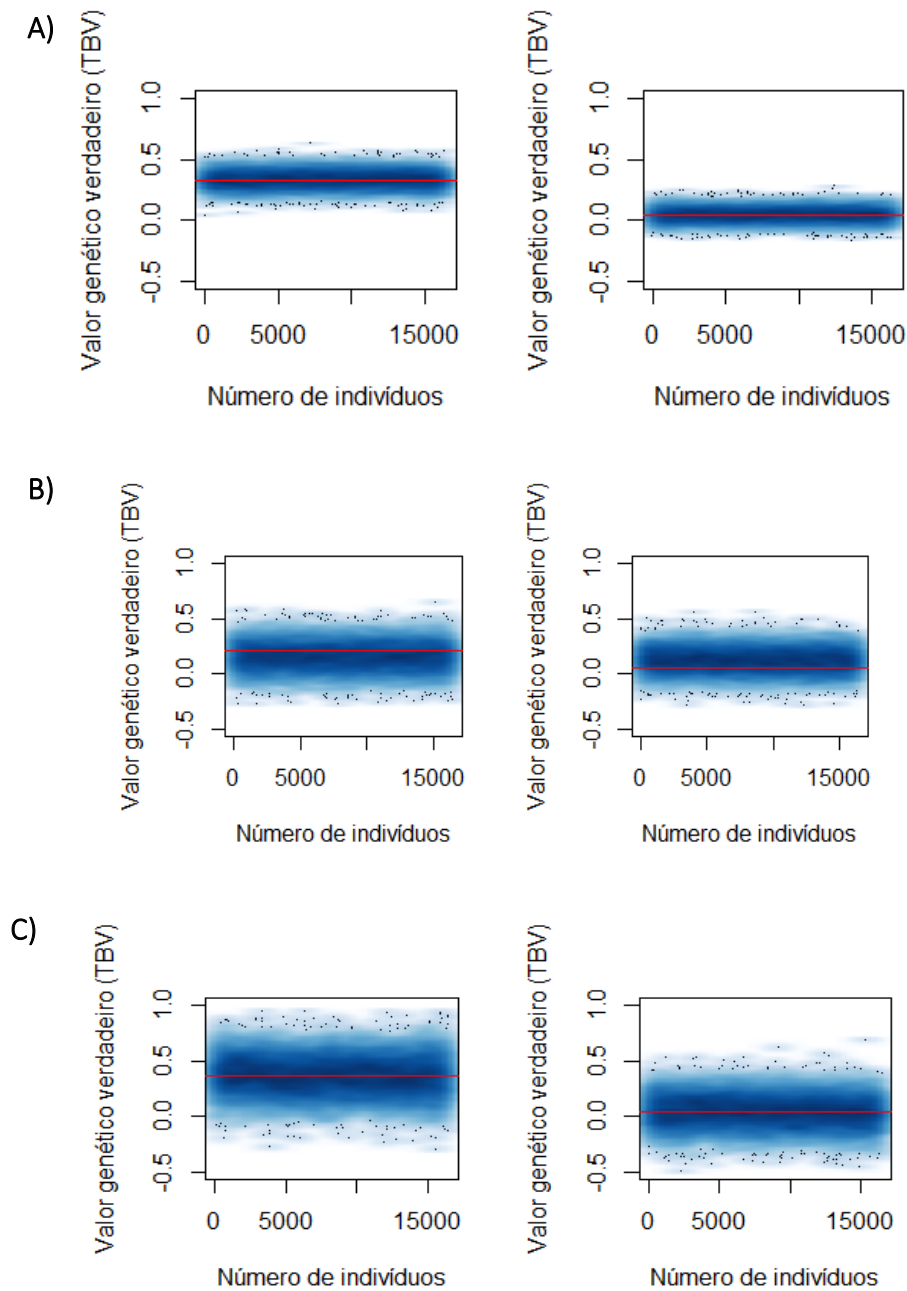


Figura 5. Gráfico de densidade dos TBVs simulados para efeito genético puramente aditivo à direita e de dominância (esquerdo), no cenário $h^2 = 0,10; 0,30$ e $0,50$ (**A**, **B** e **C**, respectivamente). A linha vermelha representa a média aritmética.

Ação genética não-aditiva (desvios de dominância)

Para modelar os efeitos genéticos de dominância, os elementos de incidência da matriz **D** de dominância, foram codificados como 0, 1 e 0, para os genótipos *A1A1*, *A1A2/A2A1* e *A2A2*, respectivamente. De acordo com Abdollahi-Arpanahi et al. (2020), eventos genéticos como a

dominância incompleta podem ser parcialmente captados em um modelo aditivo, dessa forma, a estrutura de simulação foi construída para sobredominância (*overdominance*).

Os parâmetros utilizados nessa etapa foram semelhantes à simulação dos efeitos aditivos. Dado que os desvios de dominância, do ponto de vista estatístico, representam a interação entre os alelos, o valor médio é obtido por meio do desvio médio entre genótipos heterozigotos e homozigotos (Falconer e Mackay, 1996). Dessa forma, para modelar a ação de dominância, a distribuição seguiu o estudo de Abdollahi-Arpanahi et al. (2020), sendo $d_0 \sim N(\mu = 0, \sigma = 0,50)$.

Cada efeito de dominância foi formulado como sendo $d = \alpha * (\sum d_0)$, onde d é o desvio de dominância; α os efeitos de substituição alélica anteriormente calculado (efeito aditivo); e d_0 é o efeito de dominância dimensionado para cada QTN, distribuídos ao longo do genoma. Assim como para o cenário aditivo, os TBVs para os desvios de dominância foram obtidos multiplicando os códigos genotípicos aos efeitos de dominância de QTN. Os fenótipos para efeito de dominância+aditivo (Tabela 1), foram calculados somando a matriz dos efeitos genético aditivos, com os valores genéticos de QTN para dominância, adicionados a um valor residual normalmente distribuído $e_i \sim N(0, \sigma_e^2)$. A descrição dos valores está apresentada na tabela 6.

$$y_i = \sum_{j=1}^m X_{ij}\alpha_j + \sum_{j=1}^m Z_{ij}d_j + e_i$$

onde y_i é o fenótipo para cada indivíduo i ; X_{ij} e Z_{ij} ($i = 1, \dots, n$; $j = 1, \dots, m$) são os componentes de incidência da matriz de marcadores genéticos para os efeitos aditivos e de dominância (α_j e d_j , respectivamente) e e_i é o valor residual aleatório, sendo σ_e^2 a variância residual.

Tabela 1. Número de observações (N), média, desvio padrão (DP), variância (VAR), mínimo (MIN) e máximo (MAX) para os fenótipos simulados em cada cenário genético (herdabilidade, h^2), sob efeito genético puramente aditivo e de desvio de dominância.

Cenário (h^2)	N	Média	DP	VAR	MIN	MAX
0,10	16.500	0,326	0,333	0,111	-1,094	1,799
0,30	16.500	0,468	0,390	0,151	-1,057	2,021
0,50	16.500	0,359	0,265	0,070	-0,756	1,335
Ação genética de dominância						
0,10	16.500	0,369	0,292	0,085	-0,918	1,485
0,30	16.500	0,734	0,389	0,152	-0,831	2,193
0,50	16.500	0,399	0,349	0,121	-0,960	1,681

3.2.3. Dados simulados

Os componentes de (co)variância foram estimados usando o método de máxima verossimilhança restrita com o algoritmo de informação média sob os modelos genômicos-poligênicos, no software AIRMLF90 (Miszta et al., 2018). Uma amostra de 1500 indivíduos, pertencentes a última geração de acasalamento (geração 10), foi utilizada para as análises de estimação e predição dos valores genéticos genômicos (GEBV). A matriz de relacionamento genômico \mathbf{G} foi utilizada seguindo um modelo genético univariado ajustado:

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

onde \mathbf{y} é o vetor de fenótipos simulados; \mathbf{X} é o designer da matriz de efeitos fixos; $\boldsymbol{\beta}$ é o vetor de efeitos fixos; \mathbf{Z} é o designer da matriz de efeitos aleatórios; \mathbf{a} é o vetor dos efeitos aleatórios; $\boldsymbol{\varepsilon}$ é o vetor dos efeitos residuais, com $\boldsymbol{\varepsilon} \sim \text{NID}(0, \mathbf{I}\sigma^2)$. O vetor $\boldsymbol{\beta}$ inclui os efeitos de sexo dos indivíduos e assumimos $\mathbf{a} \sim \text{N}(0, \mathbf{G}\sigma_a^2)$.

A estrutura de variância-covariância genômica foi construída a partir dos animais que possuíam fenotipagem e genótipos para um painel com 50.000 SNPs. Os programas da família BLUPF90 suportam uma etapa única para análises genômicas (*single-step* GBLUP; ssGBLUP), na qual uma matriz \mathbf{A} de relacionamento aditivo de pedigree é combinada com a matriz \mathbf{G} para criar uma matriz única de relacionamento genômico \mathbf{H} (Legarra, et al., 2009; Miszta et al., 2009; Aguilar et al., 2010).

Assim, a inversa de \mathbf{H}^{-1} pode ser expressa como:

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

em que $\mathbf{G}_w = w\mathbf{G} + (1 - w)\mathbf{A}_{22}$; \mathbf{G}_w é a matriz final dos relacionamentos genômicos; w o peso dos marcadores moleculares, que representa a proporção da variância genética aditiva dos SNPs; \mathbf{A}_{22} uma submatriz de \mathbf{A} para os animais genotipados; \mathbf{G} a matriz de relacionamento genômico.

Como nessa etapa de avaliação todos os animais foram genotipados, nos interessa apenas utilizar a matriz \mathbf{G} . Portanto, omitir \mathbf{A}^{-1} funciona como opção no software RENUMF90 e dessa forma, \mathbf{G}_w^{-1} é preparada combinada com \mathbf{A}_{22} , uma vez que essa última não é avaliada quando $\mathbf{G}_w = 0,99\mathbf{G} + (1 - 0,99)\mathbf{A}_{22} = \mathbf{G}_w = 0,99 - 0,01 \mathbf{a}_{11}$ (Massuda, 2019; Christensen e Lund, 2010). A constante w recebeu um ponderar igual a 0,99 como forma de mimetizar o uso da matriz de pedigree (\mathbf{A}) e assim, o software realizar um pseudo *single-step* GBLUP (Aguilar, 2014).

Dessa forma, a matriz \mathbf{G} foi calculada conforme VanRaden (2008) utilizando as informações dos marcadores moleculares, na qual o numerador é análogo à matriz de relacionamento aditivo de pedigree:

$$\mathbf{G} = \frac{\mathbf{Z}\mathbf{Z}'}{2\sum_i p_j(1 - p_j)}$$

onde \mathbf{Z} contém os genótipos ajustados com a frequência alélica e p_j a frequência dos alelos do marcador j . Aqui, o coeficiente de endogamia genômico individual é amostrado como $\mathbf{G}_{jj} - 1$, e o relacionamento entre os indivíduos j e k são obtidos pela divisão dos elementos \mathbf{G}_{jk} pela raiz quadrada da diagonal \mathbf{G}_{jj} e \mathbf{G}_{kk} .

Para as análises que incluem o efeito de dominância, o modelo utilizado segue:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{a} + \mathbf{W}\mathbf{d} + \boldsymbol{\varepsilon}$$

onde \mathbf{y} é o vetor de fenótipos simulados; \mathbf{X} é o designer da matriz de efeitos fixos; $\boldsymbol{\beta}$ é o vetor de efeitos fixos; \mathbf{Z} e \mathbf{W} são os designers da matriz de efeitos aleatórios; \mathbf{a} e \mathbf{d} são os vetores dos efeitos genéticos aditivos e de desvio de dominância, respectivamente; $\boldsymbol{\varepsilon}$ é o vetor dos efeitos residuais, com $\boldsymbol{\varepsilon} \sim \text{NID}(0, \mathbf{I}\sigma^2)$. O vetor $\boldsymbol{\beta}$ inclui os efeitos de sexo dos indivíduos e assumimos $\mathbf{a} \sim \mathbf{N}(0, \mathbf{G}\sigma_a^2)$ e $\mathbf{d} \sim \mathbf{N}(0, \mathbf{D}\sigma_D^2)$.

A matriz de relacionamento para efeitos de dominância foi construída conforme Vitezica et al. (2013), utilizando como controle de qualidade MAF de 0,05. A matriz seguiu sendo:

$$\text{cov}(\mathbf{v}) = \frac{\mathbf{W}\mathbf{W}'}{\sum(2p_iq_i)^2} \sigma_D^2 = \mathbf{D}\sigma_D^2$$

onde a matriz \mathbf{D} é obtida semelhante a matriz \mathbf{G} clássica de VanRaden (2008), na qual $\text{cov}(\mathbf{v})$ é a covariância dos desvios de dominância; \mathbf{W} são os efeitos de SNPs; p e q as frequências alélicas do marcador i .

Esse método é equivalente ao modelo clássico, que explicam dado fenótipo (\mathbf{y}) em função das variâncias estimadas. Contudo, a interpretação nesse caso, deve-se aos efeitos genômicos de desvio de dominância dos marcadores e para obter as predições dos GEBVs, esse efeito precisa ser incorporado como um efeito aleatório extra. O pacote AGHmatrix (Amadeu et al., 2016) do

software R (R Code Team, 2020) foi utilizado para construir as matrizes de relacionamento genômico para os desvios de dominância.

O controle de qualidade (QC) dos genótipos foi aplicado pelo *default* do software AIREMLF90, na qual foram excluídos os marcadores com *call-rate* menor que 90%, frequência alélica menor que 0,05 (*minor allele frequency*, *MAF*) e aqueles com frequência de heterozigotos maior que 0,15 do que o esperado pelo equilíbrio de Hardy-Weinberg. Cada repetição nos três cenários avaliados, eram compostos por 50.200 SNPs e após QC, restaram aproximadamente 45.000 ± 1.500 SNPs para os três cenários avaliados.

3.2.4. Dados reais

Fenótipos

Todos os procedimentos com os animais seguiram os padrões éticos de pesquisa, estabelecidos pela Universidade de São Paulo e demais instituições envolvidas, presente na proposta intitulada “Incorporação de informações genômicas no estudo de características de composição e qualidade do leite de vacas”, protocolo nº 2017.5.1197.11.3, aprovada pela comissão de ética no uso de animais – CEUA da ESALQ/USP.

O banco de fenótipos era composto por 7.664 vacas da raça Holandesa, de primeira a décima lactação, e 127.432 registros mensais para as características de produção e qualidade do leite. As mensurações eram oriundas de três fazendas brasileiras localizadas no Estado de São Paulo e uma no Paraná, coletados entre os anos de 2012 e 2016 em parceria com a Clínica do Leite, pertencente ao Departamento de Zootecnia da Escola Superior de Agricultura “Luiz de Queiroz”, da Universidade de São Paulo. Nesses rebanhos, as fêmeas eram criadas em sistema free-stall e o manejo de ordenha ocorria três vezes ao dia sob sistema mecânico. Adicionalmente, havia registros genealógicos disponíveis para todos os rebanhos.

Inicialmente foi realizado uma checagem de consistência de dados de forma individual para produção de leite no dia do controle leiteiro (PL, kg/dia), porcentagem de proteína (PROT, %), gordura (GORD, %), caseína (CAS, %) e conteúdo de ácido graxo poli-insaturado (POLI, %). Essas características foram selecionadas devido ao seu impacto econômico no sistema de produção leiteira e na saúde humana. Para os níveis aceitáveis de produção de leite e sólidos, foi utilizado como referência a diretrizes de registros fenotípicos para gado leiteiro (ICAR, 2017). Além disso, foram excluídos os animais sem nenhuma informação de fenótipo, com idade superior a 10 anos, dias em lactação menor que 5 e maior que 305 dias e aqueles animais que não possuíam informações de data de nascimento, coleta da amostra de leite e de parto, e os registros

com produção de leite abaixo de 3,0 e acima de 99,0 kg de leite/dia e para proteína e gordura, foi considerado os registros com desvio-padrão de até $\pm 3,0$, sendo uma variação aproximada de 1,0 à 9,0% e para gordura de 1,5 à 12,0% (índices para vacas de alta produção de leite). A estatística descritiva para as informações de fenótipo pode ser consultada na tabela 2.

A partir disso, os grupos de contemporâneos (GC) foram formados concatenando fazenda, mês e ano de parto e da coleta da amostra do leite. Uma análise de variância, ANOVA, foi realizada por meio do software R para testar o nível de significância das variáveis incluídas no GC. Os três fatores testados foram significativamente diferentes de zero pelo teste F, portanto as variáveis agrupadas foram utilizadas como efeito fixo nas análises genéticas subsequentes.

Tabela 2. Número de observações (N), média, desvio-padrão (DP), valores de mínimo e máximo obtidos para produção de leite (PL), concentração de proteína (PROT), gordura (GROD), caseína (CAS) e conteúdo de ácido graxo poli-insaturado (POLI).

Característica	N	Média	DP	MIN	MAX
PL (kg/dia)	15.780	32,19	11,53	3,02	80,10
PROT (%)	17.894	3,13	0,36	1,40	5,83
GORD (%)	17.809	3,49	0,79	1,50	11,92
CAS (%)	13.623	2,38	0,32	0,92	4,68
POLI (%)	12.438	0,16	0,05	0,01	1,09

Genótipos

Originalmente, o genoma da população era composto por 1.157 animais amostrados, sendo que 768 vacas foram genotipadas em painel de baixa densidade (Illumina Bovine LD BeadChip, 6K), com 6.909 polimorfismos de nucleotídeo único (Single Nucleotide Polymorphism, SNP) e 389 vacas genotipadas em painel de média densidade, composto por 47.843 SNPs (GeneSeek Genomic Profile Bovine 50K). Tais informações já estão disponíveis, tendo sido obtidas em projetos anteriores (CNPq 482327/2010-3; 484560/2012-3). Nas fazendas residentes do Estado de São Paulo foram coletadas amostras biológicas de 930 animais e na fazenda do Paraná, foram 227 vacas.

Foram coletadas amostras de bulbos pilosos extraídos da vassoura da cauda e a extração do material genético realizado por meio do Kit NucleoSpin Tissue (Macharey-Nagel GmbH & Co. KG, Duren, Alemanha) e a leitura dos chips de genotipagem feita no equipamento HiScanTMSQ System (Illumina, Inc., 2011). A extração de DNA e leitura dos chips foram realizadas no Laboratório de Biotecnologia Animal (Departamento de Zootecnia, ESALQ/USP).

Após a obtenção das informações genômicas, os dados genotípicos foram enriquecidos por análise de imputação, conforme detalhado por Petrini et al. (2015) e Jung et al. (2019). Para esse procedimento, a população referência foi formada por touros, que eram pais das vacas genotipadas, que possuíam genótipo para 79.294 SNPs nas plataformas Illumina BovineSNP50 (Illumina, San Diego, UEA) ou GeneSeek Genomic Profiler HD (Neogen Agrigenomics, Lexington, EUA). Essa análise de imputação foi desenvolvida pelo Animal Genomics and Improvement Laboratory (Agriculture Research Service, USDA) sob encargo do Dr. George Wiggans, que utilizou o programa findhap.f90 (VanRaden, 2015).

Subsequente a análise de imputação, oito animais com genótipos duplicados ou com parentesco incompatível foram excluídos, dessa forma 1.149 animais com informações de 79.294 marcadores moleculares restaram no banco de dados original. Na intenção de reduzir o esforço computacional e aumentar a estabilidade nas estimativas dos efeitos dos SNPs nas avaliações genômicas, foi aplicado um controle de qualidade (QC) que excluiu 7.899 marcadores com call-rate menor que 90%, 32 SNPs monomórficos e 1.241 com frequência alélica (minor allele frequency, MAF) menor que 0,02. Além disso, foi utilizado como critério de frequência de heterozigotos, valores maiores que 0,15 do que o esperado quando em equilíbrio de Hardy-Weinberg, no entanto nenhum marcador foi excluído.

Adicionalmente, também foram excluídos das análises os marcadores presentes nos cromossomos sexuais e 56 animais que apresentaram conflito de parentesco, após o primeiro corte realizado na análise de imputação. O QC foi realizado de forma específica para cada característica, no programa PREGSF90 (Aguilar et al., 2014). Dessa forma, para a produção de leite, foram efetivados 1.096 animais e 70.193 marcadores moleculares do tipo SNP.

3.2.5. Análises genéticas

Para as análises genéticas foram considerados apenas os animais que possuíam genótipo e fenótipo após os controles de qualidade, utilizando o ferramental GBLUP (Aguilar et al., 2014), da família de programas do software BLUPF90 (Misztal et al., 2018). Foram ajustados dois modelos mistos univariado e com medidas repetidas, para efeitos fixos e aleatórios:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{a} + \mathbf{S}\mathbf{c} + \boldsymbol{\varepsilon} \quad (\text{MGA})$$

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{a} + \mathbf{W}\mathbf{d} + \mathbf{S}\mathbf{c} + \boldsymbol{\varepsilon} \quad (\text{MGAD})$$

em que \mathbf{y} é o vetor de observações fenotípicas; \mathbf{X} é a matriz de incidência para os efeitos fixos de grupo de contemporâneos (GC); $\boldsymbol{\beta}$ é o vetor dos efeitos fixos; \mathbf{Z} a matriz de incidência para o efeito genético aditivo aleatório; \mathbf{a} é o vetor de soluções para efeito aleatório; \mathbf{W} é a matriz de incidência para os efeitos genéticos de dominância aleatório; \mathbf{d} é o vetor de soluções para o genético de dominância; \mathbf{S} é a matriz de incidência para o efeito aleatório de ambiente permanente; \mathbf{c} é o vetor de soluções para o efeito aleatório de ambiente permanente; $\boldsymbol{\varepsilon}$ é o vetor para os efeitos residuais.

Os efeitos seguiram distribuição:

$$\mathbf{a} \sim \mathbf{N}(\mathbf{0}, \mathbf{G}\sigma_a^2), \mathbf{d} \sim \mathbf{N}(\mathbf{0}, \mathbf{D}\sigma_d^2), \mathbf{e} \sim \mathbf{NID}(\mathbf{0}, \mathbf{I}\sigma_e^2)$$

onde \mathbf{G} é a matriz de relacionamento genômico para os efeitos aditivos do marcador molecular; \mathbf{D} a matriz de relacionamento genômico para os efeitos de dominância; e \mathbf{I} a matriz identidade para os efeitos residuais.

Os componentes de (co)variância foram estimados usando o método de máxima verossimilhança restrita com o algoritmo de informação média sob os modelos genômicos-poligênicos, no programa AIREMLF90, simulando uma etapa única de avaliação genômica (*single-step* GBLUP), por meio do programa BLUPF90 (Misztal et al., 2018). Por tanto, os modelos MGA e MGAD foram ajustados da mesma forma que para a etapa com os dados simulados, assim como a construção das matrizes de relacionamento genômico para o efeito de dominância.

3.2.6. Avaliação da acurácia e viés de predição

Em estudos de simulação de dados, é possível obter os valores genéticos verdadeiros (*True Breeding Value*, TBVs) e para verificar a habilidade dos modelos aditivos e aditivo-dominante, a acurácia foi estimada empiricamente por meio da correlação de Pearson entre TBVs e os valores genéticos genômicos (*Genomic Estimated Breeding Value*, GEBV) preditos via GBLUP. Todos os animais utilizados nessa etapa possuíam informação de genótipo e fenótipo, e a partir disso, os valores de correlação foram obtidos como:

$$\rho_{x,y} = \frac{cov(x,y)}{\sigma_x\sigma_y}$$

onde $\rho_{x,y}$ é a correlação entre TBV e GEBV; x e y representam o TBV e GEBV, respectivamente; cov é a covariância; σ o desvio-padrão.

O viés de predição foi dado como o desvio médio do coeficiente de regressão entre TBV e GEBV $b(TBV, GEBV)$ de todos os indivíduos avaliados, em que coeficientes iguais a 1 indicam que não há viés da predição dos valores genéticos genômicos, entretanto estimativas acima ou abaixo de 1 indicam que as predições dos GEBVs foram subestimadas ou superestimadas, respectivamente. Esse método é semelhante a estudos com populações simuladas, que objetivaram a comparação entre as predições em diferentes procedimentos de análise genômica em animais e plantas (Daetwyler et al., 2013; van den Berg et al., 2019; Gowan et al., 2019). As estimativas e respectivos erros-padrão foram obtidos pela média dos coeficientes de regressão de 20 repetições, em cada cenário genético.

Para comparar a acurácia e viés de predição entre os modelos com dados reais, foram gerados dois conjuntos de dados para produção de leite, concentração de gordura, proteína, caseína e conteúdo de ácido graxo poli-insaturado, sendo 1) conjuntos de dados completo, na qual todos os animais que possuíam genótipos e fenótipos foram considerados como a população de treinamento; e 2) um conjunto de dados reduzido, em que um-terço dos animais genotipados eram de primeira lactação e não possuíam fenótipo, isto é, para esses indivíduos o fenótipo foi zerado (população de validação).

Dessa forma, a habilidade de predição ou acurácia foi estimada pela correlação de *Pearson* (r) entre os valores genéticos genômicos (GEBV) preditos para a população de treinamento e o fenótipo corrigido para efeitos fixos no grupo de validação (Legarra et al., 2008):

$$r = cor(GEBV, y^*)$$

em que r é o valor de correlação entre o valor genético genômico (GEBV) e o fenótipo ajustado para o efeito de fixo de grupo de contemporâneos.

O ajuste do fenótipo foi realizado por meio de um modelo linear simples, com um vetor de média dos valores fenotípicos da população de treinamento e o grupo de contemporâneos como o efeito fixo. Dessa forma, o fenótipo ajustado foi obtido por:

$$y^* = \mu + y - X\hat{\beta} - \varepsilon$$

sendo y^* o fenótipo ajustado/corrigido; μ vetor de média geral para o fenótipo da população de treinamento; y é o vetor de fenótipos da população de treinamento; X é a matriz de incidência para os efeitos fixos; $\hat{\beta}$ é o vetor de efeitos fixos (grupo de contemporâneos); ε é o vetor de resíduos.

O grau de inflação ou deflação dos modelos (viés de predição), foram obtidos de um modelo de regressão linear univariada do fenótipo ajustado, em função do GEBV predito na população de treinamento, para cada característica (Aliloo et al., 2016). A interpretação ocorreu por meio dos coeficientes de regressão estimados, na qual valores iguais ou próximos de 1 indicam que não houve subestimação dos valores preditos (valores > 1) ou superestimação (valores < 1).

3.3. RESULTADOS

3.3.1. Componentes de variância

O objetivo com a simulação foi verificar o impacto em incluir os efeitos de dominância nos modelos de predição genômica tradicionais e qual a contribuição para a estimativa dos componentes de variância. As estimativas de variância estão apresentadas nas tabelas 3 (dados simulados) e 4 (dados reais).

Tabela 3. Componentes de variância para fração genética aditiva e de desvios de dominância, variância residual, fenotípica e herdabilidades estimadas para o modelo genético aditivo (MGA) e aditivo-dominante (MGAD), em três cenários genéticos de simulação (I, $h^2 = 0,10$; II, $h^2 = 0,30$; III, $h^2 = 0,50$).

Modelo	σ_a^2	σ_e^2	σ_p^2	h_a^2	σ_d^2	h_d^2
	Cenário I					
MGA	0,078 (0,034)	0,934 (0,045)	1,012 (0,037)	0,076 (0,033)		
MGAD	0,110 (0,054)	1,512 (0,117)	1,711 (0,064)	0,065 (0,031)	0,089 (0,085)	0,0520 (0,049)
Cenário II						
MGA	0,299 (0,054)	0,697 (0,045)	0,996 (0,040)	0,300 (0,049)		
MGAD	0,357 (0,083)	1,360 (0,111)	1,769 (0,070)	0,201 (0,044)	0,052 (0,068)	0,030 (0,040)
Cenário III						
MGA	0,541 (0,065)	0,490 (0,041)	1,516 (0,045)	0,523 (0,048)		
MGAD	0,621 (0,102)	1,128 (0,123)	1,800 (0,075)	0,345 (0,050)	0,051 (0,066)	0,0290 (0,037)

σ_a^2 variância genética aditiva; σ_e^2 variância residual; σ_p^2 variância fenotípica; h_a^2 proporção da variância genética aditiva (herdabilidade no sentido restrito); σ_d^2 variância genética de dominância; h_d^2 componente da variância de dominância sob a variância genética total.

As estimativas de variância genética foram maiores para os modelos aditivo-dominante e, entre os cenários, as estimas aumentaram na medida que a magnitude de herdabilidade foi maior. No entanto, a proporção da variância genética de dominância em relação à aditiva, foi maior para o cenário I com herdabilidade de 0,10 (8,9%), com variância fenotípica de aproximadamente $1,711 \pm 0,064$. Nos cenários II e III as variâncias fenotípicas foram de $1,769 \pm 0,070$ e $1,800 \pm 0,075$, respectivamente, com variância genética residual menor que no cenário I. As herdabilidades estimadas pelo procedimento GBLUP foram muito próximas do que foi simulado,

sendo que a herdabilidade devido ao efeito de dominância foi maior para cenário I ($0,0520 \pm 0,049$), e menor para o cenário II e III ($0,030 \pm 0,040$ e $0,0290 \pm 0,037$).

Após as análises de simulação, os modelos foram estendidos para uma fase prática. As estimativas de variância aditiva e herdabilidades aditivas para todas as características, foram semelhantes independente se o efeito de dominância foi incluído no modelo (Tabela 4).

Tabela 4. Componentes de variância, proporção da variância devido ao efeito de ambiente permanente, proporção da variância genética aditiva (h_a^2) e do efeito de dominância (h_d^2) sob a variância genética total, herdabilidade e respectivos erros-padrão (entre parênteses), para produção de leite (PL, kg/dia), concentração de proteína (PROT, %), gordura (GORD, %), caseína (CAS, %) e componente de ácido graxo poli-insaturado (POLI, %), estimados no modelo genômico aditivo (MGA) e aditivo-dominante (MGAD).

Fenótipo	σ_a^2	σ_{pe}^2	σ_e^2	σ_p^2	c^2	h_a^2	σ_d^2	h_d^2	h_{a+d}^2
MGA									
PL	6,9162 (2,1371)	21,261 (2,0561)	45,948 (0,5845)	74,125 (1,6407)	0,2868 (0,0266)	0,0933 (0,0280)			
PROT	0,0192 (0,0027)	0,0107 (0,0018)	0,0480 (0,0006)	0,0780 (0,0018)	0,1377 (0,0024)	0,2467 (0,0303)			
GORD	0,0641 (0,0122)	0,0574 (0,0095)	0,4433 (0,0053)	0,5643 (0,0092)	0,1018 (0,0169)	0,1136 (0,0206)			
CAS	0,0157 (0,0027)	0,0106 (0,0019)	0,0383 (0,0005)	0,0646 (0,0017)	0,1636 (0,0316)	0,2435 (0,0383)			
POLI	0,0002 (0,00003)	0,0001 (0,00003)	0,002 (0,00002)	0,002 (0,00003)	0,071 (0,0144)	0,080 (0,0167)			
MGAD									
PL	6,9162 (2,1362)	21,259 (2,0556)	45,948 (0,5845)	74,125 (1,6038)	0,2868 (0,0265)	0,0933 (0,0281)	0,0031 (0,0143)	0,0002 (0,0020)	0,093
PROT	0,0192 (0,0025)	0,0105 (0,0017)	0,0480 (0,0006)	0,0780 (0,0018)	0,1347 (0,0217)	0,2466 (0,0279)	0,0003 (0,0001)	0,0033 (0,0015)	0,250
GORD	0,0641 (0,0122)	0,0521 (0,0255)	0,4427 (0,0053)	0,5646 (0,0932)	0,0923 (0,0448)	0,1135 (0,0207)	0,0057 (0,0254)	0,0102 (0,0446)	0,124
CAS	0,0158 (0,0011)	0,0105 (0,0008)	0,0383 (0,0005)	0,0646 (0,0017)	0,1625 (0,0127)	0,2439 (0,0166)	0,0001 (0,0019)	0,0008 (0,0295)	0,245
POLI	0,0001 (0,00001)	0,0001 (0,00001)	0,0016 (0,00002)	0,0019 (0,00003)	0,041 (0,001)	0,074 (0,0065)	0,001 (0,00001)	0,038 (0,001)	0,112

σ_a^2 variância genética aditiva; σ_{pe}^2 variância de efeito de ambiente permanente; σ_e^2 variância residual; σ_p^2 variância fenotípica; c^2 proporção da variância fenotípica devido o efeito de ambiente permanente; h_a^2 proporção da variância genética aditiva (herdabilidade no sentido restrito); σ_d^2 variância genética de dominância; h_d^2 componente da variância de dominância sob a variância genética total; h_{a+d}^2 herdabilidade genômica obtida pela soma da h_a^2 e h_d^2 .

As estimativas de variância genética aditiva, de efeito de ambiente permanente, variância fenotípica e a proporção da variância fenotípica devido ao efeito de ambiente permanente, foram maiores para a produção de leite, que apresentou baixa magnitude de herdabilidade genômica aditiva ($0,0933 \pm 0,0280$), comparada às demais características. Maiores estimativas de

herdabilidade aditiva foram encontradas para a concentração de proteína ($0,2467 \pm 0,0303$), caseína ($0,2435 \pm 0,0383$) e gordura ($0,1136 \pm 0,0206$). O menor valor de herdabilidade foi encontrado para o conteúdo de ácido graxo poli-insaturado, sendo de $0,080 \pm 0,0167$.

Para todas as características, a inclusão dos efeitos de dominância no modelo genômico, pouco contribuiu nas estimativas dos componentes de variância (Tabela 4). A maior estimativa foi para a concentração de gordura, na qual a variância do componente genético de dominância foi de, aproximadamente, $0,0057 \pm 0,0254$, seguido da produção de leite ($0,0031 \pm 0,0143$), do conteúdo de ácido graxo poli-insaturado ($0,001 \pm 0,00001$), da concentração de proteína ($0,0003 \pm 0,0001$) e caseína ($0,0001 \pm 0,0019$).

Embora o componente genético não-aditivo pouco tenha sido captado pelo modelo, para algumas características a variância fenotípica explicada pelo efeito de dominância foi considerável. Para o ácido graxo poli-insaturado, por exemplo, a proporção do efeito de dominância sobre a variância fenotípica foi de 3,8%, com estimativa de herdabilidade genômica total 1,4% maior no modelo aditivo-dominante ($h_{total}^2 = 0,112$), do que no modelo aditivo ($h_{total}^2 = 0,080$). Além disso, não foi encontrado relato científicos de estimativas de variância para essa característica, avaliada com modelos aditivo-dominante e, portanto, esse pode ser o primeiro estudo a dissecar os efeitos genéticos aditivo e não-aditivo para o conteúdo de ácido graxo poli-insaturado em bovinos da raça Holandesa.

Os efeitos de dominância estimados para a gordura no leite, também foram substanciais para a proporção da variância fenotípica, na qual 1,02% da variabilidade fenotípica foi devido aos efeitos de dominância. Para produção de leite, proteína e caseína as proporções foram menores, sendo de 0,02%, 0,33% e 0,08%, respectivamente. As estimativas de herdabilidade obtidas pela soma da proporção da variância aditiva e de dominância, foram levemente maiores para proteína ($h_{a+d}^2 = 0,250$), gordura ($h_{a+d}^2 = 0,124$) e caseína ($h_{a+d}^2 = 0,245$), quando considerado os efeitos de dominância no modelo. Para a produção de leite, não houve diferença entre as magnitudes de herdabilidade aditiva e total.

3.3.2. Acurácia e viés de predição

Os coeficientes de regressão entre TBV (*True Breeding Value*) e GEBV para os dados simulados (Tabela 5), revelaram que para características de média a alta herdabilidade há menor viés na predição dos valores genéticos devido às estimativas serem iguais ou próximas de 1 ($b = 1,02$ e $0,98$), comparado a herdabilidade de 0,10, na qual houve subestimação das predições ($b = 1,34$). No entanto, para o cenário de baixa herdabilidade, o coeficiente de regressão foi de 0,46, maior do que encontrado nos demais cenários para o modelo aditivo-dominante. Isso indica que

para características complexas (herdabilidade baixa), incluir os efeitos de dominância no modelo podem reduzir o viés das predições.

Tabela 5. Coeficientes de regressão $b(TBV, GEBV)$ e erro-padrão (entre parênteses) para os modelos genômicos aditivo (MGA) e aditivo-dominante (MGAD) em cada cenário genético simulado (magnitudes de herdabilidade, h^2).

Modelo	Cenário genético (h^2)		
	0,10	0,30	0,50
MGA	1,34 (0,066)	1,02 (0,028)	0,98 (0,019)
MGAD	0,46 (0,064)	0,39 (0,034)	0,41 (0,027)

A acurácia de predição foi calculada por meio de correlação entre valor genético verdadeira (TBV) e valor genético genômico predito (GEBV) para os dados simulados (Tabela 6). Verificou-se que o modelo puramente aditivo teve maior acurácia na predição dos valores genéticos para o cenário na qual a característica simulada possui magnitude de herdabilidade moderada a alta (correlação de 0,68 e 0,79). Isso também ocorreu quando incluído os efeitos de dominância no modelo, no entanto, com menor acurácia (0,29 e 0,36).

Tabela 6. Habilidade de predição para os modelos genômicos aditivo (MGA) e aditivo-dominante (MGAD) em cada cenário genético simulado (magnitudes de herdabilidade, h^2).

Modelo	Cenário genético (h^2)		
	0,10	0,30	0,50
MGA	0,47	0,68	0,79
MGAD	0,17	0,29	0,36

Para o modelo aditivo-dominante os valores encontrados de acurácia e viés de predição foram baixos, isso implica na dificuldade e complexidade em ajustar os modelos para os efeitos não-aditivos. Contudo, os desvios das estimativas de regressão mostraram que para características de menor complexidade genética (herdabilidade moderada a alta), os modelos podem ser melhor ajustados, como observado no menor nos cenários II e III, enquanto incluir o efeito de dominância pode ser vantajoso para características de baixa magnitude de herdabilidade (cenário I).

Em dados reais, os resultados seguiram otimistas quanto à inclusão dos efeitos genéticos não-aditivos nos modelos genômicos. Os coeficientes de regressão do fenótipo ajustado, em valores genéticos preditos obtidos com o modelo M1 e M2 foram semelhantes ou muito próximos entre eles, mas diferentes entre as características (Tabela 7). Nenhuma das estimativas chegou a ser igual ou próximo de 1. Para a produção de leite, foi verificada uma acentuada subestimação das predições para os dois modelos testados, na qual os coeficientes de regressão foram de, aproximadamente, $1,194 \pm 0,2130$.

No entanto, para concentração de gordura os coeficientes foram de $0,806 \pm 0,1475$ para o modelo aditivo e de $0,801 \pm 0,1477$ para o modelo aditivo-dominante, sendo essas estimativas as mais próximas do desejado. Em seguida, para o conteúdo de ácido graxo poli-insaturado as estimativas foram de $0,748 \pm 0,1728$ sob o modelo aditivo e de $0,781 \pm 0,1823$ no modelo aditivo-dominante; e para a caseína o coeficiente de regressão estimado foi $0,641 \pm 0,0885$ em ambos os modelos. As estimativas mais baixas, aquelas que indicaram inflação ou superestimação das predições foi para a concentração de proteína no leite, com valores próximos de $0,370 \pm 0,0823$ e $0,341 \pm 0,0823$ para o modelo aditivo e aditivo-dominante.

Tabela 7. Coeficientes de regressão do fenótipo ajustado em função do GEBV (by*, GEBV) e erro-padrão (entre parênteses) para produção de leite (PL), concentração de gordura (GORD), proteína (PROT), caseína (CAS) e conteúdo de ácido graxo poli-insaturado (POLI), para modelo aditivo (MGA) e aditivo-dominante (MGAD).

Modelo	PL (kg/dia)	GORD (%)	PROT (%)	CAS (%)	POLI (%)
MGA	1,194 (0,2130)	0,806 (0,1475)	0,370 (0,0823)	0,641 (0,0885)	0,748 (0,1728)
MGAD	1,194 (0,2130)	0,801 (0,1477)	0,341 (0,0823)	0,641 (0,0884)	0,781 (0,1823)

As correlações entre o fenótipo ajustado e o valor genético genômico predito serviram para avaliar a capacidade ou habilidade de predição de cada modelo, para cada característica (Tabela 8). Para a maioria dos fenótipos, as estimativas de acurácia não indicaram nenhuma vantagem na melhora da capacidade preditiva com a inclusão dos efeitos de dominância. Para ácido graxo poli-insaturado, uma rápida vantagem em acurácia foi encontrada para o modelo genômico aditivo (0,224). O maior valor de acurácia foi encontrado para caseína, com estimativa próxima de 0,387 para ambos os modelos, seguido de 0,284 para produção de leite e 0,278 para gordura. Para a concentração de proteína no leite as estimativas foram menores comparada às demais características, sendo de 0,214 em ambos os modelos.

Tabela 8. Habilidade de predição (acurácia) para produção de leite (PL), concentração de gordura (GORD), proteína (PROT), caseína (CAS) e conteúdo de ácido graxo poli-insaturado (POLI), para modelo aditivo (MGA) e aditivo-dominante (MGAD).

Modelo	PL (kg/dia)	GORD (%)	PROT (%)	CAS (%)	POLI (%)
MGA	0,284	0,278	0,214	0,387	0,224
MGAD	0,284	0,278	0,214	0,387	0,222

3.4. DISCUSSÃO

3.4.1. Componentes de variância

Foi idealizado um estudo de simulação estocástica para uma aplicação mais prática e teórica a respeito da capacidade dos diferentes modelos em estimar os componentes de variância e, conseqüentemente, a predição dos valores genéticos. Era esperado que os componentes de variância estimados fossem maiores nos cenários de média e alta magnitude de herdabilidade, independente do modelo ajustado, por dois principais fatores: para atingir a herdabilidade esperada, os efeitos genéticos eram dimensionados em função da sua distribuição; além disso, em características quantitativas é esperado que a fração genética aditiva seja maior para maiores magnitudes de herdabilidade.

Em um estudo de simulação, Su et al. (2012) encontraram variância de 5,9% para efeito de dominância e herdabilidade total acima de 0,30. Vitezica et al. (2013) encontraram proporções maiores ainda, de aproximadamente 13,2%. Contudo, em Su et al. (2012) os efeitos aditivos não são tratados adequadamente pois o seu modelo não inclui parte da ação biológica de dominância. Portanto, o modelo parametrizado de Vitezica et al. (2013) é interpretável a nível real, pois o ajuste dos efeitos de dominância é dado de forma parametrizada, o que permite entender que o modelo aditivo é equivalente ao modelo aditivo-dominante.

As estimativas referentes a fração genética não-aditiva foram maiores nos cenários de menor valor de herdabilidade. Neste caso, a proporção dos efeitos de dominância sobre a variância fenotípica foi de 8,9% para o cenário I ($h^2 = 0,10$), sendo maior do que a proporção estimada para o cenário II (5,2%, $h^2 = 0,30$) e III (5,1%, $h^2 = 0,50$). Isso indica que, para características complexas, com ação genética regida por muitos genes e, principalmente, de baixa magnitude de herdabilidade, a proporção da variância fenotípica devido aos efeitos de dominância é evidenciada em maior magnitude, comparada às características de média a alta herdabilidade. Esses resultados são consistentes com outros estudos de simulação, que encontraram valores entre 4 e 7% da proporção da variância de dominância sobre a variância fenotípica (Sun et al., 2014; Wittenburg et al., 2015; Jiang et al., 2017).

Na fase seguinte, cinco características de produção e qualidade do leite foram ajustadas por meio dos modelos que consideraram efeitos genéticos não-aditivos. As estimativas de variância e herdabilidade genômica aditiva estão de acordo com Petrini et al. (2016) e Iung et al. (2019), que utilizaram o mesmo banco de dados deste estudo. Para produção de leite, foi relatado coeficientes de herdabilidade entre 0,12 e 0,13, e de 0,35 e 0,19 para proteína e gordura, respectivamente. Essas estimativas foram superiores ao relatado nesse estudo e pode estar

associada às diferenças entre os conjuntos de dados utilizados pelos autores e neste estudo, no qual resultou em estimativas de variância residual maiores.

A caseína é a variante proteica mais abundante no leite, constituindo aproximadamente 80% da porcentagem total da proteína (Gustavsson et al., 2014). Além disso, é protagonista nos processos de coagulação e rendimento na produção de derivados lácteos. Foi relatado nesse estudo, herdabilidade próxima de 0,24, sendo consistente com outros autores, que relataram estimativas de herdabilidade aditiva entre 0,09 e 0,80, obtidas em populações de gado leiteiro de diferentes raças (Schopen et al., 2009; Bonfatti et al., 2011; Samoré et al., 2012; Sanchez et al., 2017). Assim como para a concentração de gordura e proteína, os autores apontam o fator racial e o sistema de criação dos animais como principal influenciador no conteúdo relativo da caseína no leite. Dessa forma, herdabilidade de média a alta magnitude são encontradas em indivíduos de médio a alto potencial produtivo de leite e sólidos, como as raças Holandês e Jersey.

O conteúdo de ácido graxo poli-insaturado apresentou herdabilidade de baixa magnitude para ambos os modelos, um pouco menores do que encontrado na literatura, que variam entre 0,11 e 0,176 (Lung et al., 2019; Penasa et al., 2015), no entanto, foi semelhante ao relatado por Petrini et al. (2016), com magnitude próxima de 0,11. Embora as diferenças sejam pequenas, a divergência das magnitudes pode ser devido a diferença nas estruturas populacionais, como o número de indivíduos, quantidade informações utilizadas, formação ou não de grupos de contemporâneos ou por fatores ambientais, como a temperatura do ambiente e a nutrição. Além disso, o conteúdo de ácido graxo não é sintetizado pelos ruminantes e, conseqüentemente, a sua concentração no leite é ligada diretamente com as quantidades absorvidas pelo intestino, sendo altamente influenciado por fatores ambientais e externos, como a nutrição (Chilliard et al., 2004). Portanto, é aceitável que menores coeficientes de herdabilidade sejam estimados (Bastin et al., 2011; Narayana et al., 2017).

Quando incluído os efeitos de dominância, as estimas de variância aditiva praticamente não mudaram ou pouco variaram. Essas estimativas são inferiores às relatadas por Jiang et al. (2017), que encontraram variâncias de dominância de 2,5% a 4,0% para produção de leite, gordura e proteína. O mesmo para a proporção da variância de dominância sob a variância genética total (h_d^2), com estimativas relatadas entre 7,3% e 12,6% (Jiang et al., 2017). Em outro estudo, foram relatadas proporções entre 3,2% e 3,8% para essas mesmas características em vacas de raça Holandesa (Aliloo et al., 2016).

Para a caseína, Winttenberg et al. (2015) encontraram estimativas de variância aditiva e de dominância de 1,44 e 0,80, respectivamente, muito superiores ao relatado em nosso estudo, que foram de 0,0158 e 0,0001. Esses autores também relatam estimativas maiores para produção de

leite e conteúdo de proteína e gordura, variando de 1,83 a 5,58 para variância genética aditiva e de 1,16 a 1,95 para dominância, porém não encontraram estimativa a variância de dominância para a produção de leite, que aqui foi de 0,0031.

Embora haja poucos relatos referente aos modelos aditivo-dominância para vacas de raça leiteira, a expectativa é que a proporção do efeito de dominância sob a variância genética total seja de até 10%, como visto na maioria dos trabalhos publicados. As estimativas obtidas nesse estudo com a inclusão da dominância, foram menores do que os estudos em dados reais. Isso ainda não é claro, pois recentemente os efeitos não-aditivos foram incluídos em avaliações genômicas.

Na prática, maiores e melhores estimativas de variância podem ser obtidas se os componentes incluídos nos modelos genômicos forem suficientes para explicar o fenótipo avaliado. Para estudos com efeitos de dominância é fundamental que a população avaliada seja suficientemente grande e informativa quanto ao parentesco entre os animais (Varona et al., 2018). A estimativa de variância de dominância requer grandes quantidades de dados com alta proporção de irmãos-completos, sendo necessário até 20 vezes mais informações que nas avaliações apenas com efeito aditivo (Wellmann e Bennewitz, 2011; Misztal et al., 1997). Além disso, em populações altamente endogâmicas, com taxa de homozigose elevada, a capacidade em captar os efeitos decorrentes da interação entre os genes é reduzida devido à menor proporção de trechos genômicos em heterozigose (Varona et al., 1998).

Em estudo com várias características de produção de leite e de conformação, em animais da raça Simental, Ertl et al. (2014) encontraram proporções de variâncias de dominância de até 50% para produção de leite, gordura e proteína. Os autores sustentam o fato de que a diversidade genética para a raça estudada é maior comparada à raça Holandesa, assim como o tamanho efetivo e, conseqüentemente, com maior nível de heterozigosidade.

Não foi avaliado o coeficiente de endogamia, mas em uma análise exploratória completa foi verificado que a correlação média entre os coeficientes fora da diagonal na matriz de relacionamento genômico (\mathbf{G}) foi de 0,02, isso indica que a profundidade informativa dos relacionamentos é baixa e, portanto, houve poucos irmãos-completos avaliados. Entre os elementos na diagonal da matriz a correlação foi elevada (0,56), sendo um indicativo da relação entre pais e filhos elevada e, possivelmente, com alta taxa de endogamia na população. Essas podem ser as principais causas para baixas estimativas de variância de dominância obtidas pelo modelo genômico que considerou os efeitos de dominância.

3.4.2. Acurácia e viés de predição

O uso dos efeitos aditivos mais de dominância não levou a predições mais acuradas, independente das abordagens utilizadas nesse estudo. As correlações entre os valores genéticos verdadeiros (TBV) e genômico (GEBV) e entre o fenótipo ajustado (y^*) e o GEBV foram maiores para o modelo aditivo na avaliação com dados simulados e reais. Entretanto, com os dados simulados foi observado o aumento nos valores de acurácia na medida que a magnitude de herdabilidade aumentou, sendo esse comportamento consistente com os resultados de Gowane et al. (2018), com valores de acurácia de até 0,70, próximo do que foi encontrado neste estudo para o modelo genômico aditivo (0,47 a 0,79).

Há relatos que indicam ganhos em acurácia de predição de até 17% quando incluído os efeitos de dominância nos modelos com dados simulados (Wellmann e Bennewitz, 2012; Da et al., 2014) e suaves vantagens em dados reais (Sun et al., 2014; Aliloo et al., 2016). Esses autores tiveram acesso a um maior número de animais genotipados em relação a esse estudo. Por outro lado, Jiang et al. (2017) não encontraram grandes melhorias ao incluir efeitos de dominância nas predições genômicas, mesmo com um banco de genótipos composto por mais de 200.000 mil vacas da raça Holandesa genotipadas e com informações completas de pedigree. Outro ponto importante na comparação entre os modelos, é a condição genética da raça avaliada. Liu et al. (2019) estudaram os efeitos de dominância em gado de corte e obtiveram estimativas de variância de dominância de até 15,8% em relação à variância fenotípica e um aumento de 0,9% na acurácia de predição dos GEBVs via GBLUP. Em outro trabalho com bovinos leiteiros, foi relatada a redução do viés de predição e o ganho em acurácia de até 0,09 pontos para as predições genômicas em raça pura, quando uma fração da população de referência era de animais que eram cruzados (Khansefid et al., 2020).

Fica muito claro que a quantidade de informações dos bancos de dados não será determinística para a capacidade preditiva de modelos genômicos complexos, mas sim a qualidade das informações e as propriedades de ação gênica de cada característica. Por exemplo, a baixa herdabilidade estimado para o efeito não-aditivo e a falta de pares de irmãos-completos entre as populações de treinamento e validação, são as possíveis e mais prováveis causas de não haver ganhos em capacidade preditiva, quando somado os efeitos aditivos e de dominância nos modelos de predição (Ertl et al., 2014; Jiang et al., 2017). Nossos resultados concordam com esses relatos, principalmente pelo fato de que a grande maioria dos registros fenotípicos consistiram em animais meios-irmãos, mas foram contrários ao relatado por Aliloo et al (2016) e Ertl et al. (2014), que encontram melhor ajuste dos modelos quando considerado os efeitos de dominância.

Embora não tenha sido relatado ganhos positivos na acurácia de predição pelo modelo aditivo-dominante, os coeficientes de regressão revelaram uma vantagem na predição para ácido graxo poli-insaturado, quando incluído os efeitos de dominância no modelo. Sob abordagem genômica não-aditiva, as estimativas para essa característica foram menos enviesadas e indica que para esse fenótipo, que é fortemente influenciado pela nutrição, considerar os efeitos de dominância pode ser benéfico para a estimativa dos componentes de variância. esse é o primeiro estudo que decompôs os efeitos genéticos em aditivo e de dominância para o conteúdo de ácido graxo poli-insaturado.

3.5. CONCLUSÃO

A comparação entre modelo aditivo e não-aditivo sugere que, para o conteúdo de ácido graxo poli-insaturado, os efeitos de dominância devem ser considerados nas avaliações genômicas de bovinos da raça Holandesa, para a melhor predição dos valores genéticos genômicos. A variância de dominância contribuiu de 2,9 a 5,2% para a variância fenotípica em dados simulados e para as características de produção de leite, as contribuições foram entre 0,08 e 3,8%. De modo geral, as estimativas de herdabilidade pouco variaram quando somado os efeitos de dominância e aditivo, exceto para o conteúdo de ácido graxo. Este é o primeiro estudo que decompôs os efeitos genéticos em aditivo e não-aditivo para o ácido graxo poli-insaturado.

Não houve grandes avanços e contribuição dos efeitos de dominância na habilidade de predição. As predições dos valores genéticos foram menos enviesados no modelo puramente aditivo no estudo de simulação e a acurácia foi maior para os cenários com maior magnitude de herdabilidade, independente do modelo. Para as características leiteiras, o menor viés de predição foi obtido para gordura, ácido graxo e caseína, nos dois modelos genômicos. Não foi identificado ganhos positivos em acurácia de predição dos valores genéticos entre os modelos, contudo, houve menor viés de predição para o conteúdo de ácido graxo poli-insaturado quando considerado os efeitos de dominância, ao invés de utilizar apenas os efeitos genéticos aditivos.

Fica claro a necessidade de adequação dos modelos predição genômica para efeitos não-aditivos em populações de animais da mesma raça, devido a fatores como relacionamento entre os indivíduos e taxa de endogamia, apontados na discussão deste estudo.

REFERÊNCIAS

- Abdollahi-Arpanahi, R.; Gianola, D.; Peñagaricano, F. Deep learning versus parametric and ensemble methods for genomic prediction of complex phenotypes. **Genetics Selection Evolution**, v.52, n.12, 2020.
- Aguilar, I, Misztal, I.; Jhonson, D. L.; Legarra, A.; Tsuruta, S.; Lawlor, T. J. Hot topic: a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. **Journal of Dairy Science**, v.93, n. 2, p.743-752, 2010.
- Aguilar, I. 2014. Short introduction to BLUPF90 family programs. Retrieved on 20 March 2021, from http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=slides_2014_lva.pdf.
- Akanno, E. C.; Abo-Ismael, M. K.; Chen, L.; Crowley, J. J.; Wang, Z.; Li, C.; Basarab, J. A.; MacNeil, M. D.; Plastow, G. S. Modelling heterotic effects in beef cattle using genome-wide SNP-marker genotypes. **Journal Animal Science**, v.96, n.3, p.830-845, 2018.
- Alillo, H.; Pryce, J. E.; González-Recio, O.; Cocks, B. G.; Hayes, B. J. Accounting for dominance to improve genomic evaluations of dairy cows for fertility and milk productions traits. *Genetics Selection Evolution*, v.48, n.8, p.1-11, 2016.
- Amadeu, R. R.; Cellon, C.; Lara, L.; Resende, M.; Oliveira, I.; Ferrao, L.; Muñoz, P.; Garcia, A. AGHmatrix: Relationship Matrices for Diploid and Autopolyploid Species. Retrieved on 15 February 2021, from <https://github.com/prmunoz/AGHmatrix>.
- Amuzu-Aweh, E. N.; Bijama, P.; Kinghorn, P. B.; Vereijken, A.; Visscher, J.; Am van Arendonk, J.; Bovenhuis, H. Prediction of heterosis using genome-wide SNP-marker data: application to egg production traits in White Leghorn crosses. **Heredity**, v.11, p.530-538, 2013.
- Bastin, C.; Gengler, N.; Soyeurt, H. Phenotypic and genetic variability of production traits and milk fatty acid contents across days in milk for Walloon Holstein first-parity cows. **Journal of Dairy Science**, v.94, n.8, p.4152-4163, 2011.
- Bonfatti, V.; Cecchinato, A.; Gallo, L.; Blasco, A.; Carnier, P. Genetic analysis of detailed milk protein composition and coagulation properties in Simmental cattle. **Journal of Dairy Science**, v.94, n.10, p.5183-5193, 2011.
- Bradford, H. L.; Masuda, Y.; VanRaden, P. M.; Legarra, A.; Misztal, I. Modeling missing pedigree in single-step genome BLUP. **Journal of Dairy Science**, v.102, p.2336-2346, 2019.
- Cao, Z.; Guo, Y.; Yang, Q.; He, Y.; Fetouh, M. I.; Warner, R. M.; Deng, Z. Genome-wide identification of quantitative trait loci for important plant and flower traits in petunia using a high-density linkage map and an interspecific recombinant inbred population derived from *Petunia integrifolia* and *P. axillaris*. **Horticulture Research**, v.6, n.27, 2019.
- Chilliard, Y.; Ferlay, A.; Mansbridge, R. M.; Doreau, M. Ruminant milk fat plasticity: nutritional control of saturated, polyunsaturated, *trans* and conjugated fatty acids. **Annales de zootechnie**, v.49, n.3, 2000.
- Christensen, O. F.; Lund, M. Genomic prediction when some animals are not genotyped. **Genetics Selection Evolution**, v.42, n.1, 2010.

- Compton, C. W. R.; Heuer, C.; Thomsen, P. T.; Carpenter, T. E.; Phyn, C. V. C.; McDougall. Invited review: A systematic literature review and meta-analysis of mortality and culling in dairy cattle. **Journal of Dairy Science**, v.100, p.1-16, 2017.
- Da, Y.; Wang, C.; Wang, S.; Hu, G. Mixed model methods for genomic prediction and variance component estimation of additive and dominance effects using SNP markers. **PLoS ONE**, v.9, n.1, 2014.
- Daetwyler, H. D.; Calus, M. P. L.; Pong-Wong, R.; de los Campos, G.; Hickey, J. M. 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.; Hickey, J. M.; Pong-Wong, R.; Daetwyler, H. D.; Claus, M. P. Whole-genome regression and prediction methods applied to plant and animal breeding. **Genetics**, v.193, n.2, p.327-345, 2013.
- Ertl, J.; Legarra, A.; Vitezica, Z. G.; Varona, L.; Edel, C.; Emmerling, R.; Gotz, K. Genomic analysis of dominance effects on milk production and conformation traits in Fleckvieh cattle. **Genetics Selection Evolution**, v.46, n.40, 2014.
- Estimating dominance genetic variances for growth traits in American Angus males using genomic models. **Journal of Animal Science**, v.98, n.1, p.1-7, 2020.
- Falconer, D. S. and Mackay, T. F. C. (1996) **Introduction to quantitative genetics**. 4th ed.
- Garcia-Baccino, C. A.; Lourenco, D. A. L.; Miller, S.; Cantet, R. J. C.; Vitezica, Z. G. Gengler, N. Van Vleck, L. D.; MacNeil, M. D.; Misztal, E. U.; Pariacote, F. A. Influence of dominance relationships on the estimation of dominance variance with sire-dam subclasses effects. **Journal of Animal Science**, v.75, n.11, p.2885-2891, 1997.
- Gowane, G. R.; Lee, S. H.; Clark, S.; Moghaddar, N.; Al-Mamum, H. A.; van der Werf, J. H. J. Effect of selection on bias and accuracy in genomic prediction of breeding values. **Journal of Animal Breeding and Genetics**, v.136, n.5, p.390-407, 2019.
- Gustavsson, F.; Buitenhuis, A. J.; Johansson, M.; Bertelsen, H. P.; Glantz, M.; Poulsen, N. A.; Mansson, H. L.; Stalhammar, H.; Larsen, L. B.; Bendixen, C.; Paulsson, M.; André, A. Effects of breed and casein genetic variants on protein profile in milk from Swedish Red, Danish Holstein, and Danish Jersey cows. **Journal of Dairy Science**, v.97, n.6, p.3866-3877, 2014.
- Hayes, B. J.; Bowman, P. J.; Chamberlain, A. C.; Verbyla, K.; Goddard, M. E. Accuracy of genomic breeding values in multi-breed dairy cattle populations. **Genetics Selection Evolution**, v.41, n.51, 2009.
- Iversen, M. W.; Nordbø, Ø.; Gjerlaug-Enger, E.; Grindflek, E.; Lopes, M. S.; Meuwissen, T. Effects of heterozygosity on performance of purebred and crossbred pigs. **Genetics Selection Evolution**, v.51, n.8, 2019.
- Jiang, J.; Shen, B.; O'Connell, J. R.; VanRaden, P. M.; Cole, J. B.; Ma, L. Dissection of additive, dominance, and imprinting effects for production and reproduction traits in Holsten cattle. **BMC Genomics**, v.18, n.425, 2017.

- Khansefid, M.; Goddard, M. E.; Haile-Mariam, M.; Konstantinov, K. V.; Jong, G.; Jewell, E. G.; O'Connor, E.; Pryce, J. E.; Daetwyler, H. D.; MacLeod, I. M. Improving genomic prediction of crossbred and purebred dairy cattle. **Frontiers in Genetics**, v.11, 2020.
- 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.
- Liu, N.; Liu, J.; Li, W.; Pan, Q.; Liu, J.; Yang, X.; Yan, J.; Xiao, Y. Intraspecific variation of residual heterozygosity and its utility for quantitative genetic studies in maize. **BMC Plant Biology**, v.18, n.66, 2018.
- Liu, Y.; Xu, >; Wang, Z.; Xu, L.; Chen, Y.; Zhang, L.; Xu, L.; Gao, X.; Zhu, B.; Li, J. Genomic prediction and association analysis with models including dominance effects for important traits in Chinese Simmental beef cattle. **Animals**, v.9, n.12, 2019.
- Masuda, Y. **Introduction to BLUPF90 suit programs Standard edition**. University of Georgia, 2019. Retrieved from http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=tutorial_blupf90.pdf.
- Misztal, I. Estimation of variance components with large-scale dominance models. **Journal of Dairy Science**, v.80, n.5, p.965-974, 1997.
- 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.
- Misztal, I.; Tsuruta, S.; Lourenco, D.; Masuda, Y. **Manual for BLUPF90 family of programs**. University of Georgia, Athens, USA, 2018. Retrieved from http://nce.ads.uga.edu/wiki/doku.php#blupf90_family_of_programs
- Misztal, I.; Varona, L.; Culbertson, M.; Bertrand, L. K.; Mabry, J.; Lawlor, T. J.; Van Tassel, C.; P.; Gengler, N. Studies on the value of incorporating the effect of dominance in genetic evaluation of dairy cattle, beef cattle and swine. **Biotechnology Agronomy Society and Environment**, v.2, n.4, p.227-233, 1998.
- Moghaddar, N. & Van Der Werf, J. H. J. Genomic estimation of additive and dominance effects and impact of accounting for dominance on accuracy of genomic evaluation in sheep populations. **Journal Animal Breeding and Genetics**, v.134, p.453-462, 2017.
- Mrode, R. A. & Thompson, R. **Linear models for the prediction of animal breeding values**. 2th, CABI Publishing, 2005.
- Mrode, R. A.; Thompson, R. **Linear models for the prediction of animal breeding values**. 3rd edition, 2014, p.360.
- Narayan, S. G.; Schenkel, F. S.; Fleming, A.; Koeck, A.; Malchiodi, F.; Jamrozik, J.; Johnston, J.; Sargolzaei, M.; Miglior, F. Genetic analysis of groups of mid-infrared predicted fatty acids in milk. **Journal of Dairy Science**, v.100, n.6, p.4731-4744, 2017.
- NicolinI, P.; Amorín, R.; Han, Y.; Peñagarciano, F. Whole-genome scan reveals significant non-additive effects for sire conception rate in Holstein cattle. **BMC Genetics**, v.19, n.14, 2018.

- Norris, D.; Varona, L.; Ngambi, J. W.; Visser, D. P.; Mabjiorgu, C. A.; Voodewind, S. F. Estimation of the additive and dominance variances in SA Duroc pigs. **Livestock Science**, v.131, n.1, p.144-147, 2010.
- Peltier, E.; Friedrich, A.; Schacherer, J.; Marullo, P. Quantitative trait nucleotides impacting the technological performances of industrial *Saccharomyces cerevisiae* strains. **Hypothesis and Theory**, v.10, 2019.
- Perez, B. C.; Balieiro, J. C. C.; Carvalheiro, R.; Tirelo, F.; Junior, G. A. O.; Dementshuk, J. M.; Eler, J. P.; Ferraz, J. B. S.; Ventura, R. V. Accounting for population structure in selective cow genotyping strategies. **Journal Animal Breeding and Genetics**, v.136, p.23-39, 2019.
- Pliescheke, L.; Edel, C.; Pimentel, E. C. G.; Emmerling, R.; Bennewitz, J.; Gotz, K. U. Systematic genotyping of groups of cows to improve genomic estimated breeding values of selection candidates. **Genetic Selection Evolution**, v.48, n.73, 2016.
- R Corte Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrived from <https://www.R-project.org/>.
- Samoré, A. B.; Canavesti, F.; Rossoni, A.; Bagnato, A. Genetics of casein content in Brown Swiss and Italian Holstein dairy cattle breeds. **Italian journal of Animal Science**, v.11, n.2, 2011.
- Sanchez, M. P.; Ferrand, M.; Gelé, M.; Pourchet, D.; Mirand, G.; Martin, P.; Brochard, M.; Boichard, D. Short communication: Genetic parameters for milk protein composition predicted using mid-infrared spectroscopy in the French Montebéliard, Normade, and Holstein dairy cattle breeds. **Journal of Dairy Science**, v.100, n.8, p.6371-6375, 2017.
- Sargolzael, M.; Schenckel, F. S.; Jansen, G. B.; Schaeffer, L. R. Extent of Linkage Disequilibrium in Holstein Cattle in North America. **Journal of Dairy Science**, v.91, p.2106-2117, 2008.
- Schopen, G. C. B.; Heck, J. M. L.; Bovenhuis, H.; Visker, M. H. P. W.; van Valenberg, H. J. F.; van Arendonk, J. A. M. Genetic parameters for major milk proteins in Dutch Holstein-Friesians. **Journal of Dairy Science**, v.92, n.3, p.1182-1191, 2009.
- Shaeffer, L. R. **Linear models**. Cornell University, 2019, p.381.
- Su, G.; Christensen, O. F.; Ostersen, T.; Henryon, M.; Lund, M. S. Estimating additive and non-additive genetic variances and predicting genetic merits using genome-wide dense single nucleotide polymorphism markers. **PLoS ONE**, v.7, n.9, 2012.
- Sun, C.; VanRaden, P. M.; Cole, J. B.; O'Connell, J. R. Improvement of prediction ability for genomic selection of Dairy cattle by including dominance effects. **PLoS ONE**, v.9, n.8, 2014.
- Tambasco-Talhari, D.; Alencar, M. M. de; Paz, C. C. P. de; Cruz, G. M. da; Rodriguez, A. de A.; Packer, I.; U.; Coutinho, L. L.; Regitano, L. C. de A. Molecular marker heterozigosity and genetic distances as correlates of production traits in F₁ bovine crosses. **Genetics and Molecular Biology**, v.28, n.2, 2005.
- Tsairidou, S.; Allen, A. R.; Pong-Wong, R.; McBride, S. H.; Wright, D. M.; Matika, O.; Pooley, C. M.; McDowell, S. W. J.; Glass E. J.; Skuce, R. A.; Bishop, S. C.; Woolliams, J. A. An analysis of effects of heterozygosity in dairy cattle for bovine tuberculosis resistance. **Animal Genetics**, v.49, p.103-109, 2018.

- van den Berg, I.; Meuwissen, T. H. E.; MacLeod, I. M.; Goddard, M. E. Predicting the effect of reference population on the accuracy of within, across, and multibreed genomic prediction. **Journal of Dairy Science**, v.102, n.4, p.3155-3174, 2019.
- VanRaden, P. M. Efficient Methods to Compute Genomic Predictions. **Journal of Dairy Science**, v.91, p.4414-4423, 2008.
- Varona, L.; Legarra, A.; Toro, M. A. Non-additive effects in genomic selection. **Frontiers in Genetics**, v.9, n.78, 2018
- Varona, L.; Misztal, I.; Bertrand, J. K.; Lawlor, T. J. Effect of full sibs on additive breeding values under the dominance model for stature in United States Holstein. **Journal of Dairy Science**, v.81, p.1126-1135, 1998.
- Vitezica, Z.; Varona, L.; Legarra, A. On the additive and dominant variance and covariance of individuals within the genomic selection scope. **Genetics**, v.195, n.4, p.1223-1230, 2013.
- Wellman, R. & Bennewitz, J. Bayesian models with dominance effects for genomic evaluation of quantitative traits. **Genetics Research**, v.94, n.1, p.21-37, 2012.
- Wellmann, R.; Bennewitz, J. The contribution of dominance to the understanding of quantitative genetic variation. **Genetics Research**, v.93, n.2, p.139-154, 2011.
- Xiang, T, Christensen, O. F.; Vitezica, Z.; Legarra, A. Genomic model with correlation between additive and dominance effects. **Genetics**, v.209, p.711-723, 2018.
- Xiang, T.; Christensen, O. F.; Vitezica, Z. G.; Legarra, A. Genomic evaluation by including dominance effects and inbreeding depression for purebred and crossbred performance with an application in pigs. **Genetics Selection Evolution**, v.48, n.92, 2016.

4. ESTIMATIVA DA HETEROZIGOSIDADE MOLECULAR COMO RECURSO PARA GERAR HETEROSE GENÔMICA EM POPULAÇÕES DE RAÇA PURA

RESUMO

A heteroziguidade e a heterose genômica, estão assumindo importância nas avaliações de diversas espécies, devido às facilidades que a genômica trouxe para a avaliação quantitativa de características geneticamente complexas. No entanto, a heterose estimada com base nas informações do genoma pouco tem sido explorada em populações de raças ou linhagens puras. Dessa forma, o objetivo desse estudo foi investigar a possibilidade de utilizar a heteroziguidade molecular como fonte de informação para a seleção de indivíduos com potencial para gerar maior e melhor heterose genômica. Foram simulados três cenários genéticos, com diferentes magnitudes de herdabilidade. Além disso, foram simulados os efeitos aditivos e não-aditivos, para 500 loco genômicos (*Quantitative Trait Nucleotide*, QTNs) selecionados aleatoriamente. Foram calculadas as taxas de heteroziguidade e a heterose genômica individual, e os coeficientes de regressão foram utilizados para verificar as associações entre estes fatores. Adicionalmente, uma adaptação do GWAS foi realizada para captar as variações genéticas não-aditivas do genoma, por um modelo aditivo e aditivo-dominante para as características de produção e qualidade do leite. Em média, para o cenário de menor herdabilidade (0,10), a taxa de heteroziguidade foi de 35,6%, e níveis próximos foram encontrados para os demais cenários (35,4% e 35,7%). O cenário de maior herdabilidade (0,50) apresentou os maiores valores médios de heterose genômica (17,0%), comparado aos demais que variou de 3,02 a 11,0%. Os coeficientes de regressão revelaram que a heteroziguidade tem impacto positivo sobre a heterose genômica para a maioria das populações simuladas. Em dados reais, a maioria das regiões importantes para explicar a variância genética são as mesmas entre os modelos testados, exceto para o ácido graxo poli-insaturado, que apresentou a maior quantidade de regiões ao contabilizar os efeitos de dominância. Para produção de leite e ácido graxo poli-insaturado, as estimativas de regressão entre heteroziguidade e fenótipo foram favoráveis e significativas, indicando o aumento na produtividade se houver aumento da heteroziguidade. Adicionalmente, a maior proporção de marcadores associados significativamente ao fenótipo foi por meio do modelo aditivo-dominante e oito genes foram listados como adjacentes a estas regiões. A heteroziguidade avaliada de forma empírica e por meio de dados reais revelou resultados otimistas quanto à possibilidade de selecionar os animais de raça pura, para gerar maior e melhor heterose genômica. Além disso, incluir os efeitos de dominância no modelo de GWAS adaptado pode contribuir para a identificação de marcadores associados ao fenótipo, de forma semelhante ou superior ao modelo aditivo adaptado.

Palavras-chave: Avaliação genômica. Ácido graxo. GWAS. Heterozigose. Produção de leite.

ABSTRACT

Heterozygosity and genomic heterosis are assuming importance in the evaluation of several species, due to the facilities that genomics has brought to the quantitative evaluation of genetically complex traits. However, heterosis estimated based on genome information has not been explored in populations of pure breeds or lines. Thus, the aim of this study was to investigate the possibility of using molecular heterozygosity as a information for the selection of individuals with the potential to generate greater and better genomic heterosis. Three genetic scenarios were simulated, with different magnitudes of heritability. In addition, additive and non-additive effects were simulated for 500 randomly selected genomic loci (Quantitative Trait Nucleotide, QTNs). Rates of heterozygosity and individual genomic heterosis were calculated, and regression coefficients were used to verify associations between these factors. Additionally, an adaptation of the GWAS was performed to capture the non-additive genetic variations of the genome, by an additive and additive-dominant model for milk yield and quality traits. On average, for the lower heritability scenario, the heterozygosity was 35.6%, and similar levels were found for the other scenarios (35.4 and 35.7%). The scenario of greater heritability had the highest mean values of genomic heterosis (17.0%), compared to the others, which ranged from 3.02 to 11.0%. Regression coefficients revealed that heterozygosity has a positive impact on genomic heterosis for most simulated populations. In real data, most of the regions important to explain the genetic variance are the same among the models, except for the polyunsaturated fatty acid, which presented the largest number of regions when accounting for dominance effects. For milk and polyunsaturated fatty acid production, the regression estimates between heterozygosity and phenotype were favorable and significant, indicating an increase in productivity if there is an increase in heterozygosity. Additionally, the largest proportion of markers significantly associated with the phenotype was through the additive-dominant model and eight genes were listed as adjacent. The heterozygosity evaluated empirically and through real data revealed optimistic results regarding the possibility of selecting purebred animals to generate greater and better genomic heterosis. Furthermore, including the dominance effects in the adapted GWAS model can contribute to the identification of markers associated with the phenotype, similarly or better than the adapted additive model.

Keywords: Genomic evaluation. Fatty acid content. GWAS. Heterozygosis. Milk yield.

4.1. INTRODUÇÃO

A maior quantidade de estudos que visam os modelos genômicos não -aditivos, são para a seleção de animais para o cruzamento entre raças, principalmente na produção de bovinos de corte, aves e suínos (Gengler et al., 1997; Norris et al., 2010; Vitezica et al., 2013; Amuzu-Aweh et al., 2013; Tsairidou et al., 2018; Akanno et al. 2018). Isso é conveniente quando assumido que os desvios de dominância não são um parâmetro de herança genética, mas as principais fontes genéticas para gerar heterose. Contudo, é sabido que estes efeitos podem contribuir de forma

substancial para a variabilidade genética e fenotípica de uma população, além de serem a base para compreensão da arquitetura genética de características complexas (Nicolini et al., 2018).

Na criação de bovinos leiteiros, o cruzamento entre raças pode aumentar a velocidade de aprimoramento das características, dependendo do critério utilizado para a seleção dos animais de raças puras. Isto é consistente quando realizado, por exemplo, o cruzamento entre as raças Jersey e Holandesa para melhorar as concentrações de gordura e proteína do leite, em indivíduos com maior volume de produção de leite (Sorensen et al., 2008, Kargo et al., 2021). Estas são informações valiosas quanto ao progresso genético em populações cruzadas, contudo, será benéfica se a seleção dos progenitores for de acordo com o potencial genético do indivíduo (Clasen et al., 2017; Saborío-Montero et al., 2018). Estes autores dão suporte para a hipótese de que o grupo de fatores genéticos e genômicos associados à heterozigosidade devem ser informativos quanto ao fenótipo avaliado.

A heterozigosidade do loco genômico e os efeitos de dominância são as bases moleculares para a heterose, apresentando-se fortemente associadas ao fenótipo (Tambasco-Talhari et al., 2005; Deng e Fu, 1998; Xiao et al., 1995). Dessa forma, entende-se que os efeitos de dominância estimados para cada loco ou QTL (*Quantitative Trait Loci*) poderiam refletir a ocorrência da heterose genômica e, portanto, ser utilizada como critério de seleção em populações de raça pura, para os indivíduos com maior potencial genômico para gerarem maior e melhor heterose. No entanto, esta é uma hipótese não explorada em sistemas de acasalamento e deve ser considerada devido à complexidade de muitas das características de produção de leite, que são parcialmente afetadas por efeitos genéticos não-aditivos.

Dessa forma, o objetivo deste estudo foi investigar a possibilidade em utilizar os fatores de heterozigosidade molecular como fonte de informação para a seleção de indivíduos com potencial genético para gerar maior e melhor heterose genômica, em populações simuladas e reais de bovinos da raça Holandesa. Além disso, foi proposta uma análise de associação ampla do genoma, adaptada para captar variações não-aditivas de regiões com possível vantagem heterozigótica, testado por um modelo aditivo e outro aditivo-dominante.

4.2. MATERIAL E MÉTODOS

4.2.1. Simulação da estrutura populacional e do genoma

A estrutura de população e as informações genômicas foram simuladas utilizando o software QMSim (Sargolzaei e Schenkel, 2009). Três cenários genéticos foram simulados para imprimir fenótipos com magnitude de herdabilidade (h^2) distintos, sendo 0,10, 0,30 e 0,50, que

compreendem, basicamente, características de produção e qualidade do leite. Foi simulada uma população de gado leiteiro relativamente pequena, semelhante a trabalhos anteriores (Perez et al., 2018; Compton et al., 2017 e Plieshke et al., 2016), considerando 20 repetições.

Em seguida, foi criada uma população histórica composta por 2.000 animais não relacionados e com proporção sexual equilibrada (50% de machos e 50% de fêmeas). Esses animais foram acasalados aleatoriamente por 1.000 gerações e um gargalo genético foi introduzido a partir da geração 1.001, reduzindo a população histórica de 2.000 para 200 animais por 20 gerações. O gargalo genético é utilizado para criar um cenário que exista desequilíbrio de ligação (LD) consistente na população simulada (Bradford et al., 2019).

A partir dessa população, uma segunda foi gerada e chamada de expansão, na qual os fundadores eram 100 machos e 100 fêmeas, amostrados aleatoriamente da geração 1.020 da população histórica. Essa população em expansão foi acasalada aleatoriamente por mais 60 gerações, que deu origem a 10.000 filhos, sendo 150 machos e 9.850 fêmeas na última geração (geração 1.080). A partir da população de expansão, foi criada uma população denominada recente, na qual foram selecionados aleatoriamente 50 machos e 1.500 fêmeas da geração 1.080, sendo que os acasalamentos foram administrados em 10 gerações.

Em cada geração, eram formados aproximadamente 764 machos e 891 fêmeas, que ao final totalizaram 16.500 animais em 10 ciclos reprodutivos. Foi utilizado uma taxa de descarte de 40% de machos e 20% fêmeas, para que dessa forma, a cada geração os progenitores fossem substituídos ao ponto de que todos pudessem ter a oportunidade de deixar uma progênie. Assim, a cada ciclo reprodutivo, 20 reprodutores e 300 matrizes eram substituídos e seus registros genealógicos anotados. Uma seleção aleatória dos animais foi adotada como parâmetro, no intuito de trabalhar com uma população que não sofreu nenhum tipo de seleção artificial, proporcionando a mesma oportunidade para gerar descendentes.

O genoma foi simulado para 29 pares de cromossomos autossômicos (Figura 6), com comprimento total de 2.319 cM, imitando o genoma bovino. As características foram controladas por 500 QTLs bialélicos (*Quantitative Trait Locus*), que foram espaçados de forma aleatória por todo o genoma. Além disso, os efeitos de QTL foram simulados semelhante a Bradford et al. (2019), na qual seguiram uma distribuição gama de parâmetro 0,4 para os três cenários genéticos.

Adicionalmente, havia 50.200 marcadores moleculares do tipo SNP (*Single Nucleotide Polymorphism*) bialélicos, distribuídos equidistantes por todo o genoma. Para SNP e QTL, a frequência alélica foi de 0,50 para iniciar a população histórica e $2,5 \times 10^{-5}$ de taxa de mutação, sendo a posição de ambos distribuída aleatoriamente ao longo do genoma.

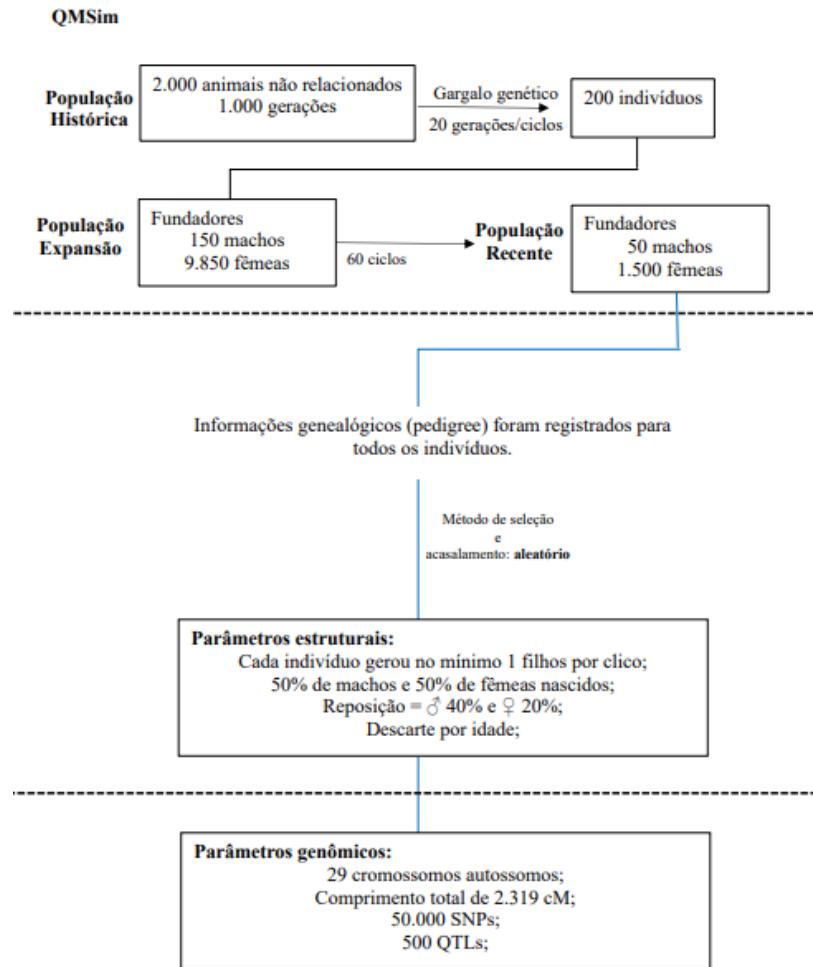


Figura 6. Esquema ilustrativo do processo de simulação.

4.2.2. Simulação dos desvios de dominância

Para modelar os efeitos genéticos de dominância, os elementos de incidência da matriz \mathbf{D} de dominância foram recodificados como 0, 1 e 0, para os genótipos $A1A1$, $A1A2/A2A1$ e $A2A2$, respectivamente. De acordo com Abdollahi-Arpanahi et al. (2020), eventos genéticos como a dominância incompleta podem ser parcialmente captados em um modelo aditivo, dessa forma, a estrutura de simulação foi construída para sobredominância (*overdominance*). Dado que os desvios de dominância, do ponto de vista estatístico, representam a interação entre os alelos, o valor médio é obtido por meio do desvio médio entre genótipos heterozigotos e homozigotos (Falconer e Mackay, 1996). Dessa forma, para modelar a ação de dominância, a distribuição segue o estudo de Abdollahi-Arpanahi et al. (2020), sendo $d \sim N(\mu = 0, \sigma = 0,50)$.

Cada efeito de dominância foi formulado como sendo $d = a * (\sum d_k)$, onde d é o desvio de dominância; a os efeitos de substituição alélica (efeito aditivo); e d_k é a dominância para k QTN, distribuídos ao longo do genoma. Os valores genéticos verdadeiros (TBVs) para os desvios de

dominância foram obtidos pela multiplicação dos códigos genotípicos (genótipos pelos efeitos de dominância de cada QTN selecionado. Os efeitos de dominância foram distribuídos de forma que, aproximadamente, 50% fossem positivos e 50% negativos.

Os fenótipos para efeito de dominância foram calculados somando a matriz dos efeitos genético aditivos, com os valores genéticos de QTN para dominância, adicionados a um valor residual normalmente distribuído $e_i \sim N(0, \sigma_e^2)$:

$$y_i = \sum_{k=1}^m X_{ik} \alpha_k + \sum_{k=1}^m Z_{ik} \alpha_k + e_i$$

onde y_i é o fenótipo para cada indivíduo i ; X_{ik} e Z_{ik} ($i = 1, \dots, n$; $k = 1, \dots, m$) são os componentes de incidência da matriz de marcadores genéticos para os efeitos aditivos e de dominância (α_k e d_k , respectivamente) e e_i é o valor residual aleatório, sendo σ_e^2 a variância residual.

4.2.3. Cálculo da heterozigose molecular e seus fatores

O nível de heterozigose foi calculado com base nas informações disponíveis 50.000 mil marcadores moleculares do tipo SNP, categorizados como homozigotos e heterozigotos (Tsairidou et al., 2018; Liu et al., 2018). Para avaliar os níveis de heterozigosidade da população, 500 QTNs foram selecionados e os respectivos desvios de dominância foram computados para os genótipos heterozigotos de cada indivíduo. Por exemplo, supondo que o efeito de dominância de um dado loco genômico foi de 0,50 ou -0,50, o genótipo heterozigoto para esse mesmo loco será igual ao seu valor, seguindo para cada indivíduo (i) para todos os k QTNs avaliados.

Na teoria clássica de Falconer e Mackay (1989), se em um determinado loci o efeito de dominância for ausente ($d = 0$) e/ou o loco apresentar desvio de dominância negativo, não haverá heterose. Portanto, esse conceito estendido para um nível molecular, na qual para um animal i , a soma dos desvios de dominância de k QTNs foi, respeitosamente, chamada de heterose genômica. Isto posto, foi considerado como uma desvantagem heterozigótica o loco que apresentou desvio de dominância negativo, dessa forma, avalia-se que para gerar maior e melhor heterose, os indivíduos que apresentam heterozigosidade e heterose genômica positiva serão preferíveis quanto a seleção para acasalamento.

A heterozigosidade foi obtida pela divisão do número total de genótipos heterozigotos de cada indivíduo pelo total de QTNs avaliados (Cao et al., 2019). Adicionalmente, o nível de heterozigose foi logrado pela soma dos efeitos de dominância do loco genômico, dividido pelo total de QTNs avaliados (Iversen et al., 2019; Tambasco-Talhari et al., 2005). Análises de

regressão linear univariada foram aplicadas para cada repetição, nos três cenários genéticos simulados. Para melhor compreender os efeitos da heterozigosidade e as ocorrências de heterose genômica, cada repetição foi considerada como uma população de indivíduos aptos à seleção, dessa forma foi possível verificar as (des)vantagens heterozigóticas.

O modelo de regressão utilizado foi: $y_i = \alpha + \beta x_i + \varepsilon_i$; onde y_i é uma variável resposta na i -ésima observação; x_i é a variável regressora na i -ésima observação; ε_i variável de resíduo; α e β são os coeficientes de regressão a serem estimados, que representam o intercepto e o coeficiente angular da reta de regressão, respectivamente. Três modelos foram aplicados: **1)** heterose \sim heterozigosidade; **2)** fenótipo \sim heterozigosidade; **3)** fenótipo \sim heterose. Para testar o efeito da heterozigosidade e seus fatores em cada parâmetro regredido, os coeficientes de regressão estimados foram divididos pelo desvio-padrão de cada parâmetro (Iversen et al., 2019). Em dados reais, foram utilizadas as informações de heterozigosidade das regiões genômicas que foram mais importantes para explicar a variância genética aditiva da característica, em função do fenótipo ajustado para os efeitos fixos.

4.2.4. Análise de associação ampla do genoma (GWAS) adaptado para heterozigosidade

Uma adaptação da análise de associação ampla do genoma (*Genome-wide Association Study*, GWAS) foi realizada para verificar as variações genéticas não-aditivas dos marcadores moleculares associados ao fenótipo. Essa etapa foi idealizada para complementar a etapa de simulação, com o objetivo de identificar possíveis regiões genômicas e marcadores moleculares em vantagem heterozigótica, para serem selecionados.

A análise de GWAS adaptado foi realizado para cinco características de relevante impacto econômico para a bovinocultura leiteira, sendo elas a produção de leite no dia do controle leiteiro (PL, kg/dia), porcentagem de proteína (PROT, %), gordura (GORD, %), caseína (CAS, %) e conteúdo de ácido graxo poli-insaturado (POLI, %) por estar associado aos aspectos de saúde do consumidor. Todos os procedimentos com os animais seguiram os padrões éticos de pesquisa, estabelecidos pela Universidade de São Paulo e demais instituições envolvidas, presente na proposta intitulada “Incorporação de informações genômicas no estudo de características de composição e qualidade do leite de vacas”, protocolo n° 2017.5.1197.11.3, aprovada pela comissão de ética no uso de animais – CEUA da ESALQ/USP.

Fenotipagem

O banco de fenótipos era composto por 7.664 vacas da raça Holandesa e 127.432 registros mensais para as características de produção e qualidade do leite. As mensurações eram oriundas de três fazendas brasileiras localizadas no Estado de São Paulo e uma no Paraná, coletados entre os anos de 2012 e 2016 em parceria com a Clínica do Leite, pertencente ao Departamento de Zootecnia da Escola Superior de Agricultura “Luiz de Queiroz”, da Universidade de São Paulo. Nesses rebanhos, as fêmeas eram criadas em sistema *free-stall* e o manejo de ordenha ocorria três vezes ao dia sob sistema mecânico. Adicionalmente, havia registros genealógicos disponíveis para todos os rebanhos.

Como forma de controle de possíveis erros amostrais, foi realizada uma checagem de consistência dos dados fenotípicos, genealógicos e do genoma dessa população. Foram excluídos os animais sem nenhuma informação de fenótipo, com idade superior a 10 anos, dias em lactação menor que 5 e maior que 305 dias e aqueles animais que não possuíam informações de data de nascimento, coleta da amostra de leite e de parto, e os registros com produção de leite abaixo de 3,0 e acima de 99,0 kg de leite/dia e para proteína e gordura, foi considerado os registros com desvio-padrão de até $\pm 3,0$, sendo uma variação aproximada de 1,0 à 9,0% e para gordura de 1,5 à 12,0% (ICAR, 2017). A estatística descritiva para os fenótipos é apresentada na tabela 9.

Tabela 9. Número de observações (N), média, desvio-padrão (DP), valores de mínimo, máximo e herdabilidade genômica aditiva (h^2) e o respectivo desvio-padrão (em parênteses) para produção de leite (PL), concentração de proteína (PROT), gordura (GROD), caseína (CAS) e conteúdo de ácido graxo poli-insaturado (POLI).

Característica	N	Média	DP	MIN	MAX	h^2
PL (kg/dia)	15.780	32,19	11,53	3,02	80,10	0,21 (0,059)
PROT (%)	17.894	3,13	0,36	1,40	5,83	0,31 (0,037)
GORD (%)	17.809	3,49	0,79	1,50	11,92	0,18 (0,032)
CAS (%)	13.623	2,38	0,32	0,92	4,68	0,34 (0,051)
POLI (%)	12.438	0,16	0,05	0,01	1,09	0,09 (0,001)

A partir disso, os grupos de contemporâneos (GC) foram formados concatenando fazenda, mês e ano de parto e da coleta da amostra do leite. Uma análise de variância, ANOVA, foi realizada por meio do software R versão 4.1.0 para testar o nível de significância das variáveis incluídas no GC. Os três fatores testados foram significativamente diferentes de zero pelo teste F , portanto as variáveis agrupadas foram utilizadas como efeito fixo nas análises genéticas subsequentes. As estimativas de componentes de variância foram obtidas pelo método de

máxima verossimilhança restrita com o algoritmo de informação média sob o modelo genômico-poligênico (M1), no programa AIREMLF90, por meio do programa BLUPF90 (Misztal et al., 2018).

Informações genômicas

O genoma da população era composto por 1.157 animais amostrados, sendo que 768 vacas foram genotipadas em painel de baixa densidade (Illumina Bovine LD BeadChip, 6K), com 6.909 polimorfismos de nucleotídeo único (*Single Nucleotide Polymorphism*, SNP) e 389 vacas genotipadas em painel de média densidade, composto por 47.843 SNPs (GeneSeek Genomic Profile Bovine 50K). Tais informações já estão disponíveis, tendo sido obtidas em projetos anteriores (CNPq 482327/2010-3; 484560/2012-3). Foram coletadas amostras de bulbos pilosos extraídos da vassoura da cauda e a extração do material genético realizado por meio do Kit NucleoSpin Tissue (Macharey-Nagel GmbH & Co. KG, Duren, Alemanha) e a leitura dos chips de genotipagem feita no equipamento HiScanTMSQ System (Illumina, Inc., 2011). A extração de DNA e leitura dos chips foram realizadas no Laboratório de Biotecnologia Animal (Departamento de Zootecnia, ESALQ/USP).

Adicionalmente, os dados genotípicos foram enriquecidos por análise de imputação, conforme detalhado por Petrini et al. (2015) e Iung et al. (2019). Para esse procedimento, a população referência foi formada por touros, que eram pais das vacas genotipadas, que possuíam genótipo para 79.294 SNPs nas plataformas Illumina BovineSNP50 (Illumina, San Diego, EUA) ou GeneSeek Genomic Profiler HD (Neogen Agrigenomics, Lexington, EUA). Essa análise de imputação foi desenvolvida pelo Animal Genomics and Improvement Laboratory (Agriculture Research Service, USDA) sob encargo do Dr. George Wiggans, que utilizou o programa findhap.f90 (VanRaden, 2015).

Após a análise de imputação, oito animais com genótipos duplicados ou com parentesco incompatível foram excluídos, dessa forma 1.149 animais com informações de 79.294 marcadores moleculares restaram no banco de dados original. Foi aplicado um controle de qualidade (QC) no intuito de reduzir erros de genotipagem e o esforço computacional das análises. Nesse controle, foram excluídos 7.899 marcadores com *call-rate* menor que 90%, 32 SNPs monomórficos e 1.241 com frequência alélica (*minor allele frequency*, MAF) menor que 0,02. Além disso, foi utilizado como critério de frequência de heterozigotos, valores maiores que 0,15 do que o esperado quando em equilíbrio de Hardy-Weinberg, no entanto nenhum marcador foi excluído.

Adicionalmente, também foram excluídos das análises os marcadores presentes nos cromossomos sexuais e 56 animais que apresentaram conflito de parentesco, após a análise de

imputação. O QC foi realizado de forma específica para cada característica, no programa PREGSF90 (Aguilar et al., 2014). Dessa forma, para a produção de leite, foram efetivados 1.096 animais e 70.193 marcadores moleculares do tipo SNP.

4.2.5. Delineamento da análise de GWAS adaptado

No modelo puramente aditivo, assume-se que há uma mudança linear no fenótipo em relação ao menor número de cópias do alelo (Vidotti et al., 2019), dessa forma, a matriz de marcadores do tipo SNP é codificada como 0 (A_1A_1), 1 (A_1A_2 e A_2A_1) e 2 (A_2A_2). Por outro lado, para o modelo de (des)vantagem do heterozigoto na análise de associação genômica, uma codificação específica foi utilizada, na qual os genótipos homozigotos (A_1A_1 ou A_2A_2) foram considerados como tendo o mesmo efeito, enquanto os heterozigotos um efeito diferente (Tsairidou et al., 2018; Vidotti et al., 2019). Posto isso, a matriz de SNPs foi recodificada de forma especial, sendo 0 para os homozigotos (A_1A_1 ou A_2A_2) e 1 para os heterozigotos (A_1A_2 ou A_2A_1).

Para verificar possíveis mudanças quanto à estimação dos efeitos dos SNPs, um modelo genético aditivo (M1) e outro que contabiliza o efeito de dominância (M2), foram ajustados:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{a} + \mathbf{S}\mathbf{c} + \boldsymbol{\varepsilon} \quad (\text{M1})$$

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{a} + \mathbf{W}\mathbf{d} + \mathbf{S}\mathbf{c} + \boldsymbol{\varepsilon} \quad (\text{M2})$$

em que \mathbf{y} é o vetor de observações fenotípicas; \mathbf{X} é a matriz de incidência para os efeitos fixos de grupo de contemporâneos (GC); $\boldsymbol{\beta}$ é o vetor dos efeitos fixos; \mathbf{Z} a matriz de incidência para o efeito genético aditivo aleatório; \mathbf{a} é o vetor de soluções para efeito aleatório; \mathbf{W} é a matriz de incidência para os efeitos genéticos de dominância aleatório; \mathbf{d} é o vetor de soluções para o efeito genético de dominância; \mathbf{S} é a matriz de incidência para o efeito aleatório de ambiente permanente; \mathbf{c} é o vetor de soluções para o efeito aleatório de ambiente permanente; $\boldsymbol{\varepsilon}$ é o vetor para os efeitos residuais.

A primeira etapa do GWAS consistiu na obtenção dos valores genéticos genômicos (GEBV) por meio dos modelos supracitados do programa BLUPF90 (Misztal et al., 2009) que, além de fornecer as matrizes genômicas e as soluções dos GEBVs, estima o nível de significância (*p-value*) dos SNPs. A equação utilizada para computar esses efeitos segue como:

$$\hat{u} = \mathbf{DZ}'[\mathbf{ZDZ}']^{-1}\hat{\mathbf{a}}_g$$

em que \hat{u} é o vetor de efeito para cada SNP; \mathbf{D} é a matriz diagonal contendo ponderadores para o efeito de cada SNPs; \mathbf{Z} é a matriz de genótipos; $\hat{\mathbf{a}}_g$ é o vetor contendo os valores genéticos genômicos dos animais genotipados.

Considerando que para esse procedimento o genoma foi modificado para heteroziguidade, a proporção da variância explicada por n SNP foi calculada com base em janelas de 40 marcadores adjacentes para $-\log_{10}(p\text{-value})$, em vez de ser individual (Tsairidou et al., 2018). A partir das estimativas dos efeitos para o SNP, a variância explicada para os SNPs foi estimada seguindo Wang et al. (2014):

$$\frac{Var(a_i)}{\sigma_a^2} \times 100 = \frac{Var(\sum_{j=1}^{40} Z_j \hat{u}_j)}{\sigma_a^2} \times 100$$

em que a_i é o valor genético de cada região genômica que consistiu em 40 SNPs adjacentes; σ_a^2 é a variância genética aditiva total; Z_j é o vetor que contém o genótipo do SNP para todos os indivíduos; \hat{u}_j é o efeito de cada marcador dentro da região genômica estipulada.

As estimativas dos efeitos de SNP foram realizadas de forma iterativa para três repetições (3 rounds) e foi utilizado um vetor de uns (n) como ponderador de SNP. Essa metodologia é conhecida como GWAS de etapa única ponderado (*Weighted single-step GWAS*, WssGWAS), que permite possíveis variâncias desiguais entre os marcadores para uma melhor acurácia na estimação dos efeitos dos SNPs (Wang et al., 2012).

Portanto, se a estrutura de população é pouco informativa ou quando o número de indivíduos genotipados é pequeno para uma característica poligênica, o *single-step* GWAS pode apresentar um desempenho superior comparado às abordagens tradicionais (Marques et al., 2018). Além disso, a utilização do ponderador pode minimizar a identificação de falsos positivos associados ao fenótipo (Habier et al., 2011). Esse mesmo procedimento foi adotado, por exemplo, em espécies como bovinos leiteiros (Tiezzi et al., 2015), suínos (Marques et al., 2018), frangos (Fragomeni et al., 2014) e ovinos (Rovadoscki et al., 2018).

Além dos efeitos dos marcadores, em estudos de associação ambiciosos que interessam identificar regiões genômicas ou QTLs de efeito importante para a característica avaliada, os valores de p obtidos para o SNP são fundamentais para tal propósito. Além disso, recentes *upgrades* no pacote de programas BLUPF90, tornaram a etapa única de associação genômica uma estratégia eficiente para lidar com grandes conjuntos de dados e modelos genéticos complexos

(Aguilar et al., 2019). Dessa forma, para saber se o efeito de substituição do alelo difere de zero, os valores de p foram obtidos com base no erro de predição a matriz de (co)variâncias para os SNPs:

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

onde Φ é uma função normal padrão cumulativa; $sd(\hat{a}_i)$ é a raiz quadrada da variância do erro de predição (PEV) para j -ésimo efeito de SNP obtidos a partir da inversa da matriz de coeficientes das equações de modelos mistos.

Uma vez que a variância para os SNPs foi calculada, para a proporção da variância genética foi considerada as regiões que explicaram pelo menos 1% da variância genética aditiva para o fenótipo e os limites de significância de associação do marcador com fenótipo de $-\log_{10}(0,05)$ e de $-\log_{10}(0,08)$, a nível de cromossomo e genoma, respectivamente. Todas as etapas de estimação dos efeitos de SNP foram realizadas apenas com animais fenotipados que possuíam informação genômica, por meio do programa POSTGSF90 (Aguilar et al., 2014).

Adicionalmente, foi realizada a anotação dos genes e a obtenção dos termos de *Gene Ontology* (GO *Terms*) por meio da ferramenta BioMart da plataforma Ensembl (<https://www.ensembl.org/biomart>), para os SNPs que atingiram os limites de significância pré-estabelecidos na análise de GWAS adaptado. A versão do genoma de referência utilizado foi a ARS-UCD1.2 para *Bos Taurus* (bos Tau9), com última atualização no ano de 2018 (https://www.ncbi.nlm.nih.gov/assembly/GCA_002263795.2). Foram considerados genes importantes aqueles com relevância funcional para a produção de leite e constituintes sólidos.

4.3. RESULTADOS

4.3.1. GWAS adaptado

Não há uma razão específica e nem um relato consistente a respeito do tamanho da janela de SNPs a serem considerados nas avaliações de associação genômica, portanto, os modelos aditivos e aditivos-dominantes foram ajustados considerando uma janela móvel de 40 SNPs adjacentes, para explicarem a variância genética aditiva do fenótipo.

Produção de leite

No modelo aditivo, um total de 25 regiões estavam dentro do limite de 1% para 40 marcadores adjacentes, localizados nos cromossomos (chr) 1, 3, 7, 9, 11, 12, 16, 20 e 29 (Tabela 10). Para as regiões do genoma formado por janelas com 40 SNPs adjacentes, 15 marcadores apresentaram maior impacto sobre o fenótipo, na qual explicaram cerca de 41,85% da variância genética aditiva, com variação de 1,06 a 2,39%. Quando ajustado o modelo aditivo-dominante, 10 regiões explicaram mais de 1% da variância genética aditiva, com total de aproximadamente 35,73%. A maior quantidade de regiões acima do limite de variação e com maiores porcentagens da variância foram registrados quando utilizado o modelo aditivo, no entanto, várias dessas regiões também foram identificadas no modelo aditivo-dominante.

Tabela 10. Identificação do SNP (inicial_final), tamanho da janela (quantidade de SNP), cromossomo (CHR), posição inicial e final da janela no cromossomo (POS_{inicial}_POS_{final}, Mpb) e porcentagem da variância genética aditiva explicada (% σ_a^2) por janela de 40 marcadores adjacentes avaliados no modelo aditivo e aditivo-dominante, para a produção de leite (kg/dia).

Identificação do SNP	Janela	CHR	POS _{inicial} _POS _{final}	% σ_a^2
Modelo aditivo				
ARS-BFGL-NGS-27691_BovineHD1100014331*	40	11	47378387_48847011	2,36
Hapmap45307-BTA-65775_ARS-BFGL-NGS-35587*	40	29	40988761_42330369	2,17
BovineHD0300009648_ARS-BFGL-NGS-42747*	40	3	30484376_31575300	1,76
BovineHD0900019394_BovineHD0900019816*	40	9	69124337_70477400	1,70
Hapmap39216-BTA-107320_ARS-BFGL-NGS-4427	40	11	80160161_81840635	1,66
BovineHD0700007089_ARS-BFGL-NGS-33569*	40	7	24494025_25725796	1,59
BovineHD1200014085_BTBTB-00493207*	40	12	50850657_52114901	1,53
ARS-BFGL-NGS-17688_BovineHD1200005150	40	12	15767379_16896392	1,50
BovineHD1600007188_ARS-BFGL-NGS-54368*	40	16	25152110_26870109	1,46
BovineHD1600005651_BovineHD1600006144*	40	16	19740175_21187584	1,35
Hapmap44892-BTA-50337_Hapmap38791-BTA-89194	40	1	10724714_13177821	1,33
ARS-BFGL-BAC-36190_BovineHD2000007814	40	20	24156101_26083947	1,11
Hapmap23765-BTA-151182_BTBTB-01444868	40	11	22485728_23678911	1,10
BovineHD2500010289_BovineHD2500010568	40	25	36560015_37411428	1,09
BovineHD2000018509_ARS-BFGL-NGS-3962	40	9	64464857_65955264	1,06
%Total				41,85
Modelo aditivo-dominante				
ARS-BFGL-NGS-27691_BovineHD1100014331*	40	11	47378387_48847011	1,89
BovineHD1600007188_ARS-BFGL-NGS-54368*	40	16	25152110_26870109	1,62
Hapmap45307-BTA-65775_ARS-BFGL-NGS-35587*	40	29	40988761_42330369	1,61
BovineHD0900019394_BovineHD0900019816*	40	9	69124337_70477400	1,33
ARS-BFGL-NGS-17688_BovineHD1200005150	40	12	15767379_16896392	1,32
BovineHD0300009648_ARS-BFGL-NGS-42747*	40	3	30484376_31575300	1,30
BovineHD0700007089_ARS-BFGL-NGS-33569*	40	7	24494025_25725796	1,25
BovineHD1600005651_BovineHD1600006144*	40	16	19740175_21187584	1,24
BovineHD1200014085_BTBTB-00493207*	40	12	50850657_52114901	1,23
BTA-107309-no-rs_BovineHD1100023515	40	11	80182206_81943182	1,13
%Total				35,73

*Janela em comum entre os modelos.

A região dos SNPs ARS-BFGL-NGS-27691_BovineHD1100014331, por exemplo, está presente em ambos os modelos. Esse marcador está localizado no cromossomo 11, na janela

47378387_48847011 Mpb (mega pares de base) e apresentou a maior proporção da variância explicada no modelo aditivo (2,36%), comparado ao modelo aditivo-dominante (aproximadamente 1,68%). Embora não tenha sido associado significativamente com o fenótipo, devido a sua contribuição para explicar a variância genética e por estar em evidência quanto à condição de heterozigiosidade estabelecida nesta avaliação, esse marcador pode ser um, dentre os demais, candidato para explorar a heterose à nível molecular para a produção de leite.

Para a produção de leite, nenhum SNP alcançou o limite de significância para associação com o fenótipo no modelo aditivo. No entanto, quando incluído no modelo os efeitos de dominância os SNPs 36618 (BTA-30857-no-rs) na posição 76081859 no cromossomo 12, 39373 (BovineHD1300023864) no cromossomo 13 e posição 81532044 Mpb, e o SNP 50658 (BovineHD1900002699) posicionado à 9334455 Mpb no cromossomo 19, foram significativos a nível de todo o cromossomo com $-\log_{10} P = 5,49$, $P = 5,08$ e $P = 5,52$, respectivamente (Figura 7). Além disso, os SNPs 35490 no cromossomo 12 (posição 27982437 Mpb) e o SNP 68277 no cromossomo 29 (posição 47672750 Mpb) estiveram muito próximos do limite de significância a nível cromossômico, com $-\log_{10} P$ de 1,50 e 2,70, respectivamente.

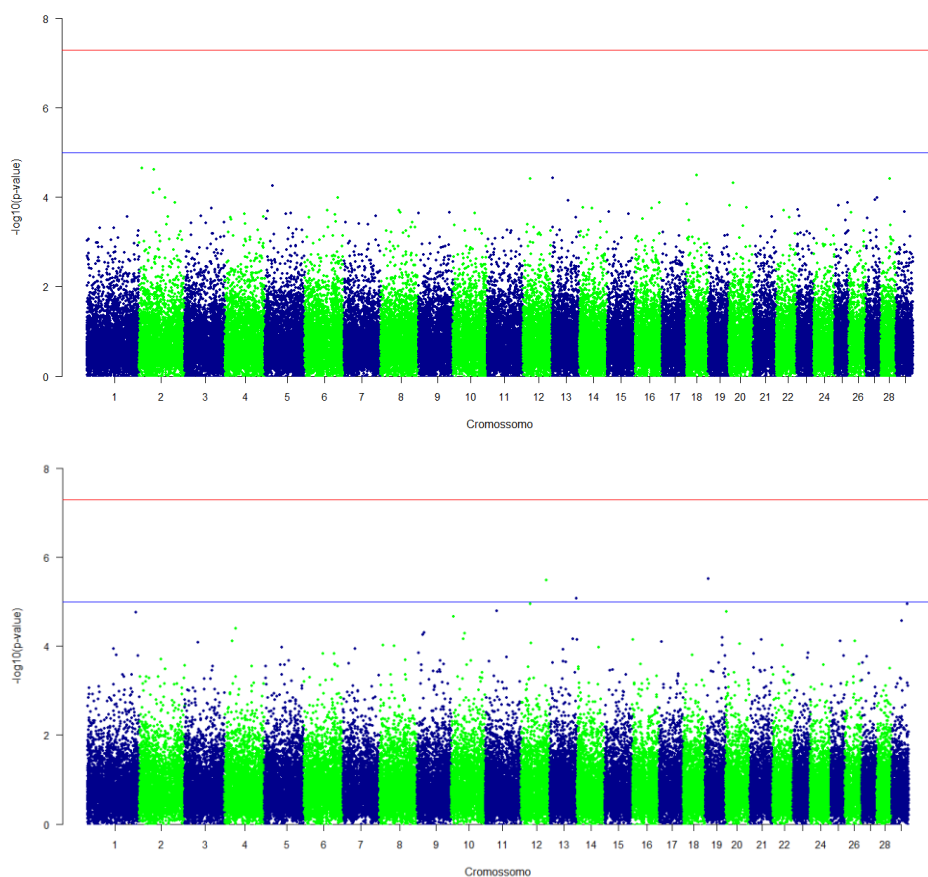


Figura 7. Gráfico do tipo Manhattan para a associação dos marcadores com a produção de leite com limites de $-\log_{10}$ (0,05) para o cromossomo (linha azul) e $-\log_{10}$ (0,08) a nível de genoma (linha vermelha), para o modelo aditivo (gráfico superior) e aditivo-dominante (gráfico inferior).

Gordura

Houve pouca variação entre os modelos ajustados referente às regiões genômicas de maior impacto no fenótipo. As regiões com porcentagem da variância explicada maior que 1% são as mesmas entre o modelo aditivo e aditivo-dominante (Tabela 11). Estas explicaram cerca de 37,87% da variância genética da gordura no leite e todos estão localizados nos cromossomos 3, 5, 6, 7, 9, 13, 14,15 e 17, sendo que a região dos marcadores BovineHD1400024339_ ARS-BFGL-NGS-22866 (chr 14, posição 146715 a 1681783 Mpb) teve o maior impacto sobre a variância, com porcentagem de 26,23% e 26,28% no modelo aditivo e aditivo-dominante, respectivamente.

Tabela 11. Identificação do SNP (inicial_final), tamanho da janela (quantidade de SNP), cromossomo (CHR), posição inicial e final da janela no cromossomo (POSinicial_POSfinal, Mpb) e porcentagem da variância genética aditiva explicada ($\% \sigma_a^2$) por janela de 40 marcadores adjacentes avaliados no modelo aditivo e aditivo-dominante, para a concentração de gordura no leite.

Identificação do SNP	Janela	CHR	POS _{inicial} _POS _{final}	$\% \sigma_a^2$
Modelo aditivo				
BovineHD1400024339_ ARS-BFGL-NGS-22866*	40	14	146715_1681783	26,23
BovineHD0300011574_ BTA-86782-no-rs*	40	3	37209678_39904161	2,25
BovineHD0500026491_ BovineHD0500026737*	40	5	92860844_93719732	1,44
BTA-21929-no-rs_ ARS-BFGL-NGS-13988*	40	17	49224250_50522879	1,42
BovineHD0700005017_ ARS-BFGL-NGS-12159*	40	7	16632028_17996900	1,20
ARS-BFGL-NGS-40765_ BovineHD1500021864*	40	15	72976813_74332306	1,12
BovineHD1400000788_ Hapmap23302-BTC-052123*	40	14	2840340_3819126	1,09
BovineHD1300018940_ BPI-3*	40	13	66163790_67177128	1,07
ARS-BFGL-NGS-18608_ BTA-102674-no-rs*	40	9	68260591_69659376	1,02
BovineHD0600023575_ Hapmap27109-BTC-060711*	40	6	84000396_85422786	1,01
%Total				37,85
Modelo aditivo-dominante				
BovineHD1400024339_ ARS-BFGL-NGS-22866*	40	14	146715_1681783	26,28
BovineHD0300011574_ BTA-86782-no-rs*	40	3	37209678_39904161	2,26
BovineHD0500026491_ BovineHD0500026737*	40	5	92860844_93719732	1,46
BTA-21929-no-rs_ ARS-BFGL-NGS-13988*	40	17	49224250_50522879	1,40
BovineHD0700005017_ ARS-BFGL-NGS-12159*	40	7	16632028_17996900	1,20
ARS-BFGL-NGS-40765_ BovineHD1500021864*	40	15	72976813_74332306	1,11
BovineHD1400000788_ Hapmap23302-BTC-052123*	40	14	2840340_3819126	1,08
BovineHD1300018940_ BPI-3*	40	13	66163790_67177128	1,06
ARS-BFGL-NGS-18608_ BTA-102674-no-rs*	40	9	68260591_69659376	1,02
BovineHD0600023575_ Hapmap27109-BTC-060711*	40	6	84000396_85422786	1,01
%Total				37,87

*Janela em comum entre os modelos.

Para ambos os modelos foi registrado diferentes SNPs associados significativamente a nível de todo o cromossomo, sendo quatro quando utilizado o modelo aditivo e três com o modelo aditivo-dominante (Figura 8). Entretanto, nenhum marcador atingiu o limite de significância a nível de todo o genoma. No modelo aditivo, os marcadores associados são o SNP 47614 (ARS-BFGL-NGS-57149), 60993 (Hapmap30536-BTA-137818), 64514 (BovineHD2700003888) e 66133 (BTB-00981253); localizados nos cromossomos 17 (posição 54062683 Mpb), 24 (posição 42090544 Mpb), 27 (posição 14315992 Mpb) e 28 (posição

22219573 Mpb), respectivamente. Para o modelo aditivo-dominante, os marcadores estão localizados nos cromossomos 13 (posição 2545062 Mpb), 21 (posição 2498610 Mpb) e 22 (posição 19093982 Mpb); SNPs 37040 (BovineHD4100009853), 54416 (BovineHD2100000332) e 56908 (BTB-00839323), respectivamente.

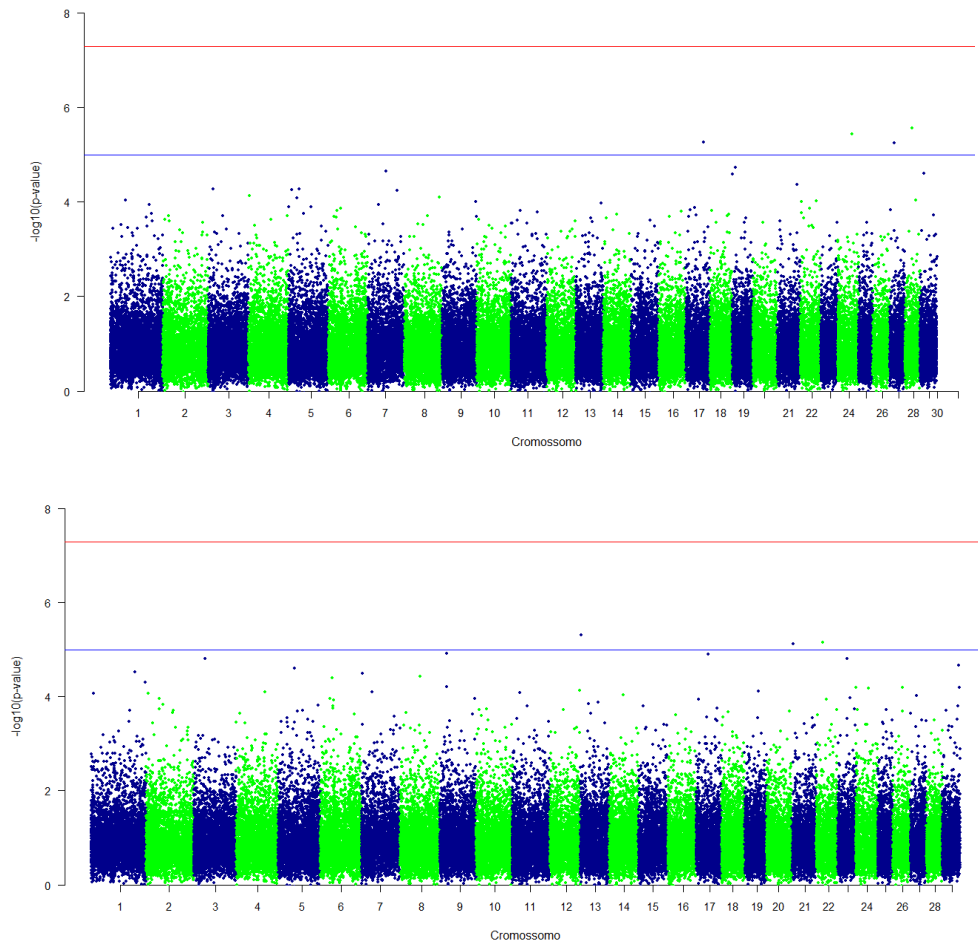


Figura 8. Gráfico do tipo Manhattan para a associação dos marcadores com a concentração de gordura no leite, com limites de $-\log_{10}(0,05)$ para o cromossomo (linha azul) e $-\log_{10}(0,08)$ a nível de genoma (linha vermelha), para o modelo aditivo (gráfico superior) e aditivo-dominante (gráfico inferior).

Proteína

Tanto no modelo aditivo quanto no aditivo-dominante, oito regiões apresentaram maior impacto sobre a variância genética aditiva para a concentração de proteína no leite, localizados nos cromossomos 1, 7, 9, 15, 19, 22, 27 e 28 (Tabela 12). Para o modelo aditivo essa proporção foi de 13,92%, cerca de 2,03% maior comparado ao modelo aditivo-dominante (porcentagem absoluta de 11,89%).

Tabela 12. Identificação do SNP (inicial_final), tamanho da janela (quantidade de SNP), cromossomo (CHR), posição inicial e final da janela no cromossomo (POSinicial_POSfinal, Mpb) e porcentagem da variância genética aditiva explicada ($\% \sigma_a^2$) por janela de 40 marcadores adjacentes avaliados no modelo aditivo e aditivo-dominante, para a concentração de proteína no leite.

Identificação do SNP	Janela	CHR	POS _{inicial} _POS _{final}	$\% \sigma_a^2$
Modelo aditivo				
ARS-BFGL-NGS-39645_BovineHD2200012346*	40	22	40916767_42054278	3,12
BTA-34638-no-rs_ARS-BFGL-NGS-39750*	40	1	71190486_73143183	2,60
BovineHD1500009999_BovineHD1500010339*	40	15	35856607_37078463	2,01
ARS-BFGL-NGS-49079_ARS-BFGL-NGS-34350*	40	9	6450799_8405821	1,65
BovineHD1900016587_BovineHD1900017041*	40	19	57724102_59090292	1,30
ARS-BFGL-NGS-17601_BovineHD2800012477*	40	28	42346190_43243315	1,17
Hapmap43946-BTA-80204_BovineHD0700028201*	40	7	92574720_94403482	1,06
Hapmap24215-BTA-163266_BovineHD2700003645*	40	27	11736207_13523197	1,01
%Total				13,92
Modelo aditivo-dominante				
ARS-BFGL-NGS-39645_BovineHD2200012346*	40	22	40916767_42054278	3,12
BTA-34638-no-rs_ARS-BFGL-NGS-39750*	40	1	71190486_73143183	2,60
BovineHD1500009999_BovineHD1500010339*	40	15	35856607_37078463	2,02
ARS-BFGL-NGS-49079_ARS-BFGL-NGS-34350*	40	9	6450799_8405821	1,64
BovineHD1900016587_BovineHD1900017041*	40	19	57724102_59090292	1,30
ARS-BFGL-NGS-17601_BovineHD2800012477*	40	28	42346190_43243315	1,17
Hapmap43946-BTA-80204_BovineHD0700028201*	40	7	92574720_94403482	1,05
Hapmap24215-BTA-163266_BovineHD2700003645*	40	27	11736207_13523197	1,01
%Total				11,89

*Janela em comum entre os modelos.

Um total de oito marcadores foram associados significativamente com a concentração de proteína no leite, quando incluído os efeitos de dominância no modelo (Figura 9).

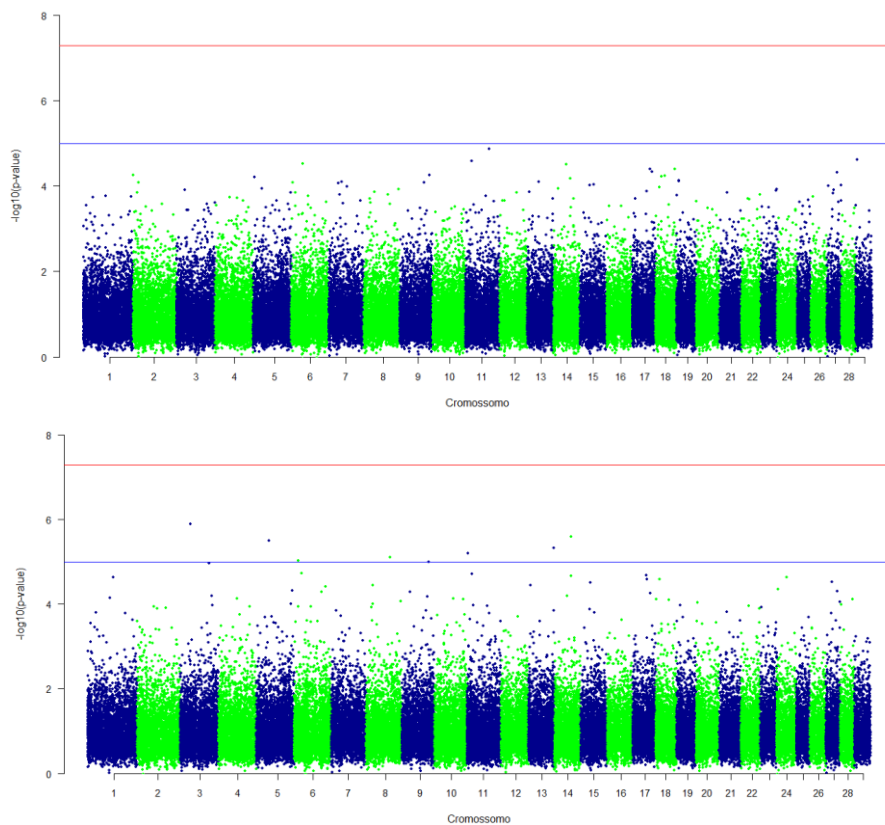


Figura 9. Gráfico do tipo Manhattan para a associação dos marcadores com a concentração de proteína no leite, com limites de $-\log_{10}(0,05)$ para o cromossomo (linha azul) e $-\log_{10}(0,08)$ a nível de genoma (linha vermelha), para o modelo aditivo (gráfico superior) e aditivo-dominante (gráfico inferior).

Os SNPs associados foram o SNP 8209 (ARS-BFGL-NGS-85942), 14706 (BovineHD0500011755), 17341 (ARS-BFGL-NGS-117043), 24767 (BTB-00359112), 28104 (BTB-00402806), 31812 (ARS-USMARC-Parent-DQ837646-rs29012894), 39350 (ARS-BFGL-NGS-25084) e o SNP 40339 (BovineHD1400008879); localizados, respectivamente, nos cromossomos 3 (posição 32529922 Mpb), 5 (posição 40936260 Mpb), 6 (posição 13415776 Mpb), 8 (posição 75072384 pb), 9 (posição 85181556 Mpb), 11 (posição 1714947 Mpb), 13 (posição 81126899 Mpb) e 14 (posição 29039917 Mpb). Para ajuste com um modelo puramente aditivo, nenhum SNP atingiu o limite de significância estabelecido.

Caseína

Para a concentração de caseína no leite, as regiões identificadas de maior impacto sob a variância genética aditiva foram semelhantes para os dois modelos ajustados (Tabela 13).

Tabela 13. Identificação do SNP (inicial_final), tamanho da janela (quantidade de SNP), cromossomo (CHR), posição inicial e final da janela no cromossomo (POSinicial_POSfinal, Mpb) e porcentagem da variância genética aditiva explicada ($\% \sigma_a^2$) por janela de 40 marcadores adjacentes avaliados no modelo aditivo e aditivo-dominante, para a concentração de caseína no leite.

Identificação do SNP	Janela	CHR	POS _{inicial} _POS _{final}	$\% \sigma_a^2$
Modelo aditivo				
ARS-BFGL-NGS-49079_ARS-BFGL-NGS-34350*	40	9	6450799_8405821	3,10
BTA-34638-no-rs_ARS-BFGL-NGS-39750*	40	1	71190486_73143183	2,94
BovineHD1500009999_BovineHD1500010339*	40	15	35856607_37078463	2,21
ARS-BFGL-NGS-22306_BovineHD0300029063*	40	3	99320849_100987554	1,85
BTB-00258161_BTA-106726-no-rs*	40	8	32526176_34413649	1,72
Hapmap35988-SCAFFOLD166019_6545_BovineHD2500006819*	40	25	22329517_23907681	1,63
ARS-BFGL-NGS-26423_ARS-BFGL-NGS-1045*	40	15	74763309_75726555	1,60
BovineHD0800024468_ARS-BFGL-NGS-105269*	40	8	80869483_81631309	1,58
BTB-00396685_BovineHD0900017275*	40	9	60328344_61996687	1,55
BovineHD1600016812_BovineHD1600017107*	40	16	58664360_59498662	1,50
ARS-BFGL-NGS-12851_ARS-BFGL-NGS-3758*	40	22	40936789_42071416	1,49
ARS-BFGL-NGS-41958_BT-00954129*	40	27	11669357_13486135	1,38
BTB-01734642_BTA-41946-no-rs*	40	1	29520198_31171747	1,28
ARS-BFGL-NGS-78057_BT-01382552*	40	6	47353083_49140436	1,25
BovineHD1100005191_BovineHD1100005496*	40	11	16528054_17815368	1,15
ARS-BFGL-NGS-103036_BovineHD0400027828*	40	4	96201713_98497972	1,07
BovineHD1900016587_BovineHD1900017041*	40	19	57724102_59090292	1,04
%Total				28,34
Modelo aditivo-dominante				
ARS-BFGL-NGS-49079_ARS-BFGL-NGS-34350*	40	9	6450799_8405821	3,10
BTA-34638-no-rs_ARS-BFGL-NGS-39750*	40	1	71190486_73143183	2,93
BovineHD1500009999_BovineHD1500010339*	40	15	35856607_37078463	2,20
ARS-BFGL-NGS-22306_BovineHD0300029063*	40	3	99320849_100987554	1,85
BTB-00258161_BTA-106726-no-rs*	40	8	32526176_34413649	1,72
Hapmap35988-SCAFFOLD166019_6545_BovineHD2500006819*	40	25	22329517_23907681	1,63
ARS-BFGL-NGS-26423_ARS-BFGL-NGS-1045*	40	15	74763309_75726555	1,61
BovineHD0800024468_ARS-BFGL-NGS-105269*	40	8	80869483_81631309	1,58
BTB-00396685_BovineHD0900017275*	40	9	60328344_61996687	1,55
BovineHD1600016812_BovineHD1600017107*	40	16	58664360_59498662	1,50
ARS-BFGL-NGS-12851_ARS-BFGL-NGS-3758*	40	22	40936789_42071416	1,48
ARS-BFGL-NGS-41958_BT-00954129*	40	27	11669357_13486135	1,38
BTB-01734642_BTA-41946-no-rs*	40	1	29520198_31171747	1,27
ARS-BFGL-NGS-78057_BT-01382552*	40	6	47353083_49140436	1,25
BovineHD1100005191_BovineHD1100005496*	40	11	16528054_17815368	1,14
ARS-BFGL-NGS-103036_BovineHD0400027828*	40	4	96201713_98497972	1,07
BovineHD1900016587_BovineHD1900017041*	40	19	57724102_59090292	1,04
%Total				28,32

*Janela em comum entre os modelos. Em negrito as janelas em comum com a avaliação da concentração de proteína no leite.

Os 17 marcadores que apresentaram variância superior a 1% estão localizados no cromossomo 1, 3, 4, 6, 8, 9, 11, 15, 19, 22, 25 e 27, e explicam, aproximadamente, 28,3% da variância aditiva, sendo a região dos SNPs ARS-BFGL-NGS-49079_ ARS-BFGL-NGS-34350 (posição no cromossomo 6450799_8405821 Mpb) com maior impacto sobre o fenótipo (explicou 3,10% da variância genética em ambos os modelos). O segundo maior foi a região dos SNPs BTA-34638-no-rs_ ARS-BFGL-NGS-39750 com variância explicada de 2,94%, seguido da região BovineHD1500009999_ BovineHD1500010339 com 2,21%. Essas duas regiões de QTL, também apresentaram importante impacto sobre a variância explicada para a proteína no leite.

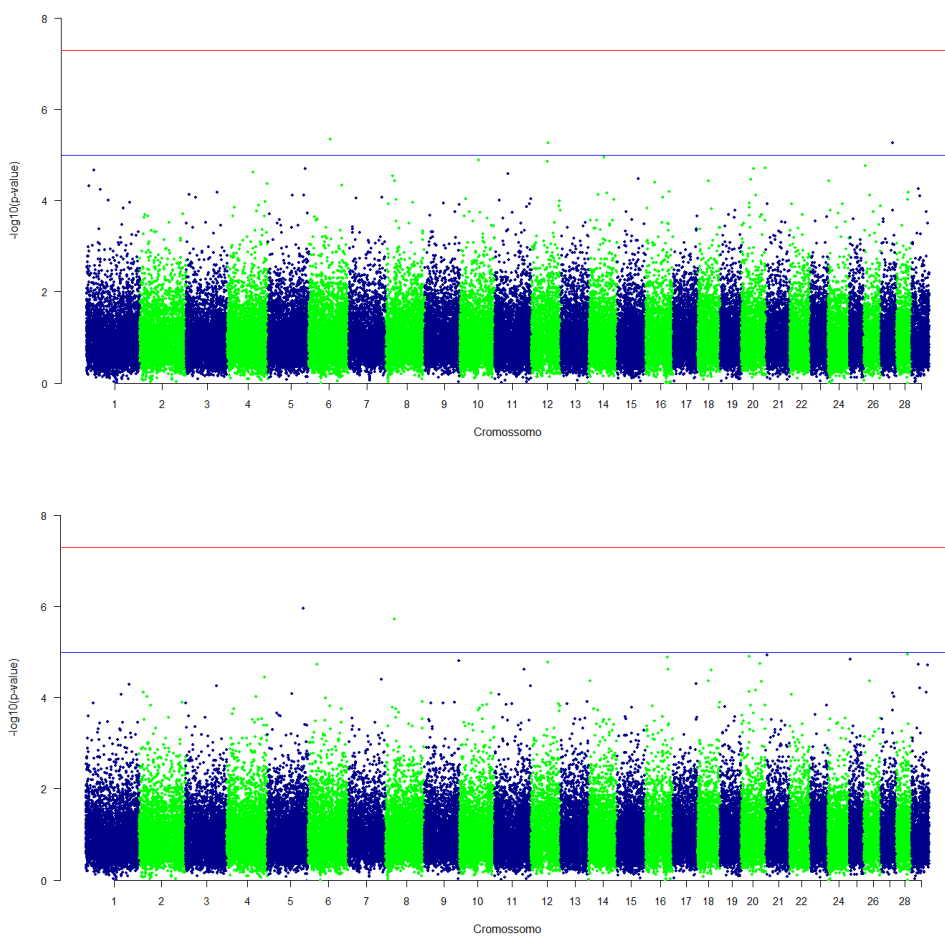


Figura 10. Gráfico do tipo Manhattan para a associação dos marcadores com a concentração de caseína no leite, com limites de $-\log_{10}(0,05)$ para o cromossomo (linha azul) e $-\log_{10}(0,08)$ a nível de genoma (linha vermelha), testados para o modelo aditivo (gráfico superior) e aditivo-dominante (gráfico inferior).

Esse achado é plausível considerando que a caseína é uma variante proteica, contudo, as regiões dos SNPs ARS-BFGL-NGS-49079 e BovineHD1500009999 parecem ter um impacto maior para a caseína, devido aos seus valores percentuais serem maiores comparado ao que foi encontrado para a concentração de proteína no leite (2,60 e 2,01%, respectivamente). Além disso, uma maior quantidade de regiões genômicas para a caseína explicaram mais de 1% da variância genética aditiva, contrapondo o que foi encontrado para a proteína.

Na análise de associação genômica, foi identificado que três SNPs atingiram o limite de significância a nível de cromossomo para o modelo aditivo e dois para o modelo aditivo-dominante (Figura 10).

Ácido Graxo Poli-insaturado

Houve uma variação considerável dentro e entre os modelos, referente às porcentagens da variância explicada pelas regiões formadas por 40 SNPs adjacentes (Tabela 14). Com a inclusão dos efeitos de dominância, 18 regiões apresentaram valores acima de 1% e explicaram cerca de 33,46%, um valor maior comparado ao modelo aditivo, na qual 17 regiões explicaram, aproximadamente, 31,29% da variância genética aditiva. Contudo, 13 regiões são comuns entre os modelos. Isso pode ser um indicativo de que essa região do genoma tem potencial para explorar a heterozigosidade para o conteúdo de ácido graxo poli-insaturado no leite (POLI).

Dentre as regiões com maior impacto sobre o fenótipo, os SNP iniciais BovineHD1400024339 no cromossomo 14 entre as janelas 146715 a 1681783 Mpb e o ARS-BFGL-NGS-18608 no cromossomo 9, posição 68260591 a 69659376 Mpb, também apresentaram impacto importante para a concentração de gordura no leite. Para POLI estas regiões explicaram cerca de 2,23% e 1,56% da variância genética no modelo aditivo, e 2,41 e 1,63% quando incluído dominância no modelo, não sendo os mais importantes para esse fenótipo. Diferente do que foi encontrado para a gordura no leite, na qual a região do SNP BovineHD1400024339 (cromossomo 14, janela 146715_1681783 Mpb) explicou 26,23% da variância genética aditiva. Nesse mesmo fenótipo, a região do SNP ARS-BFGL-NGS-18608 explicou uma proporção menor da variância genética (1,02% para os dois modelos), comparado ao ácido graxo poli-insaturado que explicou, aproximadamente, 1,56 e 1,63% da variância genética aditiva sob o modelo aditivo e aditivo-dominante, respectivamente.

Assim como para a proteína e a caseína, o ácido graxo poli-insaturado apresenta uma relação bioquímica e biológica com a concentração de gordura no leite. Dessa forma, é esperado que algumas regiões do genoma sejam comuns entre esses fenótipos e de que estes contribuam de alguma forma para a variância genética aditiva.

Foram encontrados cinco SNPs associados com o fenótipo a nível de todo o cromossomo, sendo três pelo modelo aditivo e dois com o modelo aditivo-dominante (Figura 11). No primeiro momento, foram significativos a $-\log_{10}(0,05)$ os SNPs 1314 (ARS-BFGL-NGS-43783) localizado no cromossomo 1 (posição 54347232 pb), 57885 (BovineHD2200015974) no cromossomo 22 (posição 55192606) e o marcador 61749 (BovineHD2500001118) localizado no cromossomo 25, na posição 4800735 pb. Ao incluir os

efeitos de dominância no modelo de GWAS adaptado, o SNP 15814 (Hapmap36729-SCAFFOLD106566_3245) no cromossomo 5 (posição 86395589 pb) e o SNP 28784 (ARS-BFGL-NGS-19753) localizado à 102980032 pb no cromossomo 9, foram associados a POLI. Assim como para algumas características avaliadas, aqui não foi relatado nenhuma associação do SNP com o fenótipo a nível de genoma.

Tabela 14. Identificação do SNP (inicial_final), tamanho da janela (quantidade de SNP), cromossomo (CHR), posição inicial e final da janela no cromossomo (POSinicial_POSfinal, Mpb) e porcentagem da variância genética aditiva explicada ($\% \sigma_a^2$) por janela de 40 marcadores adjacentes avaliados no modelo aditivo e aditivo-dominante, para o ácido graxo poli-insaturado.

Identificação do SNP	Janela	CHR	POS _{inicial} _POS _{final}	$\% \sigma_a^2$
Modelo aditivo				
BovineHD2600005813_ARS-BFGL-NGS-38894	40	26	22847961_24988516	3,89
BovineHD0300034950_BTBTB-01784003*	40	3	118896461_120546834	3,14
Hapmap47068-BTA-64117_BTBTB-02078366*	40	28	34912784_36299478	2,53
ARS-BFGL-NGS-119949_Hapmap38721-BTA-107281	40	15	80810699_81959353	2,52
BovineHD2900000986_BovineHD2900001270*	40	29	3439848_4457515	2,27
BovineHD1400024339_ARS-BFGL-NGS-22866*	40	14	146715_1681783	2,23
ARS-BFGL-NGS-116697_BovineHD1300002901*	40	13	9220838_10403183	1,81
ARS-BFGL-NGS-113401_BovineHD1900008474*	40	19	26735925_28206737	1,57
ARS-BFGL-NGS-18608_BTA-102674-no-rs*	40	9	68260591_69659376	1,56
Hapmap55168-rs29017467_Hapmap48168-BTA-106480*	40	22	40159906_41633675	1,48
ARS-BFGL-NGS-47808_BovineHD2000009942*	40	20	33569467_34603769	1,41
BovineHD1100011547_BovineHD1100012061*	40	11	39194600_41327004	1,35
ARS-BFGL-NGS-89598_BovineHD1800001965*	40	18	5010603_5936520	1,29
BovineHD1700015926_ARS-BFGL-NGS-783*	40	17	53928009_55324818	1,09
BovineHD1800003462_BovineHD1800003863*	40	18	9749447_10827317	1,09
ARS-BFGL-NGS-114558_ARS-BFGL-NGS-115817	40	3	48432235_49898786	1,05
ARS-BFGL-BAC-13009_BTA-23305-no-rs	40	11	42618575_43853640	1,01
%Total				31,29
Modelo aditivo-dominante				
Hapmap53405-rs29014382_BTBTB-00934644	40	26	22467951_24855737	4,07
BovineHD0300034950_BTBTB-01784003*	40	3	118896461_120546834	3,21
Hapmap47068-BTA-64117_BTBTB-02078366*	40	28	34912784_36299478	2,69
Hapmap38032-BTA-37826_BovineHD1500024523	40	15	81025239_82292544	2,46
BovineHD2900000986_BovineHD2900001270*	40	29	3439848_4457515	2,44
BovineHD1400024339_ARS-BFGL-NGS-22866*	40	14	146715_1681783	2,41
ARS-BFGL-NGS-116697_BovineHD1300002901*	40	13	9220838_10403183	2,01
ARS-BFGL-NGS-18608_BTA-102674-no-rs*	40	9	68260591_69659376	1,63
ARS-BFGL-NGS-113401_BovineHD1900008474*	40	19	26735925_28206737	1,61
Hapmap55168-rs29017467_Hapmap48168-BTA-106480*	40	22	40159906_41633675	1,54
ARS-BFGL-NGS-89598_BovineHD1800001965*	40	18	5010603_5936520	1,49
ARS-BFGL-NGS-47808_BovineHD2000009942*	40	20	33569467_34603769	1,45
BovineHD1100011547_BovineHD1100012061*	40	11	39194600_41327004	1,24
BovineHD1800003462_BovineHD1800003863*	40	18	9749447_10827317	1,11
BovineHD1700015926_ARS-BFGL-NGS-783*	40	17	53928009_55324818	1,05
BovineHD0300014907_BovineHD0300015402	40	3	48950219_50608425	1,04
BovineHD1300014256_BovineHD1300014593	40	13	48540388_50502256	1,01
BovineHD0500018417_BovineHD0500018755	40	5	65378999_66734512	1,00
%Total				33,46

*Janela em comum entre os modelos. Em negrito as janelas em comum com a concentração de gordura no leite.

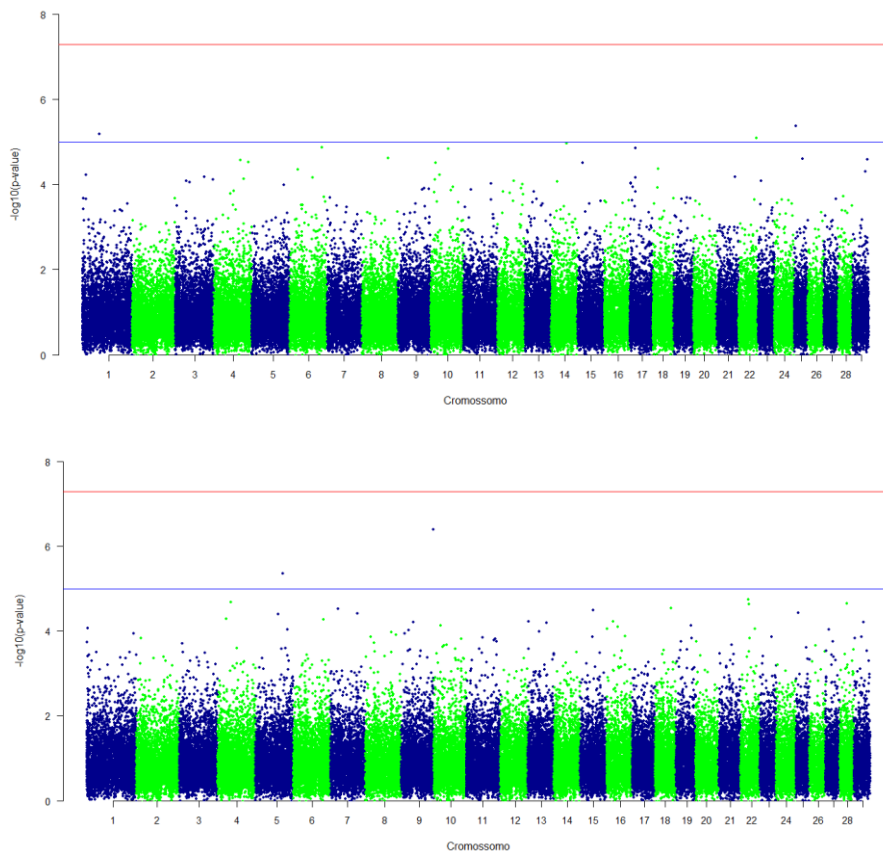


Figura 11. Gráfico do tipo Manhattan para a associação dos marcadores com o conteúdo de ácido graxo poli-insaturado no leite, com limites de $-\log_{10}(0,05)$ para o cromossomo (linha azul) e $-\log_{10}(0,08)$ a nível de genoma (linha vermelha), testados para o modelo aditivo (gráfico superior) e aditivo-dominante (gráfico inferior).

4.3.2. Genes adjacentes a região do SNP significativo

O total de oito genes foram identificados para oito SNPs associados com a produção de leite, gordura, proteína, caseína e ácido graxo poli-insaturado. A relação de genes e suas principais funções é apresentada na tabela 15. Os genes identificados desempenham função de estrutura celular e codificação de proteínas diversas. Nenhum deles apresenta função biológica direta com os fenótipos, contudo, para a proteína, o gene SLC2A13 atua de forma interativa nos processos de degradação de proteína e peptídeos e o gene NPHP1 no metabolismo, localização e transporte de diversas proteínas. Para a produção de leite, o gene RENF43 também desempenha função de metabolismos e de degradação de proteínas, além de participar de diversos processos regulatórios proteicos.

Para a gordura, foi identificado o gene GRM7 com função de composição lipídica de membrana celular de proteínas e que também pode atuar como receptor de glutamato, que é um aminoácido importante para o metabolismo humano, que participa de vias metabólicas, como gliconeogênese e glicólise que são os principais processos geradores de energia para o tecido, por meio da formação e quebra da glicose. Para caseína, o gene TACC1 também chamado de gene

transformados da proteína 1, atua na divisão e proliferação de células e é um sítio de ligação com receptores hormonais como glicocorticoides, que atuam no metabolismo de carboidratos e proteínas. Dentre os marcadores associados com o ácido graxo poli-insaturado, apenas o gene SOX5 foi identificado. Esse gene atua em funções associadas ao desenvolvimento de cartilagem e em alterações celulares ao estímulo do fator de crescimento transformador beta.

Tabela 15. Nome do SNP, cromossomo (CHR), posição do marcador no cromossomo (POS), nome do gene e sua principal função, para marcador molecular que atingiu limite de significância em cada característica avaliada.

Nome do SNP	CHR	POS	Gene	Principal função
Produção de leite				
BovineHD1900002699	19	9334455	<i>RNF43</i>	Gene de codificação de proteína, componente de membrana e está associado a proteína ubiquitina, com várias funções regulatórias e de degradação de outras proteínas.
Gordura				
BTB-00839323	22	19093982	<i>GRM7</i>	Receptor de glutamato; Componente lipídico de membrana de proteínas.
Proteína				
ARS-BFGL-NGS-85942	3	32529922	<i>CD53</i>	Componente de membrana com uma bicamada fosfolipídica.
BovineHD0500011755	5	40936260	<i>SLC2A13</i>	Atividade de transporte de solutos entre membranas e interação seletiva nos processos de proteases e peptidases.
BTB-00359112	8	75072384	<i>NFX1</i>	Interação entre íons de zinco e atividade de metabolismos de proteínas e ácidos nucleicos.
ARS-USMARC-Parent-DQ837646-rs29012894	11	1714947	<i>NPHP1</i>	Metabolismo, localização e transporte de proteínas.
Caseína				
BovineHD2700009481	27	33945088	<i>TACC1</i>	Gene transformador da proteína 1, ligado a atividades de divisão e proliferação celular, além de ser sítio de ligação com receptores hormonais como, estrogênio e glicocorticoides.
Ácido graxo poli-insaturado				
Hapmap36729-SCAFFOLD106566_3245	5	86395589	<i>SOX5</i>	Atua no processo do desenvolvimento ou extensão da produção de cartilagem, processo no qual o resultado específico é a progressão da cartilagem ao longo do tempo, desde sua formação até a estrutura madura. Atua no processo que resulta em uma mudança no estado ou atividade de uma célula como resultado de um estímulo do fator de crescimento transformador beta.

4.3.3. Heterozigose genômica e seus fatores – Simulação

O nível de heterozigose média do loco genômico foi calculado pela divisão da soma dos efeitos de dominância do QTN selecionado, pelo total de QTNs avaliados (Tabela 16). Neste contexto, os efeitos ou desvios de dominância negativos foram considerados no cálculo, por se tratar de uma fonte de informação referente à (des)vatagem heterozigótica. Em média, o nível de heterozigose molecular foi de 0,2%, com desvio-padrão de 0,07 (DP: $\pm 0,07$) e valor máximo de 22,8% para o cenário com magnitude de herdabilidade de 0,10. Para os cenários genéticos de maior magnitude de herdabilidade (h^2), os parâmetros médios foram maiores, sendo de 0,7% (DP: $\pm 0,18$) para h^2 de 0,30 (máximo de 66,4%) e de 0,6% (DP: $\pm 0,22$) para h^2 de 0,50 (máximo de 83,3%).

A heterozigosidade foi considerada como uma proporção de heterozigotos de cada indivíduo, isto é, a soma dos genótipos heterozigotos, dividido pelo total de loci marcadores avaliados. Em média, as populações simuladas para o cenário de $h^2 = 0,10$, apresentaram uma taxa de heterozigosidade média de 35,6% $\pm 1,00$. Níveis muito próximos foram encontrados para os demais cenários, na qual para $h^2 = 0,30$ a taxa média foi de 35,4% $\pm 0,5$ e para $h^2 = 0,50$ a média foi, aproximadamente, 35,7% $\pm 0,7$.

Tabela 16. Média (%), desvio-padrão (SD, %), valores de mínimo (MIN, %) e máximo (MAX, %) para o nível de heterozigose molecular, heterozigosidade e heterose genômica individual para os três cenários genéticos simulados para magnitude de herdabilidade (h^2) igual a 0,10, 0,30 e 0,50.

Parâmetro	Média	SD	MIN	MAX
$h^2 = 0,10$				
Heterozigose	0,2	7,0	-23,0	22,8
Heterozigosidade	35,6	1,0	32,8	37,6
Heterose genômica	3,2	6,0	-19,6	25,8
$h^2 = 0,30$				
Heterozigose	0,7	18,3	-64,1	66,4
Heterozigosidade	35,4	0,05	33,2	37,5
Heterose genômica	11,0	10,8	-27,1	49,5
$h^2 = 0,50$				
Heterozigose	0,6	22,1	-62,4	83,3
Heterozigosidade	35,7	0,07	33,1	37,9
Heterose genômica	17,0	13,9	-45,0	60,1

Nota: os valores apresentados nessa tabela foram calculados com base nos valores médios para 20 repetições, em cada cenário simulado.

Para a heterose genômica, o cenário de $h^2 = 0,50$ apresentou os maiores valores médios, comparado aos demais. Neste, a média foi de 17,0% (DP: $\pm 13,9\%$), com máxima de 60,1%. Em seguida, para $h^2 = 0,30$ a média foi aproximadamente 11,0% (DP: $\pm 10,8\%$) e valor máximo de

49,5%. Menores taxas foram encontradas para o cenário de menor magnitude de herdabilidade ($h^2 = 0,10$), na qual a heterose máxima foi de 25,8% e média populacional de 3,2% (DP: $\pm 6,0\%$).

Coefficientes de regressão da heterose genômica em heterozigidade

Na tabela 17 é apresentado os coeficientes de regressão para o nível de heterozigidade em função da heterose genômica. Aqui, é tratado como uma população cada ciclo (repetição) de simulação. Para o cenário I ($h^2 = 0,10$) as estimativas de regressão foram em direção favorável para a maioria das populações, ou seja, a direção para a qual é esperado para que a mudança de uma unidade de heterozigidade molecular, possa alterar positivamente uma unidade de heterose genômica. Em I, algumas estimativas foram substâncias. Por exemplo, na população 5, foi estimada a redução proporcional de 1% de heterose, em relação ao aumento de 1% de heterozigidade. Para a população 7 o oposto é verdadeiro, na qual o aumento em unidade de heterozigidade, estima-se o aumento em 1,34 unidades de heterose. Para as populações 8, 9, 11 e 14, as estimativas também foram positivas (0,59, 0,97, 0,52 e 0,55, respectivamente), assim como para a população 1, no entanto com um coeficiente de regressão estimado de menor valor (0,20). Nas demais populações, as estimativas foram negativas ou não significativamente diferentes de zero ($P > 0,05$).

Nos cenários II e III ($h^2 = 0,30$ e $0,50$, respectivamente), foi obtida uma maior proporção de estimativas favoráveis e com maior magnitude, comparado ao cenário I. Em II, o maior coeficiente de regressão estimado foi de 3,04 (população 19), seguido de 2,05 para população 9; 1,68 para 7; 1,63 para 17; 1,45 e 1,05 para as populações 11 e 20, respectivamente. No cenário III, a maior estimativa foi de 2,92 na população 20, seguido de 2,22; 2,20 e 1,00 para as populações 9, 7 e 16, respectivamente. Ao avaliar o conjunto de populações para os três cenários, notou-se uma inversão no sentido de algumas estimativas. Por exemplo, para a população 12 às estimativas foram de -0,80 para o cenário II e de 0,60 para o cenário III, ambas com coeficiente de regressão significativo. Esse comportamento foi semelhante para as populações 13, 14 e 18.

Ao dividir as estimativas dos coeficientes de regressão pelo desvio-padrão da heterozigidade, é possível obter uma indicação da importância da heterozigidade entre as populações (Tabela 17). Isso indicou que a heterozigidade teve maior efeito nas populações: a) 7 (4,80), 9 (3,47), 14 (2,10), 8 (2,09); 11 (1,99) e 10 (1,03) pertencentes ao cenário I; b) 19 (6,07), 9 (4,13), 7 (3,53), 17 (3,25), 11 (2,91), 20 (2,43), 1 (1,64), 13 (1,61), 16 (1,43), 10 (1,04) para o cenário II; c) 20 (4,69), 7 (3,54), 3 (3,43), 14 (3,01), 8 (2,09), 16 (1,63), 10 (1,41), 12 (1,10) e 1 (1,18) para o cenário III. Para as demais, não houve vantagem da heterozigidade e para aquelas populações com efeito positivo, a estimativa do coeficiente de regressão não foi significativo.

As estimativas de regressão diferiram entre as populações de cada cenário, e entre os cenários os coeficientes foram semelhantes em módulo quanto ao sentido da estimativa de regressão. Para uma situação na qual a magnitude da herdabilidade foi alta, as estimativas foram superiores comparado ao parâmetro de herdabilidade de magnitude baixa a moderada.

Tabela 17. Coeficientes de regressão e erro-padrão para heterose genômica (Het_r) em função do nível de heterozigidade individual (Het_z), para 20 ciclos de simulação para herdabilidades (h^2) de 0,10, 0,30 e 0,50.

População	$E(\text{Het}_r/\text{Het}_z) = \alpha + \beta\text{Het}_z$														
	$h^2 = 0,10$					$h^2 = 0,30$					$h^2 = 0,50$				
	β	EP	α	$\sigma_{\text{Het}_r}^1$	$\beta/\sigma_{\text{Het}_r}^2$	β	EP	α	$\sigma_{\text{Het}_r}^1$	$\beta/\sigma_{\text{Het}_r}^2$	β	EP	α	$\sigma_{\text{Het}_r}^1$	$\beta/\sigma_{\text{Het}_r}^2$
1	0,20*	0,095	-0,06	0,27	0,71	0,85*	0,178	0,08	0,52	1,64	0,75*	0,223	-0,13	0,64	1,18
2	-0,69*	0,095	0,08	0,27	-2,56	-0,91*	0,165	0,05	0,48	-1,91	-1,58*	0,219	0,25	0,63	-2,53
3	0,06	0,093	0,10	0,27	0,22	-0,26	0,148	0,00	0,43	-0,61	-0,07	0,206	-0,04	0,59	-0,12
4	-0,13	0,106	0,00	0,30	-0,42	-1,22*	0,185	-0,05	0,53	-2,31	-0,43	0,238	0,22	0,68	-0,63
5	-1,03*	0,091	0,03	0,26	-3,97	-2,38*	0,158	0,07	0,46	-5,17	-3,23*	0,196	0,13	0,57	-5,70
6	0,17	0,099	0,03	0,29	0,57	-0,68*	0,180	0,30	0,50	-1,35	-0,14	0,220	0,05	0,63	-0,23
7	1,34*	0,097	-0,02	0,28	4,80	1,69*	0,162	0,05	0,48	3,53	2,20*	0,217	0,03	0,62	3,54
8	0,59*	0,097	0,15	0,28	2,09	0,43*	0,168	-0,07	0,48	0,89	1,32*	0,218	0,03	0,63	2,09
9	0,97*	0,096	0,12	0,28	3,47	2,05*	0,175	-0,03	0,50	4,13	2,22*	0,222	0,10	0,65	3,43
10	0,26*	0,086	0,08	0,25	1,03	0,48*	0,165	0,03	0,46	1,04	0,81*	0,198	0,09	0,58	1,41
11	0,52*	0,091	-0,11	0,26	1,99	1,45*	0,173	-0,27	0,50	2,91	0,56*	0,202	-0,07	0,60	0,95
12	0,01	0,087	0,10	0,25	0,02	-0,80*	0,167	0,15	0,48	-1,67	0,60*	0,194	-0,18	0,54	1,10
13	-0,04	0,096	-0,12	0,28	-0,15	0,75*	0,162	-0,11	0,47	1,61	-1,59*	0,060	0,09	0,72	-2,21
14	0,55*	0,091	0,04	0,26	2,10	-0,57*	0,181	0,18	0,51	-1,11	1,94*	0,224	-0,04	0,65	3,01
15	0,05	0,091	0,00	0,26	0,18	-2,85*	0,159	0,08	0,46	-6,19	-1,24*	0,220	-0,11	0,62	-1,99
16	-0,22*	0,091	0,06	0,27	-0,80	0,69*	0,166	-0,12	0,48	1,43	1,00*	0,217	-0,17	0,62	1,63
17	-0,98*	0,084	0,07	0,24	-4,09	1,63*	0,172	0,02	0,50	3,25	-0,43	0,227	0,21	0,65	-0,66
18	0,10	0,089	-0,12	0,25	0,38	0,84*	0,162	0,03	0,46	1,82	-2,51*	0,204	-0,14	0,59	-4,28
19	-0,82*	0,097	-0,04	0,29	-2,99	3,04*	0,173	-0,08	0,50	6,07	0,33	0,220	-0,07	0,64	0,52
20	0,14	0,088	-0,10	0,25	0,55	1,05*	0,151	-0,03	0,43	2,43	2,92*	0,219	0,14	0,62	4,69

Hetr: heterose genômica; **Hetz**: heterozigidade; **h²**: parâmetro de herdabilidade; **β** : estimativa; **EP**: erro-padrão; **α** : intercepto; ¹Desvio-padrão da heterose; ²Coefficiente de regressão dividido pelo desvio-padrão da heterose genômica; *Coefficiente de regressão significativamente diferente de zero ($P > 0,05$).

Coeficiente de regressão do fenótipo em função dos fatores de heterozigidade

Para um modelo genético aditivo-dominante, os fenótipos foram simulados como sendo a soma do valor genético verdadeiro calculado para QTNs de efeito aditivo e de desvio de dominância, associados a um vetor de valores residuais. Com o intuito de verificar possíveis relações entre esses fenótipos e os fatores heterozigóticos, os coeficientes de regressão do fenótipo sob a heterozigidade e, do fenótipo em função da heterose genômica são apresentados nas tabelas 18 e 19, respectivamente.

No cenário com herdabilidade baixa (I, $h^2 = 0,10$), os coeficientes de regressão do fenótipo em função da heterozigidade, foram favoráveis para as populações 7 e 11 (1,39 e 1,04, respectivamente) e para as populações 1, 5, 8, 17 e 20 as estimativas foram negativas. Para as outras populações, as estimativas não foram significativas para $P > 0,05$. Nos cenários II ($h^2 = 0,30$) e III ($h^2 = 0,50$), foram encontradas estimativas positivas e maiores, comparado ao cenário

I. Em II, o maior coeficiente estimado foi para a população 9 (3,85), seguido de 1 (1,70), 7 (1,65), 12 (1,08), 14 (0,89), 17 (0,74) e 18 (0,51). Para a população 5 e 6, os coeficientes foram altos, porém negativos (-3,12 e -1,35).

Tabela 18. Coeficientes de regressão e erro-padrão para o fenótipo (fen) em função do nível de heterozigidade individual (Hetz), para 20 ciclos de simulação com magnitude de herdabilidade (h^2) 10, 30 e 50%.

$$E(\text{fen}/\text{Hetz}) = a + \beta\text{Hetz}_i$$

População	$h^2 = 0,10$					$h^2 = 0,30$					$h^2 = 0,50$				
	β	EP	α	σ_{fen}^1	β/σ_{fen}^2	β	EP	α	σ_{fen}^1	β/σ_{fen}^2	β	EP	α	σ_{fen}^1	β/σ_{fen}^2
1	-1,19*	0,489	1,33	1,40	-0,85	1,70*	0,476	1,62	1,39	1,22	0,67	0,454	2,08	1,31	0,52
2	-0,27	0,478	0,750	1,36	-0,20	-0,64*	0,467	-0,83	1,34	-0,47	-0,77	0,458	-1,89	1,31	-0,59
3	-0,89	0,459	0,87	1,31	-0,68	0,06	0,431	0,12	1,25	0,04	0,35	0,483	-0,24	1,39	0,25
4	-0,55	0,467	0,56	1,33	-0,41	-1,05*	0,489	0,21	1,39	-0,75	-0,81	0,501	0,82	1,43	-0,56
5	-0,86*	0,436	0,67	1,25	-0,69	-3,12*	0,445	-1,03	1,29	-2,41	-3,56*	0,445	-1,10	1,28	-2,78
6	0,81	0,470	0,67	1,36	0,60	-1,35*	0,510	0,30	1,43	-0,94	-0,17	0,491	-1,62	1,40	-0,12
7	1,39*	0,462	-0,04	1,32	1,05	1,65*	0,441	2,35	1,30	1,27	2,81*	0,464	2,60	1,33	2,11
8	-0,99*	0,464	0,67	1,34	-0,74	0,05	0,479	0,15	1,38	0,04	2,57*	0,464	-0,61	1,35	1,91
9	0,53	0,455	1,08	1,32	0,40	3,85*	0,471	1,10	1,33	2,89	1,96*	0,451	2,62	1,31	1,49
10	-0,15	0,413	0,21	1,18	-0,13	0,09	0,470	1,70	1,31	0,07	1,31*	0,426	1,58	1,24	1,05
11	1,04*	0,460	-0,67	1,31	0,80	0,05	0,473	1,38	1,36	0,03	1,64*	0,436	-1,05	1,29	1,28
12	0,35	0,440	1,53	1,25	0,28	1,08*	0,480	-1,49	1,38	0,78	2,21*	0,457	0,08	1,28	1,72
13	-0,26	0,473	0,85	1,36	-0,19	0,04	0,470	1,04	1,35	0,03	0,08	0,120	-1,15	1,35	0,06
14	0,14	0,437	0,86	1,25	0,11	0,89*	0,492	-0,05	1,39	0,64	0,62	0,484	2,04	1,39	0,44
15	-0,10	0,445	0,41	1,29	-0,08	-0,29	0,429	-0,64	1,23	-0,23	1,09*	0,499	-2,13	1,41	0,78
16	-0,55	0,434	0,54	1,29	-0,42	-0,70*	0,455	-0,28	1,32	-0,53	-0,46	0,484	0,43	1,37	-0,34
17	-1,40*	0,444	0,17	1,26	-1,11	0,74*	0,446	3,29	1,29	0,57	-0,39	0,484	-0,94	1,39	-0,28
18	-0,27	0,424	-0,34	1,21	-0,22	0,51*	0,469	-0,04	1,35	0,37	-0,68	0,446	-1,43	1,28	-0,53
19	-0,62	0,486	-0,59	1,43	-0,43	0,19	0,482	2,83	1,38	0,14	0,20	0,462	1,11	1,34	0,15
20	-0,96*	0,454	-0,47	1,31	-0,73	-0,21	0,425	1,74	1,22	-0,17	0,66	0,443	4,31	1,25	0,52

fen: fenótipo; Hetz: heterozigidade; h^2 : parâmetro de herdabilidade; β : estimativa; EP: erro-padrão; α : intercepto. ¹Desvio-padrão fenotípico; ²Coefficiente de regressão dividido pelo desvio-padrão fenotípico; *Coefficiente de regressão significativamente diferente de zero ($P > 0,05$).

No cenário III, valores de estimativa superiores e com sentido favorável para a heterozigidade foram encontrados para as populações 7, 8, 9, 10, 11, 12 e 15, que variaram de 1,09 a 2,81. No mais, salvo a população 5 (-3,56), as estimativas não foram significativas. Entre os três cenários, nota-se que para herdabilidade de magnitude moderada à alta, as estimativas foram maiores, comparado ao cenário de menor herdabilidade. Além disso, quando o coeficiente de regressão foi alto e positivo, a heterozigidade teve maior efeito nos fenótipos simulados.

Para algumas populações, percebe-se um comportamento crescente entre os coeficientes de regressão estimados. Para a população 7, por exemplo, a estimativa variou de 1,39 a 2,81, uma diferença substancial de 1,42 unidades entre os cenários I e III. Da mesma forma, o efeito da heterozigidade sob o fenótipo, foi 1,06 unidades maior no cenário III, em relação ao I. Um

fato curioso foi observado na população 9, na qual para um cenário de $h^2 = 0,30$, a estimativa foi maior que para o cenário de alta herdabilidade. Para esse caso, a diferença é maior ainda do que encontrado anteriormente, 1,96 vezes maior no cenário II ($\beta = 1,96$), comparado ao III ($\beta = 3,85$).

Adicionalmente, para as populações na qual, o aumento em unidade de heterozigidade causaria decréscimo em unidade do fenótipo, o efeito da heterozigidade foi negativo, isto é, nestas situações é possível que a ocorrência de uma desvantagem quanto ao fator de heterozigidade, independentemente da magnitude da herdabilidade.

A respeito dos coeficientes de regressão do fenótipo em heterose genômica (Tabela 19), as estimativas foram maiores que 0,90 para praticamente todas as populações, nos três cenários avaliados. Salvo exceções na qual os coeficientes foram negativos ou não significativos, o efeito da heterose foi muito semelhante entre os cenários e menores comparado ao fator de heterozigidade.

Tabela 19. Coeficientes de regressão e erro-padrão para o fenótipo (fen) em função da heterose genômica (Hetr), para 20 repetições de simulação com magnitude de herdabilidade (h^2) 10, 30 e 50%.

População	$E(\text{fen}/\text{Hetr}) = a + \beta\text{Hetr}_i$														
	$h^2 = 0,10$					$h^2 = 0,30$					$h^2 = 0,50$				
	β	EP	a	σ_{fen}^1	$\beta/\sigma_{\text{fen}}^2$	β	EP	a	σ_{fen}^1	$\beta/\sigma_{\text{fen}}^2$	β	EP	a	σ_{fen}^1	$\beta/\sigma_{\text{fen}}^2$
1	1,01*	0,039	0,91	1,40	0,72	0,98*	0,019	1,85	1,40	0,71	0,91*	0,014	2,20	1,31	0,70
2	0,97*	0,038	0,81	1,36	0,71	0,95*	0,021	-0,80	1,36	0,71	0,93*	0,015	-1,87	1,31	0,71
3	1,07*	0,037	0,44	1,31	0,82	1,03*	0,021	0,24	1,31	0,82	1,11*	0,016	-0,04	1,39	0,80
4	0,99*	0,034	0,40	1,33	0,74	0,99*	0,019	0,32	1,33	0,71	1,03*	0,014	0,47	1,43	0,72
5	1,04*	0,036	0,71	1,25	0,83	0,99*	0,020	-1,36	1,25	0,77	0,98*	0,016	-2,26	1,28	0,77
6	1,06*	0,036	0,87	1,36	0,78	1,09*	0,020	-0,24	1,36	0,76	1,01*	0,016	-1,67	1,40	0,72
7	0,91*	0,036	0,03	1,32	0,69	0,95*	0,020	2,33	1,32	0,73	1,01*	0,015	2,76	1,33	0,76
8	1,01*	0,036	-0,03	1,34	0,75	1,08*	0,021	0,07	1,34	0,79	1,01*	0,015	-0,20	1,35	0,75
9	0,98*	0,036	0,81	1,32	0,74	1,01*	0,019	1,75	1,32	0,76	0,97*	0,014	2,46	1,31	0,74
10	0,99*	0,036	-0,01	1,18	0,84	1,06*	0,021	1,52	1,18	0,81	1,03*	0,015	1,66	1,24	0,83
11	1,00*	0,038	-0,38	1,31	0,77	-0,03	0,021	1,40	1,31	-0,02	0,99*	0,015	-0,58	1,29	0,77
12	0,98*	0,039	1,56	1,25	0,78	0,06*	0,022	-1,11	1,25	0,04	1,01*	0,017	0,82	1,28	0,79
13	1,00*	0,037	0,89	1,36	0,74	0,04	0,023	1,05	1,36	0,03	0,01	0,015	-1,11	1,35	0,01
14	0,95*	0,036	0,68	1,25	0,76	-0,02	0,021	0,26	1,25	-0,02	0,01	0,017	2,26	1,39	0,00
15	1,00*	0,037	0,36	1,29	0,78	0,03	0,021	-0,71	1,29	0,03	0,00	0,018	-1,74	1,41	0,00
16	1,00*	0,036	0,36	1,29	0,78	-0,02	0,021	-0,54	1,29	-0,01	0,01	0,017	0,26	1,37	0,01
17	0,99*	0,040	-0,05	1,26	0,79	0,00	0,020	3,56	1,26	0,00	-0,01	0,017	-1,08	1,39	-0,01
18	-0,05	0,037	-0,44	1,21	-0,04	0,03	0,023	0,13	1,21	0,02	-0,03	0,017	-1,71	1,28	-0,02
19	0,99*	0,038	-0,47	1,43	0,70	0,00	0,021	2,91	1,43	0,00	-0,04*	0,016	1,18	1,34	-0,03
20	1,02*	0,040	-0,77	1,31	0,78	-0,02	0,022	1,68	1,31	-0,02	0,03	0,016	4,51	1,25	0,02

fen: fenótipo; **Hetr:** heterose genômica; **h^2 :** parâmetro de herdabilidade; **β :** estimativa; **EP:** erro-padrão; **a :** intercepto; ¹Desvio-padrão fenotípico; ²Coefficiente de regressão dividido pelo desvio-padrão fenotípico; *Coeficiente de regressão significativamente diferente de zero ($P > 0,05$).

No cenário de herdabilidade 0,10, as estimativas favoráveis variaram de 0,97 a 1,07, com efeito da heterose sob o fenótipo variando de 0,70 a 0,84. Variações muito próximas foram encontradas para o cenário III (0,99 a 1,11) e valores menores para o cenário II (0,06 a 1,08). No entanto, a proporção de

populações com estimativas desfavoráveis e não significativas, são maiores nos cenários na qual o valor de herdabilidade é moderado e alto.

Os fatores de heterozigose contribuíram de forma importante para o fenótipo simulado. De alguma forma, o nível de heterozigosidade se apresentou mais relevante, comparada à heterose. Se tratando de um estudo empírico para animais de raça pura, os resultados revelaram que a seleção dos indivíduos por meio da heterozigosidade avaliada no loco genômico seja preferível, se o objetivo for buscar animais para gerar maior e melhor heterose molecular.

4.3.4. Exploração da heterozigosidade no genoma real

Na tabela 20 são apresentados a estatística descritiva e o nível do fator de heterozigosidade nas regiões genômicas que explicaram mais de 1% da variância genética, para cada característica. O nível de heterozigosidade do loco genômico foi calculado como uma proporção de genótipos de heterozigotos para cada de QTLs, dividido pelo total de QTLs avaliados em cada característica. Para gordura, aproximadamente 50,5% da heterozigosidade média foi encontrada em 10 QTLs de maior impacto sobre a variância genética. Em seguida, a concentração de caseína com proporção de heterozigotos de 50,4% em 17 QTLs; ácido graxo POLI com 50,3% em 22 QTLs, sendo essa a característica com maior número de regiões genômicas identificadas pelo GWAS adaptado; proteína com 50,1% em oito QTLs; e a menor proporção para a produção de leite, com aproximadamente 49,4% de heterozigose, estimada para 16 QTLs.

Adicionalmente, foi estimada a heterozigosidade para cada indivíduo e obtido a média desse fator para a população avaliada (Tabela 21). Para isso, foi realizada a contagem de genótipos heterozigotos das regiões genômicas e dividido pelo total de marcadores moleculares avaliados. As taxas variaram de 0 a 1, sendo que para gordura e proteína houve indivíduos com taxa de heterozigosidade máxima (100%). Para as demais características a variação foi de 0 a 0,875 (0 a 87,5%), uma variação menor em relação a gordura e proteína. Em média, a maior proporção de genótipos heterozigotos foi para a proteína, com aproximadamente 37,9% de heterozigosidade na população, seguido da produção de leite (37,7%), ácido graxo (37,3%), gordura (36,9%) e caseína (35,3%).

Os níveis médios de heterozigose do loco genômico e a heterozigosidade baseada nos marcadores dos locos genômicos, foram uniformes entre as características e relativamente altas considerando uma população de indivíduos de raça pura. Isso pode ser um indicativo de que as regiões identificadas pelo GWAS adaptado possuem taxas de heterozigose superiores e com uma possível vantagem heterozigótica para os fenótipos, comparada as regiões do genoma que não atingiram o limite de significância pré-estabelecido.

Tabela 20. Número de animais avaliados, número de regiões genômicas que explicaram mais de 1% da variância genética aditiva, valores de mínimo (MIN), máximo (MAX) e média para os fatores de heteroziguidade calculado no loco genômico e individualmente para cada animal, para produção de leite (PL, kg), porcentagem de gordura (GORD, %), proteína (PROT, %), caseína (CAS, %) e conteúdo de ácido graxo poli-insaturado (POLI, %).

Heteroziguidade no loco						
Característica	Número de animais	Número de regiões	MIN	MAX	Média	
PL (kg/dia)	1092	16	0,467	0,550	0,494	
GORD (%)	1092	10	0,439	0,585	0,505	
PROT (%)	1092	8	0,486	0,541	0,501	
CAS (%)	951	17	0,445	0,583	0,504	
POLI (%)	1089	22	0,470	0,568	0,503	
Heteroziguidade individual						
Característica	Número de animais	Número de regiões	MIN	MAX	Média	
PL (kg/dia)	1092	16	0	0,875	0,377	
GORD (%)	1092	10	0	1	0,369	
PROT (%)	1092	8	0	1	0,379	
CAS (%)	951	17	0	0,882	0,353	
POLI (%)	1089	22	0	0,864	0,373	

Para melhor compreender as possíveis ações da heteroziguidade individual sob o fenótipo, foram obtidos coeficientes de regressão entre o fenótipo ajustado e o nível de heteroziguidade individual (Tabela 21). As estimativas de coeficiente de regressão em heteroziguidade estavam em direção favorável para a produção de leite ($1,9642 \pm 0,3376$), proteína ($0,0026 \pm 0,0092$) e para ácido graxo poli-insaturado ($0,0059 \pm 0,0021$). Para as demais características, as estimativas indicaram que o aumento em heteroziguidade poderá reduzir em algumas unidades a concentração de gordura ($-0,0048 \pm 0,0256$) e caseína ($-0,0094 \pm 0,0115$).

Tabela 21. Coeficientes de regressão para a heteroziguidade (Hetz) em função do fenótipo ajustado (y^*) para a produção de leite (PL, kg/dia), gordura (GORD, %), proteína (PROT, %), caseína (CAS, %) e conteúdo de ácido graxo poli-insaturado (POLI, %).

Características	$E(y^*/Hetz) = a + \beta Hetz_i$				
	β	EP	α	$\sigma_{y^*}^1$	$\beta/\sigma_{y^*}^2$
PL (kg/dia)	1,9642*	0,3376	31,3916	7,7663	0,2530
GORD (%)	-0,0048	0,0265	3,4958	0,6922	-0,0007
PROT (%)	0,0026	0,0092	3,1269	0,2521	0,0103
CAS (%)	-0,0094	0,0115	2,3875	0,2280	-0,0412
POLI (%)	0,0059*	0,0021	0,1559	0,0404	0,1460

Hetz: heteroziguidade; β : estimativa; **EP:** erro-padrão; α : intercepto; ¹Desvio-padrão do fenótipo ajustado (y^*); ²Coefficiente de regressão dividido pelo desvio-padrão do fenótipo ajustado; *Coeficiente de regressão significativamente diferente de zero ($P > 0,05$).

Estimativas substâncias foram encontradas para a produção de leite. Por exemplo, é previsto que o aumento em 1% da heteroziguidade, cause aumento em 1,94 kg de leite por dia. Assim também deve acontecer para o ácido graxo, mas em menor proporção, sendo o aumento em unidade de heteroziguidade poderá resultar no aumento de 0,0059% de conteúdo de ácido graxo poli-insaturado. Para as outras características, os coeficientes não foram significativamente diferentes de zero ($P\text{-value} < 0,05$).

Ao dividir o coeficiente de regressão estimado pelo desvio-padrão do fenótipo ajustado, observa-se que o nível de heteroziguidade teve maior impacto sobre o volume de produção de leite e para o conteúdo de ácido graxo poli-insaturado. Para a porcentagem de proteína também foi encontrado efeito de heteroziguidade relativamente importante, mas a estimativa de regressão não foi significativa.

4.4. DISCUSSÃO

Neste capítulo foi proposto explorar a heteroziguidade molecular e seus fatores, como um possível critério na seleção de animais em populações de raça pura, para gerar maior e melhor heterose. Inicialmente, foi utilizado a simulação de dados para avaliar as relações entre os fatores de heteroziguidade em diferentes cenários genéticos. O modelo de dominância de Soresen et al. (2008) define o grau de heterose de uma raça oriunda de cruzamento, a chance de um loco genômico específico possuir genes das diferentes raças utilizadas para o cruzamento. Por sua vez, isto é definido como a heteroziguidade da raça, calculado como a proporção de contribuição genética ou heterozigótica das linhagens parentais. Portanto, em cruzamentos, a heterose da população F_1 é considerada como sendo os efeitos devido a dominância. Neste estudo, a heterose genômica foi calculada com base nos desvios de dominância simulados e, quando realizada a regressão entre a proporção de genótipos heterozigotos em heterose, foram encontradas estimativas de coeficientes positivos e significativos, independente do cenário (Tabela 17).

Para algumas situações, ficou claro que o aumento da heteroziguidade poderá promover aumentos positivos e substâncias em heterose genômica. Nestes casos, é possível que haja indivíduos com maior afinidade genética para gerar maior e melhor heterose, se selecionados para os acasalamentos. Além disso, os cenários de herdabilidade baixa e moderada apresentaram as maiores estimativas de regressão e os maiores valores de importância da heteroziguidade para explicar a heterose genômica. Isto pode indicar que, para características complexas, a heteroziguidade individual deverá ser o fator de maior impacto sobre a heterose.

Nas tabelas 18 e 19, observa-se que para os cenários de menor magnitude de herdabilidade, as estimativas de regressão do fenótipo em heterose genômica foram favoráveis e

significativas em maior quantidade, comparado ao cenário de maior herdabilidade. A hipótese de que a arquitetura genética do fenótipo influenciará nos fatores de heterozigosidade, concorda com outros estudos (Kargo et al., 2021; Saborío-Montero et al., 2018; Clasen et al., 2017). Além disso, foram poucas as estimativas de regressão significativas e favoráveis do fenótipo em função da heterozigosidade em todos os cenários, não sendo possível atribuir grandes mudanças no fenótipo, apenas considerando a heterozigosidade. Contudo, o fator de heterozigosidade foi importante para o fenótipo, principalmente no cenário com maior magnitude de herdabilidade.

Para identificar regiões do genoma com variações não-aditivas, uma análise de GWAS foi adaptada para os efeitos heterozigóticos. Por meio de uma análise de regressão linear do fenótipo e da heterozigosidade estimada em algumas regiões genômicas, foram encontrados coeficientes de regressão favoráveis e significativas para a produção de leite e o conteúdo de ácido graxo poli-insaturado, com expectativa de que, o aumento em heterozigosidade, poderá aumentar nos níveis de produção destas características (Tabela 21). Está hipótese concorda com estudos anteriores, que relataram vantagens ao incluir efeitos de dominância e heterozigosidade em programas de acasalamento em diferentes espécies de animais (Alillo et al., 2017; Gutiérrez-Gil et al., 2014; Bagheri et al., 2013; Tambasco-Talhari et al., 2005). Em média, a heterozigosidade foi entre 0,353 e 0,377, e concordam com as estimativas médias de 0,37 no estudo de Keller et al (2017), em populações de vacas da raça holandesa.

Estudos anteriores retrataram a base molecular da heterose, correlacionada com a heterozigosidade do loco genômico e com forte associação com os efeitos de dominância e sobredominância (Tambasco-Talhari et al., 2005; Xiao et al., 1995). Ao contabilizar uma matriz de relacionamento genômico para os efeitos de dominância no modelo de GWAS, foram identificadas algumas regiões extras, ao invés de um modelo somente aditivo. É possível que as associações entre heterozigosidade e fenótipo não sejam devido à dominância de um único loco, pois os homozigotos para o alelo dominante poderiam compartilhar o mesmo fenótipo para os heterozigotos (Deng e Fu, 1998). Dessa forma, se vários loci estivessem envolvidos na expressão superior dos heterozigotos, a correlação entre heterozigosidade e fenótipo existiria (Tambasco-Talhari et al., 2005). De acordo com os resultados deste estudo, sugere-se que a seleção baseada nas regiões genômicas seria o mais adequado e, portanto, uma possível estratégia para direcionar os acasalamentos entre os indivíduos com potencial para gerar heterose genômica favorável em sistemas de criação de animais de raças puras.

Não há estudos diretamente comparáveis a este, mas os resultados que foram relatados devem circundar a possível ocorrência de heterose em nível molecular, em populações resultantes de acasalamentos entre indivíduos da mesma raça. Além disso, a heterozigosidade avaliado no

loco genômico deve ser um parâmetro adequado se o objetivo dos acasalamentos for à heterose, em características de produção e qualidade do leite, pois foi relatado como um fator importante para explicar o fenótipo e a ocorrência de heterose genômica.

Adicionalmente, por meio do GWAS adaptado foram identificados marcadores moleculares do tipo SNP associados às características de produção e qualidade do leite. Entre os modelos testados, ao contabilizar os efeitos de dominância uma maior quantidade de SNPs foram associados ao fenótipo, comparado ao modelo aditivo. A análise de anotação de genes foi realizada apenas para as regiões genômicas destes SNPs e, dessa forma, foi gerada uma lista com oito genes adjacentes à região destes marcadores. A maioria dos genes identificados pouco se relacionam funcionalmente e biologicamente com a característica de interesse, sendo as principais funções como constituintes de membrana celular, codificação de proteínas, transporte de solutos e receptores hormonais.

No entanto, para o ácido graxo poli-insaturado o gene *SOX5* possui uma função importante no processo de desenvolvimento e extensão da produção de cartilagem, resultante do estímulo de um fator de crescimento transformador beta. Em mamíferos, principalmente em bovinos leiteiros, esse fator de crescimento quando presente no leite tem a função de estimular o crescimento de células especialmente do tecido conectivo, além de participar da formação de ossos e cartilagens, sendo o foco em estudos voltados a saúde humana (Gauthier et al., 2006; Silva et al., 2016). Fatores intimamente ligados a saúde do consumidor são as principais motivações nas avaliações do conteúdo de ácido graxos no leite e este estudo deve ser o primeiro a registrar o gene *SOX5* para o ácido graxo poli-insaturado em bovinos leiteiros, em uma região com possível vantagem heterozigótica.

4.5. CONCLUSÃO

A heterozigosidade avaliada de forma empírica e por meio de dados reais revelou resultados otimistas quanto à possibilidade de selecionar os animais de raça pura para gerarem maior e melhor heterose genômica. Os coeficientes de regressão estimados mostraram que a heterozigosidade avaliada para um grupo de loci genômicos deve indicar associações favoráveis com a heterose genômica, estimada pelos desvios de dominância genômicos e, portanto, deve representar a ocorrência de heterose molecular se selecionado os animais de acordo com a taxa de heterozigosidade genômica.

Adicionalmente, a adaptação do GWAS para captar variações genéticas não-aditivas revelou regiões genômicas extras quando contabilizada uma matriz de relacionamento genômico de dominância, comparado ao modelo aditivo, com proporção da variância

explicada pelos SNPs adjacentes igual ou superior em quase todas as características avaliadas. Além disso, o gene *SOX5* associado adjacente a região do SNP Hapmap36729-SCAFFOLD106566_3245 apresenta função biológica importante para a saúde humana, sendo este o primeiro registro da presença deste gene em uma região com possível vantagem heterozigótica, em vacas leiteiras. Portanto, incluir os efeitos de dominância no modelo de GWAS adaptado pode contribuir para a identificação de marcadores associados ao fenótipo, de forma semelhante ou superior ao modelo aditivo adaptado.

REFERÊNCIAS

- Abdollahi-Arpanahi, R.; Gianola, D.; Peñagaricano, F. Deep learning versus parametric and ensemble methods for genomic prediction of complex phenotypes. **Genetics Selection Evolution**, v.52, n.12, 2020.
- Aguilar, I, Misztal, I.; Jhonson, D. L.; Legarra, A.; Tsuruta, S.; Lawlor, T. J. Hot topic: a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. **Journal of Dairy Science**, v.93, n. 2, p.743-752, 2010.
- Aguilar, I. 2014. Short introduction to BLUPF90 family programs. Retrieved on 20 March 2021, from http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=slides_2014_lva.pdf.
- Aguilar, I., Misztal, I.; Tsuruta, S.; Legarra, A.; Wang, H. PREGSF90- - POSTGSF90: Computational tools for the implementation of single-step genomic selection and genome-wide association with ungenotyped individuals in BLUPF90 programs. World Congress on Genetics Applied to Livestock Production (WCGALP), Vancouver, Canada. American Society of Animal Science, 2014, Proceeding 10th. Hal-02743809. Retrieved on 20 March 2021, from <https://hal.inrae.fr/hal-02743809>.
- Aguilar, I.; Legarra, A.; Cardoso, F.; Masuda, Y.; Lourenco, D.; Misztal, I. Frequentist p-values for large-scale single step Genomic-wide association, with an application to birth weight in American Angus cattle. **Genetics Selection Evolution**, v.51, n.28, 2019.
- Alillo, H.; Pryce, J. E.; González-Recio, O.; Cocks, B. G.; Goddard, M. E.; Hayes, B. J. Including nonadditive genetic effects in mating programs to maximize dairy farm profitability. **Journal of Dairy Science**, v.100, n.2, p.1203-1222, 2017.
- Amadeu, R. R.; Cellon, C.; Lara, L.; Resende, M.; Oliveira, I.; Ferrao, L.; Muñoz, P.; Garcia, A. AGHmatrix: Relationship Matrices for Diploid and Autopolyploid Species. Retrieved on 15 February 2021, from <https://github.com/prmunoz/AGHmatrix>.
- Begheri, M.; Miraie-Ashtiani, R.; Moradi-Shahrbabak, M.; Nejati-Javaremi, A.; Pakdel, A.; von Borstel, U. U.; Pimentel, E. C. G.; Konig, S. Selective genotyping and logistic regression analyses to identify favorable SNP-genotypes for clinical mastitis and production traits in Holstein dairy cattle. **Livestock Science**, v.151, n.2-3, p.140-151, 2013.

- Cao, Z.; Guo, Y.; Yang, Q.; He, Y.; Fetouh, M. I.; Warner, R., M.; Deng, Z. Genome-wide identification of quantitative trait loci for important plant and flower traits in petunia using a high-density linkage map and an interspecific recombinant inbred population derived from *Petunia integrifolia* and *P. axillaris*. **Horticulture Research**, v.6, n.27, 2019.
- Christensen, O. F.; Lund, M. Genomic prediction when some animals are not genotyped. **Genetics Selection Evolution**, v.42, n.1, 2010.
- Clasen, J. B.; Norberg, E.; Madsen, P.; Pederson, J.; Kargo, M. Estimation of genetic parameters and heterosis for longevity in crossbred Danish dairy cattle. **Journal of Dairy Science**, v.100, n.8, p.6337-6342, 2017.
- Deng, H. W.; Fu, Y. X. Conditions for positive and negative correlations between fitness and heterozygosity in equilibrium populations. **Genetics**, v.148, n.3, 1998.
- Fragomeni, B. de O.; Misztal, I.; Lourenco, D. L.; Aguilar, I.; Okimoto, R.; Muir, W. M. Changes in variance explained by top SNP windows over generations for three traits in broiler chicken. **Frontiers in Genetics**, v.1, n.332, 2014.
- Gauthier, S. F.; Pouliot, S. F.; Maoubois, J. L. Growth factors from bovine milk and colostrum: composition, extraction and biological activities. **Le Lait**, v.86, n.2, p.99-125, 2006.
- Gutiérrez-Gil, B.; Arranz, J. J.; Pong-Wong, R.; García-Gómez, E.; Kijas, J.; Wiener, P. Application of selection mapping to identify genomic associated with dairy production in sheep. **PLoS ONE**, v.9, n.5, 2014.
- Habier, D.; Fernando, R. L.; Kizilkaya, K.; Garrick, D. J. Extension for the Bayesian alphabet for genomic selection. **BMC Bioinformatics**, v.12, n.186, 2011.
- ICAR – International Committee for Animal Recording (2017). Section 2: Cattle milk recording. Retrieved on 24 May 2021, from, <https://www.icar.org/Guidelines/02-Overview-Cattle-Milk-Recording.pdf>
- Iung, L. H. S.; Petrini, J.; Ramírez-Díaz, J.; Salvian, M.; Ravadoski, G. A.; Pilonett, F.; Dauria, B. D.; Machado, P. F.; Coutinho, L. L.; Wiggans, G. R.; Mourão, G. B. Genome-wide association study for milk production traits in a Brazilian Holstein population. **Journal of Dairy Science**, v.102, n.6, p.5305-5314, 2019.
- Kargo, M.; Clasen, J. B.; Nielsen, H. M.; Byskov, K.; Norberg, E. Short communication: Heterosis and breed effects for milk production and udder health traits in crosses between Danish Holstein, Danish Red, and Danish Jersey. **Journal of Dairy Science**, v.104, n.4, p.678-682, 2021.
- Kelleher, M. M.; Berry, D. R.; Keraney, J. F.; McParland, S.; Buckley, F.; Purifield, D. C. Inference of population structure of purebred dairy and beef cattle using high-density genotype data. **Animal**, v.11, n.1, p.15-23, 2017.
- 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.
- Marques, D. B. D.; Bastiaansen, J. W. M.; Broekhuijse, M. L. W. J.; Lopes, M. S.; Knol, E. F.; Harlizius, B.; Guimarães, S. E. F.; Silva, F. F.; Lopes, P. S. Weighted single-step GWAS and gene network analysis reveal new candidate genes for semen traits in pigs. **Genetics selection Evolution**, v.50, n.40, 2018.

- Masuda, Y. **Introduction to BLUPF90 suit programs Standard edition**. University of Georgia, 2019.
Retrieved from http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=tutorial_blupf90.pdf
- 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.
- Misztal, I.; Tsuruta, S.; Lourenco, D.; Masuda, Y. **Manual for BLUPF90 family of programs**. University of Georgia, Athens, USA, 2018.
- Nicolini, P.; Amorín, R.; Han, Y.; Peñagarciano, F. Whole-genome scan reveals significant non-additive effects for sire conception rate in Holstein cattle. **BMC Genetics**, v.19, n.14, 2018.
- Petrini, J.; Jung, L. H. S.; Rodriguez, M. A. P.; Salvian, M.; Pértille, F.; Ravadoski, G. A.; Cassoli, L. D.; Coutinho, L. L.; Machado, P. F.; Wiggans, G. R.; Mourão, G. B. **Animal Breeding and Genetics**, v.133, n.5, p.384-395, 2016.
- R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from <https://www.R-project.org/>.
- Rovadoski, G. A.; Pertile, S. F. N.; Alvarenga, A. B.; Cesar, A. S. M.; Pértille, F.; Petrini, J.; Franzo, V.; Soares, W. V. B.; Morota, G.; Spangler, M. L.; Pinto, L. F. B.; Carvalho, G. G. P.; Lanna, D. P. D. Coutinho, L. L.; Mourão, G. B. Estimates of Genomic heritability and genome-wide association study for fatty acids profile in Santa Inês sheep. **BMC Genomics**, v.19, n.375, 2018.
- Saborío-Montero, A.; Vargas-Leitón, B.; Romero-Zuñiga, J. J.; Camacho-Sandoval, J. Additive genetic and heterosis effects for milk fever in a population of Jersey, Holstein x Jersey, and Holstein cattle under grazing conditions. **Journal of Dairy Science**, v.101, n.10, p.9128-9134, 2018.
- Silva, F. L.; Pinto, M. S.; Carvalho, A. F.; Perrone, I. T. Fator de crescimento transformador beta (TGF- β). **Revista do Instituto de Laticínios Cândido Tostes**, v.70, n.4, p.226-238, 2016.
- Sorensen, M. K.; Norberg, E.; Pedersen, J.; Christensen, L. G. Invited Review: Crossbreeding in Dairy Cattle: A Danish Perspective. **Journal of Dairy Science**, v.91, n.11, p.4116-4128, 2008.
- Tambasco-Talhari, D.; Alencar, M. M.; Paz, C. C. P.; Cruz, G. M.; Rodrigues, A. A.; Packer, I. U.; Coutinho, L. L.; Regitano, L. C. A. Molecular marker heterozygosities and genetic distances as correlates of production traits in F1 bovine crosses. **Genetics and Molecular Biology**, v.28, n.2, 2005.
- Tiezzi, F.; Parker-Gaddis, K. L.; Cole, J. B.; Clay, J. S.; Maltecca, C. A genome-wide association study for clinical mastitis in first parity US Holstein cows using single-step approach and genomic matrix re-weighting procedure. **PLoS ONE**, v.10, n.2, 2015.
- Tsairidou, S.; Allen, A. R.; Pong-Wong, R.; McBride, S. H.; Wright, D. M.; Matika, O.; Pooley, C. M.; McDowell, S. W. J.; Glass, E. J.; Skuce, R. A.; Bishop, S. C.; Woolliams, J. A. An analysis of effects of heterozygosity in dairy cattle for bovine tuberculosis resistance. **Animal Genetics**, v.49, p.103-109, 2018.
- VanRaden, P. M. (2015). findhap.f90 – Find haplotypes and impute genotypes using multiple chip sets and sequence data. Retrieved from <https://aipl.arsusda.gov/software/findhap/>.

- VanRaden, P. M. Efficient Methods to Compute Genomic Predictions. **Journal of Dairy Science**, v.91, p.4414-4423, 2008.
- Vidotti, M.; S.; Lyra, D. H.; Morosini, J. S.; Granato, I.; S.; C.; Quecine, M. C.; Azevedo, J.; L.; de; Fritsche-Neto, R. Additive and heterozygous (dis)advantages GWAS models reveal candidate genes involved in the genotypic variation of maize hybrids to *Azospirillum brasiliense*. **PLoS ONE**, v.14, n.9, p.1-21, 2019.
- Vitezica, Z.; Varona, L.; Legarra, A. On the additive and dominant variance and covariance of individuals within the genomic selection scope. **Genetics**, v.195, n.4, p.1223-1230, 2013.
- Wang H.; Misztal I.; Aguilar I.; Legarra A. Genome-wide association mapping including phenotypes from relatives without genotypes. **Genetics Research**, v.94, n.2, p.73-83, 2012.
- Wang H.; Misztal I.; Aguilar I.; Legarra A.; Fernando, R. L.; Vitezica, Z.; Okimoto, R.; Wing, T.; Hawken, R.; Muir W. M. Genome-wide association mapping including phenotypes from relatives without genotypes in single-step (ssGWAS) for 6-week body weight in broiler chickens. **Frontiers in Genetics**, v.5, p.1–10, 2014.
- Xiao, J.; Li, J.; Yuan, L.; Tanksley, S. D. Dominance is the major genetic basis of heterosis in rice as revealed by QTL analysis using molecular markers. **Genetics**, v.140, n.2, 1995.

5. CONSIDERAÇÕES FINAIS

Neste estudo foram seguidas duas principais linhas de pesquisa sobre a inferência, análise e utilização de fatores genéticos não-aditivos na avaliação de vacas leiteiras. Por meio de uma simulação de dados estocástica e posterior aplicação prática, conclui-se que os efeitos genéticos de dominância devem beneficiar de alguma forma a estimação dos componentes de variância genética, como para o conteúdo de ácido graxo poli-insaturado e porcentagem de gordura no leite, relatadas com maior magnitude de herdabilidade quando contabilizado os efeitos de dominância no modelo genômico.

Adicionalmente, uma leve redução no viés de predição foi relatada para o conteúdo de ácido graxo, ao incluir os efeitos de dominância no modelo. Isto indica que os efeitos não-aditivos como a dominância, devem ser considerados nas avaliações de características complexas, de baixa magnitude de herdabilidade e altamente influenciadas pelo ambiente, mesmo não havendo aumentos ou melhorias nas estimativas de acurácia.

Os resultados apresentados são esperançosos e otimistas para a ocorrência de heterose genômica em populações de animais de raça pura, devido aos coeficientes de regressão favoráveis e significativos entre heterozigiosidade e heterose genômica, para a maioria das populações simuladas. Isto concorda com os resultados em dados reais, na qual para produção de leite e o conteúdo de ácido graxo poli-insaturado as estimativas de regressão foram positivas e significativas, em relação a heterozigiosidade estimada no conjunto de QTLs selecionados. Portanto, o GWAs adaptado foi importante para a identificação de regiões genômicas e QTLs com possível vantagem heterozigótica, sendo essa informação relevante para a seleção de grupos de vacas com potencial para gerar maior e melhor heterose genômica.