

UNIVERSIDADE DE SÃO PAULO
FACULDADE DE ZOOTECNIA E ENGENHARIA DE ALIMENTOS

ALISSON STEFANY ACERO VALDERRAMA

**Avaliação do impacto da inclusão de informação genômica na estrutura dos
grupos de pais desconhecidos em bovinos da raça Nelore**

Versão corrigida

Pirassununga

2022

ALISSON STEFANY ACERO VALDERRAMA

Avaliação do impacto da inclusão de informação genômica na estrutura dos grupos de pais desconhecidos em bovinos da raça Nelore

Versão corrigida

Texto de dissertação apresentada à Faculdade de Zootecnia e Engenharia de Alimentos da Universidade de São Paulo, como parte dos requisitos para a obtenção do título de Mestre em Ciências do programa de Mestrado em Zootecnia.

Área de Concentração: Qualidade e produtividade Animal

Orientador: Prof. Dr. Joanir Pereira Eler
(Co) orientador: Rafael Espigolan

Pirassununga

2022

Ficha catalográfica elaborada pelo
Serviço de Biblioteca e Informação, FZEA/USP,
com os dados fornecidos pelo(a) autor(a)

A173a

Acero Valderrama, Alisson Stefany
Avaliação do impacto da inclusão de informação
genômica na estrutura dos grupos de pais
desconhecidos em bovinos da raça Nelore / Alisson
Stefany Acero Valderrama ; orientador Joanir
Pereira Eler ; coorientador Rafael Espigolan. --
Pirassununga, 2022.
90 f.

Dissertação (Mestrado - Programa de Pós-Graduação
em Zootecnia) -- Faculdade de Zootecnia e
Engenharia de Alimentos, Universidade de São Paulo.

1. Avaliação Genética. 2. Bovinos de corte. 3.
Melhoramento genético. 4. Seleção Genômica. 5. single-
step. I. Pereira Eler, Joanir, orient. II.
Espigolan, Rafael, coorient. III. Título.

ALISSON STEFANY ACERO VALDERRAMA

Avaliação do impacto da inclusão de informação genômica na estrutura dos grupos de pais desconhecidos em bovinos da raça Nelore

Dissertação apresentada á Faculdade de Zootecnia e Engenharia de Alimentos da Universidade de São Paulo, como parte dos requisitos para a obtenção do Título de Mestre em ciências.

Área de concentração: Qualidade e Produtividade Animal

Data de aprovação: __/__/____.

Banca Examinadora:

Prof.(a) Dr.(a) Rafael Espigolan

Instituição: Faculdade de Zootecnia e Engenharia de Alimentos – FZEA/USP

Presidente da Banca Examinadora

Prof. (a) Dr.(a) Bárbara da Conceição Abreu Silva

Instituição: Universidade Federal de Sergipe – Departamento de Zootecnia

Prof. (a) Dr.(a) José Bento Ferraz Sterman

Instituição: Faculdade de Zootecnia e Engenharia de Alimentos – FZEA/USP

DEDICATÓRIA

Á minha mãe, Rosa Maria Valderrama, a coluna vertebral da nossa família, quem me ensinou a ser forte sem importar a distância e a soledade, por seu caráter, fé e fortaleza, por me mostrar que sempre é possível com dedicação e trabalho, pelo apoio e espera incondicional.

AGRADECIMENTOS

A Deus por minha vida e de todas as pessoas que estimo e admiro.

A minha família, por me apoiar sem importar os obstáculos, agradeço imensamente mesmo estando tão longe, pois foram parte inicial da minha formação como pessoa, com o exemplo no trabalho da terra.

Ao Prof. Dr. Joanir Pereira Eler, pela orientação, aceitando-me como sua aluna e estar disponível sempre para os questionamentos e o andamento do trabalho.

Ao Dr. Rafael Espigolan pelo conhecimento, orientação, conversas do trabalho e a sua experiência no campo da genômica.

Ao Prof. Dr. José Bento Sterman Ferraz, por ser quem me “abriu as portas” do GMA, obrigada professor por me aceitar no primeiro instante.

A Elisangela Chicaroni Mattos, que sempre esteve disponível para dar suporte, ajuda e orientação nas análises, Eli muito obrigada.

A meus amigos, que me acolheram, inspiraram e me aconselharam, que me cobraram e consolaram, com quem formamos uma família moderna no Brasil, Barbara da Conceição Abreu Silva e Fernando de Oliveira Bussiman, vocês me ensinaram a falar, a estudar, até melhorar meus defeitos com muita paciência, muito obrigada estou em dívida com vocês, as palavras não conseguem disser o quanto vocês impactaram na minha vida.

A meu namorado Arley, pelo apoio, a espera e me encorajar mesmo no distanciamento.

A Felipe Eguti de Carvalho, Carolina Assis Almeida e Fernanda Schneberger, companheiros de caminho que me deram a amizade e formam parte de momentos importantes durante este período de formação.

A Bruno Casali Sarzi, que me aceito no seu núcleo familiar e em sua casa, sempre esteve disposto a me ajudar quando precisei.

Aos colegas do GMA/FZEA-USP, que mesmo se não estivemos muito tempo junto no espaço físico, aportaram ferramentas para o desenvolvimento do trabalho e da minha experiência neste tempo.

A Faculdade de Zootecnia e Engenharia de Alimentos (FZEA-USP) por me aceitar como parte da instituição e viabilizar a conclusão do meu mestrado e minha pesquisa.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

RESUMO

ACERO, A. S. V. **Avaliação do impacto da inclusão de informação genômica na estrutura dos grupos de pais desconhecidos em bovinos da raça Nelore**. 2022. 95 f. Dissertação (Mestrado) - Faculdade de Zootecnia e Engenharia de Alimentos, Universidade de São Paulo, Pirassununga, 2022.

A utilização de reprodutores múltiplos é o sistema de manejo mais comum na produção extensiva brasileira de bovinos de corte, onde vários touros são mantidos no piquete para se acasalar com vacas durante a estação de monta. No rebanho comercial do Brasil, o percentual de animais com incerteza de paternidade pode chegar a 40%. Nesse contexto, os grupos de pais desconhecidos (UPG) foram idealizados para solucionar problemas de falta de dados no pedigree no modelo animal. Nenhuma abordagem foi aceita de forma geral, pois cada um deles tem implicações diferentes. Portanto, o objetivo geral deste projeto consiste em avaliar a aplicação dos modelos de Melhor Predição Linear Não Viesada (*Best Linear Unbiased Prediction* - BLUP) e BLUP genômico de "passo único" (*single-step Genomic Best Linear Unbiased Prediction* - ssGBLUP), incluindo UPG e considerando ou não a informação genômica em avaliações genéticas de rebanhos comerciais de bovinos da raça Nelore. Inicialmente, foram utilizados dados fenotípicos e genotípicos simulados para uma característica de baixa e moderada herdabilidade (0,10 e 0,35, respectivamente) por meio do programa QMSIM com 5 réplicas, testando os cenários de 30%, 50% e 70% de informação faltante do pai no pedigree. Os UPG foram definidos com informação de geração (UPGY, ssUGPY), e a combinação da geração e o sexo do animal (UPGS, ssUPGS). Eles foram adicionados somente na matriz de parentesco baseada nas informações de pedigree (**A**) ou na matriz de parentesco combinada (**H**). Na segunda abordagem, foram utilizados dados reais, pertencentes à Agro-Pecuária CFM que participa do programa de avaliação genética realizado pelo Grupo de Melhoramento Animal e Biotecnologia (GMAB) da Faculdade de Zootecnia e Engenharia de Alimentos (FZEA/USP). No total, a informação fenotípica de 500.000 animais (253.000 machos e 247.000 fêmeas) e um pedigree contendo 642.000 animais foram processadas. No banco de dados foram encontrados 3.174 pais e 240.733 mães com 171.406 e 584.547 progênes, respectivamente. Um total de 243.907 indivíduos com progênie e 171.115 animais sem registros de pai e mãe e 344.564 animais com paternidade desconhecida, para as características de peso á desmama (PD), ganho de peso da desmama ao sobreano ajustado aos 550 dias de idade (GDPSOB), perímetro escrotal (PE), escore de musculosidade (MUSC) e peso aos 18 meses (PES18). A estrutura de análise para os dados reais foi similar à utilizada com os dados simulados, apenas trocando a geração simulada pelo ano de nascimento. Todas as rotinas de análises foram realizadas utilizando os programas da família BLUPF90. Os resultados encontrados sugerem que os modelos com UPG e informação genômica geraram GEBVs com maior acurácia, menos viesados e mais estáveis em comparação com a utilização dos UPG sem informação genômica. Ressalta-se que os UPG devem ser definidos e atribuídos com cuidado.

Palavras-chave: Avaliação genética. Bovinos de corte. BLUP. Single-step. SNP. UPG.

Abstract

ACERO, A. S. V. **Evaluating the impact of the inclusion of genomic information on the structure of unknown parent groups in Nelore cattle.** 2022. 95 f. Dissertação (Mestrado) - Faculdade de Zootecnia e Engenharia de Alimentos, Universidade de São Paulo, Pirassununga, 2022.

Multiple sires natural mating is the most common mating system in extensive beef cow-calf production, where several sires are kept in the same paddock to breed with cows during the mating season. In Brazil commercial herds, the percentage of animals with paternity uncertainty can reach up to 40%. In this regard, the unknown parent groups (UPG) were designed to accommodate missing pedigrees and data in the animal model. No approach was generally accepted as each has different implications. Therefore, the general objective of this work is to evaluate the application of Best Linear Unbiased Prediction (BLUP) and single-step Genomic Best Linear Unbiased Prediction (ssGBLUP) models including UPG considering or not the genomic information in genetic evaluations of commercial Nelore beef cattle herds. Initially, simulated phenotypic and genotypic data were used for a trait of low and moderate heritability (0.10 and 0.35, respectively) using QMSIM software with 5 replicates, and testing scenarios with 30%, 50% and 70% of unknown sires in the pedigree. The UPG were defined with generation information (UPGY, ssUPGY), and the combination of the animal generation and sex (UPGS, ssUPGS). These ones were added only in the relationship matrix based on the pedigree information (**A**) or in the combined genetic-genomic relationship matrix (**H**). In second approach, real data, belonging to Agro-Pecuária CFM which participates of the genetic evaluation program performed by Animal Breeding and Biotechnology Group (GMAB) from Faculty of Animal Science and Food Engineering (FZEA/USP) were used. In total, the phenotypic information was 500,000 animals (253,000 males and 247,000 females) and a pedigree containing 642,000 animals. In the database was found 3,174 sires and 240,733 dams with 171,406 and 584,547 progenies, respectively. A total of 243,907 individuals with progeny and 171,115 animals without known sire and mother and 344,564 animals with unknown paternity, for the traits of weaning weight (WW), post-weaning weight gain adjusted for 550 days (WWG), scrotal circumference (SC), muscle score (MUSC) and post-yearling weight (PW). The structure of analysis of the real data was similar to that used for the simulated data swapping the simulated generation for year of birth of the animal. Analyzes were performed using BLUPF90 family programs. The results found suggest that the UPG, with genomic information, increases the accuracy and decreases the bias of the GEBVs, also showing the stability of genetic genomic values compared to the use of UPG without genomic data. It is noteworthy that UPG must be defined and assigned carefully.

Keywords: Beef cattle. BLUP. Genetic evaluation. Single-step. SNP. UPG.

SUMARIO

1	INTRODUÇÃO	12
1.1	Objetivo.....	14
1.1.1	Objetivo geral.....	14
1.1.2	Objetivos específicos.....	14
2	REVISÃO BIBLIOGRÁFICA	15
2.1	Produção de gado de corte no Brasil.....	15
2.2	Programas de melhoramento genético.....	16
2.2.1	Componentes de variância.....	18
2.2.2	Matrizes de relacionamento (matrizes A, G e H).....	19
2.2.3	Avaliação genética tradicional.....	21
2.2.4	Avaliação genômica.....	22
2.3	Paternidade desconhecida.....	24
2.3.1	Grupos de pais desconhecidos (UPG) e modelo misto.....	28
3	REFERENCIAS.....	34
4	AVALIAÇÃO DO USO DE GRUPOS DE PAIS DESCONHECIDOS EM CARACTERÍSTICAS SIMULADAS DE MODERADA E BAIXA HERDABILIDADE.....	42
	Resumo.....	42
	Abstract.....	43
4.1	Introdução.....	43
4.2	Material e métodos.....	45
4.2.1	População Simulada.....	45
4.2.2	Genoma simulado.....	46
4.2.3	Modelos.....	47
4.2.4	Cenários.....	49
4.2.5	Análise dos resultados.....	50
4.3	Resultados e discussão.....	50
4.4	Conclusão.....	60
	Referências.....	60
5	INCLUSÃO DE GRUPOS DE PAIS DESCONHECIDOS NA AVALIAÇÃO GENÉTICA-GENÔMICA EM REBANHOS COMERCIAIS DA RAÇA NELORE.....	63
	Resumo.....	63

Abstract.....	64
5.1 Introdução.....	65
5.2 Material e Métodos	66
5.2.1 Dados genealógicos e fenotípicos.....	66
5.2.2 Dados genômicos.....	69
5.2.3 Parâmetros genéticos	70
5.2.4 Modelos	71
5.2.5 Cenários testados.....	72
5.2.6 Análise dos resultados.....	73
5.3 Resultados e discussão	74
5.4 Conclusão.....	87
Referências.....	87

1 INTRODUÇÃO

As condições climáticas tropicais e um extenso território contribuíram para alcançar resultados de destaque e competitividade a nível mundial, favorecido por que, a maioria dos animais é mantida no sistema de pastejo, o que favorece o baixo custo de produção (PEREIRA et al., 2016). Desse modo, é importante ressaltar que os sistemas de produção da pecuária brasileira permitem a diversificação produtiva podendo atender diferentes mercados com requisitos específicos, que variam de cortes de carne de menor valor agregado até mercados padrão e aqueles com alto grau de qualidade (ABIEC, 2019). Portanto, pode-se afirmar que a produção brasileira de gado de corte tem um amplo potencial para atender os diferentes mercados de consumo.

Uma prática de manejo muito utilizada nos sistemas de produção extensiva de bovinos de corte é a utilização de reprodutores múltiplos (RM), em que, o grupo de touros selecionados para reprodução permanece em piquetes junto às vacas durante a estação de monta (ARAÚJO et al., 2012; TONUSSI et al., 2017). Nos rebanhos comerciais brasileiros, a porcentagem de animais com pai desconhecido (progênes de touros múltiplos) pode alcançar até 40% e em muitas propriedades, os touros são usados em uma temporada de monta e descartados logo, dificultando a seleção dos animais por características de interesse econômico (CARDOSO; TEMPELMAN, 2003; TONUSSI et al., 2017).

Os RM atualmente são formados com touros provados, e graças à avaliação genômica consegue-se o relacionamento com a progênie (ELER, 2019), no entanto, a progênie que não é genotipada tende a ter só a informação do RM o que aumenta a incerteza de paternidade mesmo com o modelo de matriz de parentesco combinada single-step (TONUSSI et al., 2017). Isto pode adicionar uma atribuição errada de parentesco (WOLAK; REID, 2017), além de conflitos com os animais com genética estrangeira, pois eles não têm conexão com a população e suas progênes seriam as primeiras avaliadas dentro do rebanho, limitando as conexões genéticas (BONIFAZI et al., 2020).

Nesse contexto, os grupos de pais desconhecidos (UPG) foram idealizados para ajustar os pedigrees e solucionar problemas de falta de dados no modelo animal (QUAAS, 1988), particularmente quando a seleção é feita em países estrangeiros e, portanto, os animais importados apresentarem média produtivas diferentes quando comparados às populações nacionais (THOMPSON, 1979). O uso de UPG também é

comum nos rebanhos onde é difícil identificar a paternidade, como no caso em rebanhos que utilizam touros múltiplos para acasalamento (TONUSSI et al., 2017), além de ser importante para modelar populações multirraciais de gado de corte (LEGARRA et al., 2007), permitindo estimativas mais realistas das tendências genéticas (THERON; KANFER; RAUTENBACH, 2002).

Em 1988, Henderson propôs a metodologia para obtenção da Melhor Predição Linear Não-Viesada (*Best Linear Unbiased Prediction* - BLUP), um modelo misto baseado na matriz de parentesco, que permite a inclusão da informação de touros múltiplos baseados em UPG. O BLUP pressupõe a estimativa de cada candidato como verdadeiro pai de um animal, portanto, os UPG podem ser incluídos nas equações de modelos mistos para levar em conta a diferença genética entre os grupos definidos em uma avaliação genética (TSURUTA et al., 2014).

Na última década, o desenvolvimento de painéis contendo informações de milhares de marcadores moleculares propiciou o aprimoramento dos tradicionais modelos para estimativa dos valores genéticos. Nesse sentido, Legarra et al. (2009) e Misztal et al. (2009) propuseram o método de Melhor Predição Linear Não Viesada genômica de "passo único" (*single-step Genomic Best Linear Unbiased Prediction* - ssGBLUP), uma metodologia que é computacionalmente eficiente e proporciona previsões mais acuradas, sendo amplamente utilizada em diversas espécies, como no gado leiteiro (AGUILAR et al., 2010; TSURUTA et al., 2011), aves (CHEN et al., 2011), suínos (CHRISTENSEN et al., 2012) e gado de corte (TONUSSI et al., 2017).

Atualmente, os modelos que consideram os UPGs nos rebanhos comerciais de gado de corte utilizando dados genômicos não são utilizados nos programas de melhoramento genético (TONUSSI et al., 2017). Ainda assim, pedigrees deficientes ou com erros são comuns nos sistemas extensivos de produção de gado de corte. Devido ao escasso número de estudos que se concentram neste problema, é um tema válido para estudo metucioso. Portanto, o propósito deste trabalho foi avaliar a aplicação dos modelos: Melhor Predição Linear Não Viesada (*Best Linear Unbiased Prediction* - BLUP) e BLUP genômico de "passo único" (*single-step Genomic Best Linear Unbiased Prediction* - ssGBLUP), incluindo UPG e considerando ou não a informação genômica na avaliação genética de rebanhos comerciais de bovinos da raça Nelore.

1.1 Objetivo

1.1.1 Objetivo geral

Avaliar a aplicação dos métodos Melhor Predição Linear Não Viesada (*Best Linear Unbiased Prediction* - BLUP) e BLUP genômico de "passo único" (*single-step Genomic Best Linear Unbiased Prediction* - ssGBLUP), incluindo UPG e considerando ou não a informação genômica em avaliações genéticas de rebanhos comerciais de bovinos da raça Nelore.

1.1.2 Objetivos específicos

- Explorar as diferentes definições de UPG, como ano de nascimento, sexo, de acordo com a característica de interesse e determinar a estratégia mais adequada para o conjunto de dados;
- Avaliar o uso de UPG nas avaliações genéticas para uma característica de baixa (0,10) e moderada herdabilidade (0,35) utilizando dados simulados e sua influência em diferentes percentagens de pais desconhecidos;
- Estimar os valores genéticos aplicando os modelos BLUP e ssGBLUP, incluindo os UPGs, considerando, dessa forma, as informações genômicas para as características de interesse econômicas;
- Avaliar os critérios de qualidade dos modelos nos cenários com inclusão de UPG com os valores genéticos preditos, utilizando ou não informação genômica.

2 REVISÃO BIBLIOGRÁFICA

2.1 Produção de gado de corte no Brasil

A pecuária de corte foi responsável por cerca de 8,7% do PIB da agropecuária brasileira no ano 2018, movimentando mais de R\$ 597,22 bilhões, valor que inclui desde os insumos utilizados na produção do gado, investimento em genética, faturamento dos animais até o total comercializado pelas indústrias e varejos, sendo o maior já registrado nos últimos dez anos (ABIEC, 2019). No abate médio anual de animais, o volume é de 43,3 milhões de cabeças, com peso médio de carcaça de 234 quilos, rendimento de carcaça entre 51% e 55% (MALAFAIA et al., 2019).

Diante da crescente demanda mundial de alimentos e proteína de origem animal de qualidade e sustentável, é importante o desenvolvimento e a implementação de novas tecnologias nos sistemas de produção pecuários para obedecer aos exigentes padrões do mercado consumidor. A produção pecuária é fundamentada de acordo com algumas áreas do conhecimento, principalmente a nutrição, sanidade, melhoramento genético e manejo. Com atenção conjunta a todas as áreas e o desenvolvimento de estratégias que envolvam desde a escolha do material genético até a padronização do produto ofertado, principalmente para o peso, idade e grau de acabamento, é possível atingir maiores índices produtivos (PINHEROS, 2011).

No intuito de aumentar a produtividade do gado de corte, uma das ferramentas principais dos programas de melhoramento genético consiste na seleção, a qual altera as frequências alélicas, selecionando animais geneticamente superiores para características de interesse econômico. Por conseguinte, os animais com maior taxa de crescimento ou precocidade sexual podem contribuir diretamente na diminuição do ciclo de produção da carne, permitindo um maior retorno econômico (BOLIGON et al., 2009).

Nesse contexto, a Agropecuária CFM vem desenvolvendo, desde 2006, pesquisas com genômica aplicada à seleção de animais superiores, tendo, recentemente, publicado o Sumário de Touros Nelore CFM 2020/2021 com animais avaliados pela metodologia do *single-step*, ou seja, utilizando informações de pedigree e genômicas (AGRO-PECUÁRIA CFM, 2020). Em 2018, surgiu a parceria entre a ABCZ (Associação Brasileira dos Criadores de Zebu) e a Embrapa somando suas

competências técnicas e seus bancos de dados para produzir uma única e conjunta avaliação genética das raças zebuínas, desenvolvendo avaliações genéticas genômicas, via “single step” de todas as características para a raça Nelore (ABCZ, 2019).

2.2 Programas de melhoramento genético

Os sistemas de produção mais eficientes são aqueles que conseguem otimizar todos os recursos disponíveis: genéticos e ambientais, com boas práticas de manejo, na reprodução, produção e o produto, sendo estes os principais componentes do ciclo produtivo em bovinos (BARBOSA, 2003). No sistema produtivo o melhoramento genético é uma ferramenta importante, pois é definido como a ciência que estuda as ações genéticas dos indivíduos e do ambiente na determinação de suas características de interesse econômico (FERRAZ; ELER, 2010). Em poucas palavras, é o encarregado, junto com a reprodução, da gestão e administração do recurso genético das espécies, esses recursos referem-se a cada animal, vegetal e/ou micróbio de algum valor (atual ou potencial) econômico, científico, social ou ambiental, que apoiam a subsistência de cada ser humano (BRASIL, 2019).

O objetivo do melhoramento genético de bovinos de corte se baseia na obtenção de animais de conformação estética desejável e que sejam capazes de produzir a maior quantidade de carne de boa qualidade por unidade de produção, em regime de pastagens e boas condições de manejo, no menor tempo e custo possíveis em cada geração. Desse modo, os programas que visam o adequado uso desses recursos e dos dados disponíveis na avaliação genética, são de extremo interesse para o aumento da eficiência e produtividade dos rebanhos brasileiros (PETRINI, 2012). Em 1990 no território brasileiro haviam mais de 60 raças de bovinos, algumas exploradas para produção comercial, no entanto, em relação ao Zebu, a raça Nelore e seus cruzamentos é o recurso genético mais utilizado para produção de carne bovina no Brasil (BARBOSA, 2003).

No Brasil no 2021, há 25 entidades que estão registradas ante o MAPA para a realização das provas zootécnicas, que podem emitir o Certificado especial de Identificação e produção (CEIP) para bovinos de corte. Os animais certificados têm os mesmos benefícios fiscais que os animais registrados em associações de criadores, independentemente de serem ou não registrados na associação. O CEIP (Certificado Especial de Identificação e Produção) estimulou os criadores a aumentar a

intensidade da seleção e tem mudando drasticamente o mercado da genética no país (CARVALHEIRO, 2014). Algumas das entidades registradas são: Gensys Consultores Associados, Conexão Delta G, Agropecuária CFM Ltda. DELTAGEN (Conexão Pecuária para o melhoramento genético), ANCP (Associação Nacional de Criadores e Pesquisadores), Associação de Criadores de Montana, IRCA Zootecnia, entre outras. Cada uma com a responsabilidade de registrar e aprovar os projetos técnicos de melhoramento genético para executar os trabalhos de registro genealógico de cada raça (BRASIL, 2021). Entretanto, há uma desvantagem, a falta de uma avaliação genética nacional única que permita uma seleção de indivíduos mais eficiente além de outros benefícios. Por outro lado, a existência de diferentes programas de melhoramento genético para a raça Nelore gera competitividade e ajuda a manter a variabilidade genética, o que acaba beneficiando a pecuária brasileira (CARVALHEIRO, 2014).

Graças a esses programas, muitos estudos no Brasil foram desenvolvidos para tornar as estimativas de componentes de variância mais robustas, principalmente, quando a espécie de interesse é a bovina. Durante este século foram analisados diferentes métodos de padronização dos pesos em diferentes idades, como por exemplo, aos 205, 365 e 550 dias de idade e constataram diferenças entre as médias dos pesos corporais calculadas pelos diferentes métodos, bem como diferenças nas estimativas de componentes de variância e de parâmetros genéticos. Assim, o método de padronização de peso pode influenciar o processo de seleção. Atualmente, a informação genômica pode contribuir com estudos envolvendo meta-análise, resultando em grande atenção nas diversas linhas de pesquisa em produção animal, pois a análise conjunta de um grande número de estudos pode ajudar na obtenção de estimativas mais acuradas de importantes parâmetros estatísticos (LÔBO; DE BITTNECOURT; PINTO, 2010).

Vários programas de melhoramento genético têm sido implementados visando à melhoria da eficiência produtiva e reprodutiva tanto na raça Nelore quanto das demais raças que compõem o rebanho nacional. Entre os principais programas, encontra-se o programa de melhoramento da Agropecuária CFM, no qual as características avaliadas são, basicamente peso à desmama (PD, Kg), ganho de peso da desmama ao sobre ano (GDPSOB, Kg), perímetro escrotal (PE, cm), escore de musculabilidade (MUSC) e peso aos 18 meses (PES18, Kg). Índice CFM no sumário de

2019, é calculado com base na seguinte equação (1), na qual os valores são padronizados:

$$I_{CFM} = 2 PD + 4 GPSOB + 2 MUSC + 2 PE \quad (1)$$

Em que, PD é a DEP para peso à desmama; GPSOB, é a DEP para ganho de peso da desmama ao sobre ano; MUSC é DEP para escores visuais de musculabilidade, medidos ao sobre ano; e PE é a DEP para perímetro escrotal medido ao sobre ano.

2.2.1 Componentes de variância

Nos rebanhos, os melhores animais irão se sobressair se houver variação nos desempenhos, logo eles estão posicionados mais distantes da média da característica mensurada, pois lembre-se que a similaridade entre parentes é devida aos genes compartilhados entre eles o que fornece a decisão de selecionar os melhores animais garantindo progênies superiores dentro da população que tem esses genes “melhoradores” (KINGHORN; WERF; RYAN, 2016). Os componentes de variância estimados permitem calcular a quanto porção do fenótipo é devida à genética principalmente a variância genética aditiva (σ_A^2) baseada no grau de similaridade entre parentes, e a herdabilidade (h^2) que é a proporção da superioridade dos pais que é transmitida para a próxima geração, estatisticamente definida como o coeficiente entre a variância genética aditiva (σ_A^2) e a variância fenotípica (σ_P^2) (FALCONER; MACKAY, 1996; KINGHORN; WERF; RYAN, 2016).

Existem vários métodos para estimação dos componentes de variância, os quais pretendem explicitar os diferentes efeitos fixos e aleatórios de modelos mistos aplicados em dados balanceados ou desbalanceados. Os métodos mais reconhecidos e utilizados são: a ANOVA (Método da Análise de Variância) desenvolvido por Fisher em 1918, os métodos de estimação I, II e III propostos por Henderson em 1953, em 1967, Hartley e Rao aplicaram o método de máxima verossimilhança (ML), o método de estimação quadrática não-viesada de norma mínima (MINQUE) por Rao em 1971, e em 1971 foi desenvolvido o método da Máxima Verossimilhança Restrita (REML) por Paterson e Thompson (MARCELINO; IEMMA, 2000).

A Máxima Verossimilhança Restrita (REML) é considerada o procedimento mais acurado para estimação dos componentes de variância, sendo o mais utilizado na estimação de parâmetros genéticos para conjunto de dados de animais de produção. Trata-se de um método que apresenta boa consistência, eficiência, gera

componentes de variância com valores dentro do espaço paramétrico e considera a perda de graus de liberdade referente aos efeitos fixos incluídos nos modelos (PATTERSON; THOMPSON, 1971). O método REML consiste na maximização da função densidade de probabilidade das observações em relação aos efeitos fixos e aos componentes de variância dos efeitos aleatórios do modelo; sendo cada observação dividida em duas partes independentes, um referente aos efeitos fixos e outra aos aleatórios, de maneira que a função densidade de probabilidade das observações é dada pela soma das funções densidade de probabilidade de cada parte (CORBEIL; SEARLE, 1976). Além disso, O REML considera as alterações na variância genética ocasionada pelo desequilíbrio de ligação, endogamia, seleção e pela tendência ou ganho genético realizado (KNOTT et al., 1995).

2.2.2 Matrizes de relacionamento (matrizes A, G e H)

As relações genéticas ou de parentesco entre os indivíduos é uma informação importante para a estimação de variâncias genéticas e herdabilidades. Nas diferentes espécies, incluindo na genética humana, as relações genéticas podem fornecer diferentes estimativas (LEGARRA, 2016) de acordo ao modelo estatístico, método de estimação, população base e o modelo de relacionamento da população. Nesse contexto, o modelo animal consiste em uma abordagem estatística bastante empregada, pois consegue separar os efeitos fixos dos aleatórios e genéticos em um fenótipo, no entanto, é fundamental a informação da característica (fenótipo), os fatores que poderiam influenciá-la e a informação de como os animais estão relacionados uns com os outros. Esta última obtida por meio das informações de pedigree (HENDERSON, 1976), dados genômicos, ou uma combinação entre ambos (LEGARRA; AGUILAR; MISZTAL, 2009; MUFF et al., 2019).

O parentesco é a relação existente entre dois indivíduos com pelo menos um ancestral em comum (ELER, 2017), a informação de linhagem (contida no pedigree por exemplo) fornece esta informação. No modelo animal, o parentesco é dado pela inversa da matriz **A** (\mathbf{A}^{-1}), que é estimada por meio das covariâncias genéticas aditivas entre os animais (ELER, 2017; WRIGHT, 1922). O método de para obtenção da matriz de parentesco baseada nas informações de pedigree é demonstrado detalhadamente por Henderson (1976a) para populações não endogâmicas usando os cálculos de covariâncias genéticas de Wright (1922) (onde **A** deve cumprir 3 propriedades) e com endogamia por Henderson (1975b).

Os coeficientes calculados de forma tradicional interpretam o compartilhamento esperado pelo genoma entre os indivíduos no pedigree, esta similaridade pode ser quantificada por análise direta de DNA, e é mensurada pela probabilidade de que os genes sejam idênticos por descendência geralmente calculadas a partir do pedigree (VANRADEN, 1992). Atualmente podemos medir o compartilhamento do genoma diretamente graças as novas tecnologias desenvolvidas como os marcadores moleculares, neste caso, a partir dos dados de polimorfismo de nucleotídeo único (SNP) de alta densidade podendo cobrir todo o genoma. Entretanto, o compartilhamento real do genoma pode diferir desses valores esperados (teoricamente) devido à informação disponível no pedigree e as contrastadas por SNP (SPEED; BALDING, 2015).

Os relacionamentos genômicos são inferidos quantificando o número de alelos compartilhados entre dois indivíduos, informados pelos SNPs (TORO et al., 2011). A princípio esses relacionamentos podem ser entendidos como “idêntico por estado” (IBS), se dois alelos escolhidos aleatoriamente, um de cada indivíduo, são idênticos, independentemente da origem. Entretanto, VanRaden (2008) propôs em seu método modificações para refletir a covariância entre indivíduos nos termos do relacionamento “idêntico por descendência” (IBD), ou seja, os alelos compartilhados, em algum momento, pertenceram a um ancestral comum, equivalente ao que ocorre no pedigree. Essas definições para IBD estariam totalmente corretas, primeiro, se o genoma considerado fosse infinito, fato que na realidade não ocorre, por tanto os relacionamentos via pedigree estarão corretos, na média. Segundo o pedigree não é infinito, limita-se pelo número de gerações observadas (THOMPSON, 2013).

Desse modo, a matriz “G” é computada de forma que na diagonal estão os números de loci homozigotos de cada indivíduo, e fora da diagonal estão os números de alelos compartilhados entre os membros da população (VANRADEN, 2007). Esses valores são uma aproximação dos relacionamentos reais mais acurada que aqueles obtidos pelo pedigree, e uma das principais propriedades para a predição genômica ser considerada melhor que análises tradicionais (FERNANDES JÚNIOR et al., 2016).

O desenvolvimento do ssGBLUP vem do fato de que normalmente apenas uma pequena parte dos animais, em uma determinada população, é genotipada (LOURENÇO et al., 2014). Desta forma, a melhor abordagem para evitar várias etapas nos procedimentos de obtenção dos GEBVs seria combinar pedigree e relações

genômicas em uma única matriz (chamada de H), e utilizá-la nas Equações do Modelo Misto.

Legarra et al. (2009) mostraram que a relação do pedigree pode ser vista como uma relação *a priori* e a relação genômica como a relação observada. Com a derivação da distribuição conjunta dos valores genéticos de todos os animais (genotipados ou não) conseguiu estender (ou imputar) a informação genômica para animais não genotipados. Isso significa que no ssGBLUP as relações de pedigree para animais não genotipados são aprimoradas pelas informações genômicas de seus parentes. Posteriormente, todas as ferramentas baseadas nas equações do modelo misto do BLUP, como a máxima verossimilhança restrita (PATTERSON; THOMPSON, 1971), podem ser facilmente convertidas ao método de passo único (2).

$$H^{-1} = A^{-1} \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & G^{-1} - A_{22}^{-1} \end{bmatrix} \quad (2)$$

Onde, A^{-1} é a inversa da matriz de relacionamento genética (pedigree), G^{-1} é a inversa da matriz de relacionamento genômica e A_{22}^{-1} é a inversa da matriz de relacionamento dos animais genotipados e que tem informação de pedigree.

2.2.3 Avaliação genética tradicional

A maioria dos programas de avaliação genética se baseia na análise estatística de grandes bancos de dados com fenótipos, utilizando o modelo misto linear para estimar os valores genéticos em candidatos à seleção (DEKKERS, 2012). Os modelos mistos analisados sob o modelo animal para a obtenção da melhor predição linear não viesada (BLUP), baseia-se em assumir os valores genéticos como efeitos aleatórios, corrigindo-se para os demais efeitos contidos no modelo, envolvendo todos os indivíduos identificados na estrutura genealógica da população (HENDERSON et al., 1959).

O BLUP ou modelo animal consiste em um modelo misto com a propriedade de estimar simultaneamente os efeitos fixos (efeitos ambientais compartilhados ou atributos individuais) e aleatórios em um fenótipo, o efeito genético (considerado também como um efeito aleatório) pode induzir semelhanças fenotípicas entre parentes (MUFF et al., 2019), inclusive ele permite a avaliação genética de animais em diferentes rebanhos, utilizando a matriz de parentesco completa, além da possibilidade de trabalhar com dados desbalanceados, estimar a tendência genética da população, utilizar informações de características correlacionadas e obter os

coeficientes de endogamia para os animais a partir da matriz **A**. Por utilizar toda a matriz de parentesco é possível estimar os valores genéticos dos animais sem nenhum registro de produção, associando-o com os valores genéticos de seus parentes, presentes na matriz de parentesco (HENDERSON, 1975a).

2.2.4 Avaliação genômica

Na aplicação da genética molecular em espécies animais, há a expectativa de que a informação ao nível do DNA incrementa de 8 a 38% de ganhos genéticos na população quando comparados aos ganhos genéticos propiciados com base apenas nos dados fenotípicos (BLUP) (HALEY; VISSCHER, 1998). As características quantitativas são geralmente afetadas por muitos genes e, desse modo, desenvolvimento de marcadores moleculares e mapas genéticos permitiu a detecção de QTL (*Quantitative Trait loci*) e o mapeamento de regiões no genoma que podem afetar tanto características quantitativas como qualitativas e a identificação de genes que controlam importantes variações genéticas, assim como mutações causais (MEUWISSEN; HAYES; GODDARD, 2001).

Os primeiros marcadores baseados em DNA utilizados foram os polimorfismos do comprimento do fragmento de restrição (RFLP- *Restriction Fragment Length Polymorphism*). Entretanto, esse tipo de marcador era lento para ser gerado e pouco informativo, contendo apenas dois alelos em qualquer população, além de necessitar de grandes quantidades da amostra de DNA. Posteriormente, a técnica da PCR (*Reação em cadeia da polimerase*) e dos microssatélites demandavam quantidades pequenas de DNA, e ofereciam mais informações sobre o genoma de vários alelos de uma população (HALEY; VISSCHER, 1998).

O DNA apresenta variações, que podem ser repetidas ao longo do genoma que são denominadas microssatélites, outra pode ser dada pela inserção ou deleção (InDels) ou em apenas uma base, esta última é denominada SNP (polimorfismo de nucleotídeo único) e é usado como um marcador molecular e estão distribuídos em todo o genoma, e para avaliação genômica pode fornecer informações não só do parentesco, mas também dos *loci* de interesse econômico (DEKKERS, 2012). A utilização de marcadores genéticos para identificar *loci* para características quantitativas e a implementação da informação gerada por esses marcadores moleculares em programas de melhoramento genético consiste no princípio da avaliação genômica (DEKKERS, 2012). A seleção genômica está se tornando

realidade em um número crescente de populações e espécies de animais e plantas (BOICHARD et al., 2016). Os primeiros trabalhos foram em gado leiteiro na seleção de touros jovens, na seleção de características de precocidade e crescimento em gado de corte, assim como em suínos e aves (MEUWISSEN; HAYES; GODDARD, 2016).

É importante salientar que a distribuição dos marcadores, não é uniforme, portanto, algumas áreas do genoma são mais densamente marcadas em comparação com outras, no entanto, as informações de relacionamento derivadas do marcador, complementam ou substituem a matriz de relacionamento padrão por relacionamentos reais, melhorando as estimativas de herdabilidade e a precisão da seleção através da limitação de erros de linhagem. As relações genômicas fornecidas pela cobertura dos marcadores de alta densidade, por exemplo, permitem estimar e controlar a consanguinidade, e, por consequência, também a depressão endogâmica e a heterose em todo o genoma ou em regiões particulares do genoma (HALEY; VISSCHER, 1998; MEUWISSEN; HAYES; GODDARD, 2001).

Dentre os modelos estatísticos que possibilitam a inclusão da informação genômica se destaca o GBLUP (Melhor Predição Linear Não-viesada Genômica, Genomic Best Linear Unbiased Prediction) que consiste em uma extensão do BLUP poligênico em que a matriz de parentesco é calculada baseada nas informações dos marcadores (matriz G) ao invés do pedigree (BOICHARD et al., 2016). A obtenção das estimativas dos efeitos dos marcadores pelo método GBLUP é caracterizada pela utilização das equações de Henderson. As vantagens da predição genômica incluem a maior precisão dos valores genéticos estimados, principalmente para características que são muito onerosas ou difíceis de medir e características que são avaliadas muito tarde na vida do animal ou limitada pelo sexo (MACNEIL, 2016), como no caso de características reprodutivas (MATILAINEN et al., 2018), eficiência alimentar e de carcaça.

Outra metodologia que possibilita a inclusão dos marcadores moleculares corresponde ao ssGBLUP (Método de Passo Único “single-step”- Genomic Best Linear Unbiased Prediction) proposta por Misztal, Legarra e Aguilar (2009), que permite a avaliação simultânea dos animais genotipados e não genotipados. O ssGBLUP consiste em usar a matriz combinada H (Matriz A e Matriz G) para estimar os valores genéticos genômicos (GEBV). A matriz H combinada pode ser obtida da seguinte forma (AGUILAR et al., 2010; MISZTAL; LEGARRA; AGUILAR, 2009):

$$H = \begin{pmatrix} \text{var}(\mathbf{u}_1) & \text{cov}(\mathbf{u}_1, \mathbf{u}_2) \\ \text{cov}(\mathbf{u}_2, \mathbf{u}_1) & \text{var}(\mathbf{u}_2) \end{pmatrix} = \begin{pmatrix} \mathbf{A}_{11} + \mathbf{A}_{12}\mathbf{A}_{22}^{-1}(\mathbf{G} - \mathbf{A}_{22})\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{G} \\ \mathbf{G}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{G} \end{pmatrix} \quad (3)$$

Onde, \mathbf{u}_1 são os animais no pedigree, \mathbf{u}_2 são os animais com genótipo, e os subíndices de \mathbf{A} são: 1 os animais que estão no pedigree e 2 os animais genotipados. No modelo animal, a inversa da matriz de parentesco (\mathbf{A}^{-1}) é substituída por a matriz \mathbf{H}^{-1} , que combina o pedigree e a informação genômica.

Não obstante, os animais da matriz \mathbf{A} e \mathbf{G} devem estar em escalas próximas, com a capacidade de conexão entre esses animais. Caso na matriz \mathbf{A} , a relação com a matriz \mathbf{G} seja incompatível, pode ocorrer uma superestimação da variância (FORNI; AGUILAR; MISZTAL, 2011), em consequência, os valores de predição genômica podem ser viesados, sendo este um dos desafios ao fazer este tipo de avaliação, para que sejam computacionalmente viáveis na contribuição no modelo misto (MATILAINEN et al., 2018). A incompatibilidade pode ser associada não somente à escala das matrizes \mathbf{A} e \mathbf{G} , mas também aos grupos genéticos de pais desconhecidos (MATILAINEN et al., 2018; MISZTAL et al., 2013a).

As vantagens da metodologia do ssGBLUP são: as estimativas de variâncias e herdabilidades genéticas ficam em uma escala semelhante (LEGARRA, 2016), valores genéticos estimados mais precisos (CHRISTENSEN; LUND, 2010), simplicidade na para assimilar os pesos automáticos para as várias fontes de informação para obtenção do valor genético (AGUILAR et al., 2010), e, com as correções para endogamia, compara animais de maneira mais justa, permitindo considerar uma heterogeneidade genômica, evitando problemas de convergência na metodologia (AGUILAR et al., 2020).

As informações genômicas consistem em uma ferramenta importante para os criadores, pois possibilitam a recuperação de relações genéticas de touros não incluídos no pedigree tradicional (por exemplo, pais desconhecidos, melhorando as decisões de acasalamento e a lucratividade do rebanho (TONUSSI et al., 2021).

2.3 Paternidade desconhecida

Parte fundamental do modelo animal é a informação do relacionamento dos indivíduos uns com outros, obtido a partir das diferentes dados disponíveis (pedigree, dados genômicos ou a sua combinação) (MUFF et al., 2019). Uma matriz de relacionamento completa permite uma avaliação conjunta, contabilizando informações

automaticamente, com facilidades de cálculos viáveis na prática (LEGARRA; AGUILAR; MISZTAL, 2009). Na realidade, bancos completos são raros, e a falta de informações sobre parentescos são comuns, tanto na avaliação dos animais como na seleção dos melhores exemplares para a reprodução e construir a nova geração melhorada com características desejáveis para a produção. Pois a falta de um pedigree completo ou ideal, significa que os valores de parentesco resultantes são arbitrários até certo ponto (SPEED; BALDING, 2015).

A paternidade incerta pode afetar as avaliações genéticas em diferentes pontos, como na estimação dos componentes de variância e nas previsões genéticas-genômicas. De maneira mais detalhada, as incertezas de paternidade, bem como os erros no pedigree e pedigrees deficientes, podem afetar:

- Variação fenotípica (σ_F^2): ao diminuir a conexão do pedigree, faz com que a variação fenotípica seja atribuída a efeitos ambientais ao invés de genéticos, o que danifica as informações-chaves que permitem que os modelos animais separem os efeitos genéticos dos ambientais (WOLAK; REID, 2017);
- Coeficiente de endogamia: é de valor limitado, pois são calculados conforme a profundidade da informação do pedigree e, portanto, fazem sentido apenas nas gerações antigas (SPEED; BALDING, 2015), porém influenciam as gerações mais recentes, podendo subestimar a depressão por endogamia (WOLAK; REID, 2017), além de que as contribuições potenciais de ancestrais desconhecidos são ignoradas e as estimativas mais precisas são de animais com linhagens completas em comparação com aqueles de linhagem desconhecida (VANRADEN, 1992). A consanguinidade média real em animais com pais não identificados pode ser semelhante a seus contemporâneos com ambos os pais conhecidos (LUTAAYA et al., 1999);
- Confiabilidade (*reliability*) e precisão (*accuracy*) da avaliação genética: quando os fenótipos dos animais não podem ser associados a fenótipos de parentes, os valores genéticos podem ser tendenciosos refletindo os efeitos ambientais do fenótipo (WOLAK; REID, 2017) e no caso de animais jovens, tenderiam a ter uma avaliação não confiável, pois não apresentam informação completa ou parcial que os conecte com os seus semelhantes e com frequência são descartados (TONUSSI et al., 2021). O viés dos valores genéticos é proporcional ao número de animais sem identificação paterna;

- Variância genética (σ_A^2): a quantificação da relação entre os indivíduos permite a estimação da variância genética a partir de uma decomposição da similaridade fenotípica entre parentes, assim formar uma linhagem populacional (WOLAK; REID, 2017), ao não ter a informação de parentesco, pode comprometer o ganho genético esperado (GÓMEZ et al., 2021);
- Dimensionamento de matrizes: com paternidade incerta o problema de dimensionamento de A e G é previsto, pois A é mais dispersa e menos informativa que G, situação que pode levar a estimativas de parâmetros tendenciosos, dependendo das frequências alélicas assumidas (MISZTAL et al., 2013b).

Para reduzir os efeitos não desejados pela falta de informação na genealogia, diferentes métodos estatísticos foram desenvolvidos, como os grupos genéticos (GG) (FAMULA; POLLAK; VAN VLECK, 1983; PEŠKOVIČOVÁ; GROENEVELD; WOLF, 2004; PETRINI et al., 2015), probabilidade de parentesco, modelo bayesiano empírico iterativo (ITER), modelo hierárquico bayesiano (HIER) (CARDOSO; TEMPELMAN, 2003) e o modelo baseado na matriz de relacionamento médio (ANRM) (HENDERSON, 1988). No entanto, é frequente ignorar esses problemas se a ancestralidade não estiver disponível, e presumir que os animais não são relacionados (VANRADEN, 1992).

No contexto de paternidade desconhecida, os reprodutores múltiplos são empregados tanto em fêmeas que foram inseminadas ou grandes lotes de fêmeas formados para a reprodução (ARAÚJO et al., 2012; CARDOSO et al., 2003; TONUSSI et al., 2017). Nos rebanhos comerciais brasileiros, a porcentagem de animais com pai desconhecido (progênes de touros múltiplos) pode alcançar até 40% e, em muitas propriedades, os touros são utilizados em uma temporada de monta e descartados em seguida, dificultando a seleção dos animais para características de interesse econômico (CARDOSO; TEMPELMAN, 2003; TONUSSI et al., 2017). A preferência pela utilização de touros múltiplos é justificada principalmente pelo fato desse sistema não requerer mão de obra qualificada, sendo uma alternativa de baixo custo comparada com outros sistemas de acasalamento. Além disso, minimiza as dificuldades de manejo comparado ao acasalamento controlado, aumentando os índices de prenhes (OLIVEIRA JÚNIOR et al., 2013; VALLE; ANDREOTTI; THIAGO, 1998).

No entanto, apesar das vantagens deste sistema de manejo, ele não permite a identificação da paternidade das progênes, o que aumenta a perda de informação no pedigree e, em consequência, compromete a avaliação genética (TONUSSI et al., 2017; VALLE, ANDREOTTI, THIAGO, 1998). Porém, métodos tem sido sugeridos para o aumento da acurácia na avaliação genética com pais desconhecidos (TONUSSI et al., 2017; TSURUTA et al., 2014; WESTELL; QUAAS; VAN VLECK, 1988), permitindo o estabelecimento das relações genéticas entre os indivíduos e, por conseguinte, a probabilidade destes animais possuírem genes em comum (PETRINI, 2012).

É importante ressaltar que, na formação dos grupos de reprodutores nos programas de melhoramento genético brasileiros, são utilizados touros avaliados geneticamente, que na sua maioria estão compostos por animais com alto valor genético em características de interesse econômico. Também se deve considerar que, com o aumento substancial da quantidade de animais genotipados nos programa de melhoramento genético, ainda existem os erros de genotipagem, que podem ser atribuídos à genótipos imputados, informações de linhagem ausentes e erros de identificação do animal ou da amostra de DNA, são fatores que impactam na compatibilidade entre as matrizes G e A nos procedimentos de avaliação genética utilizando o ssGBLUP, (MEUWISSEN; HAYES; GODDARD, 2016; TONUSSI et al., 2021) afetando diretamente as estimativas da avaliação (LEGARRA; AGUILAR; MISZTAL, 2009). A genotipagem utilizando painéis de marcadores do tipo SNP de alta densidade pode resolver a maioria dos problemas ligados a paternidade desconhecida, mas só se aplica para animais vivos ou amostras biológicas que estejam disponíveis (GÓMEZ et al., 2021).

Quando as identidades dos pais estão ausentes de forma não aleatória, o que é comum pela pré-seleção de indivíduos, o viés pode ser ainda mais grave (WOLAK; REID, 2017). Mas a atribuição dos pais desconhecidos a grupos genéticos (GG) distintos (método para estabelecimento de relações genéticas entre indivíduos) pode explicar no modelo a falta dessa informação (QUAAS, 1988) tanto em modelos sem e com informação genômica.

Na atualidade, os métodos mais utilizados nos modelos para tentar explicar a paternidade incerta são: a inclusão de um efeito fixo extra, por meio da adição dos grupos de pais desconhecidos (UPG) (QUAAS, 1988), e a utilização da abordagem dos metafundadores (LEGARRA et al., 2015). Na inclusão de efeitos extras no

modelo, não só os UPG ou pais fantasmas (*phantom groups*) podem ser contemplados, também os GG são possíveis soluções (GÓMEZ et al., 2021), pois ao relacioná-los com os animas, a correção para diferentes níveis genéticos é plausível (THERON; KANFER; RAUTENBACH, 2002; WESTELL; QUAAS; VAN VLECK, 1988). Ainda assim, o déficit de informação proveniente do pedigree pode influenciar os parâmetros do modelo animal mesmo com a inclusão dos UPG (WOLAK; REID, 2017). É importante destacar que nenhum dos métodos citados acima para lidar com a deficiência dos pedigrees é amplamente aceita, pois cada abordagem tem implicações diferentes (SPEED; BALDING, 2015).

2.3.1 Grupos de pais desconhecidos (UPG) e modelo misto

Na literatura os grupos genéticos (GG), pais “fantasmas” ou grupos de pais desconhecidos (UPG) são termos que deferem na de estrutura de cada grupo, mas no modelo misto são assumidos na mesma matriz Q como foi proposto por Quaas (1988) e Westell et al. (1988) com o objetivo de modelar as genealogias perdidas, a contribuição genética média para as progênes desses pais desconhecidos e as diferenças genéticas entre grupos de animais na avaliação genética. Em simulações, a formação de UPG visando uma estrutura apropriada para as avaliações dos rebanhos em programas de melhoramento genético animal, demonstraram que os grupos genéticos proporcionam uma melhora na predição e acurácia dos valores genéticos dos animais com paternidade incerta (OLIVEIRA JÚNIOR et al., 2013).

A inclusão de UPG para estimação de componentes de variância e estimação de valores genéticos tem sido adotado na avaliação genética de diferentes espécies, como, por exemplo, bovinos de corte, leite (PETRINI, 2012; PHOCAS; LALOË, 2004; THERON; KANFER; RAUTENBACH, 2002), suínos (PEŠKOVIČOVÁ; GROENEVELD; WOLF, 2004), aves (BERMANN et al., 2021) e ovelhas (MACEDO et al., 2020). O mérito genético de um indivíduo é a soma do efeito médio do pai, o efeito médio da mãe e o efeito médio da subpopulação da qual ele pertence, este é geralmente explicado pelo efeito fixo de grupo genético, mas se esses efeitos fossem ignorados fornece uma medida do viés da previsão (KENNEDY, 1981), mas como os animais são atribuídos a grupos de acordo com um critério específico, a abordagem permite representar o mérito genético médio de um grupo de indivíduos que não tem a informação do seus pais disponível (BERMANN et al., 2021).

Os UPG no modelo animal são definidos como efeitos fixos, pois medem o desvio médio esperado do grupo de referência (que podem ser verdadeiros fundadores, assumindo-se que o efeito médio deles é igual a zero) nos efeitos genéticos aditivos totais de cada grupo, podendo esses efeitos serem estimados explicitamente ou implicitamente (QUAAS, 1988). Wolak e Reid (2017) conceitualizaram mais detalhadamente as duas abordagens, baseados no exposto por Quass (1988). Quando a estimação do efeito do UPG é realizada como efeitos fixos de maneira separada (modelagem explícita) os indivíduos herdaram efeitos dos UPG e é mais recomendado para análises multivariadas (MISZTAL et al., 2013b). Por outro lado, quando é estimado como parte dos efeitos genéticos aditivos totais individuais, temos a modelagem implícita, e, ambas as estimativas são equivalentes do ponto de vista estatístico, gerando idênticos efeitos para os grupos.

A inclusão de UPG no BLUP é uma variação das equações do modelo misto e as soluções deste sistema modificado são funções lineares do grupo e dos valores do indivíduo, mostrando uma relação entre modelo de agrupamento e o modelo de seleção (QUAAS; POLLAK, 1981). No modelo genético básico o EBV (valor genético estimado) é formulado como a média dos seus pais mais o desvio causado pela amostragem mendeliana dos genótipos dos pais durante a gametogênese (QUAAS, 1988).

Com o objetivo de exemplificar como os UPG são incluídos no modelo misto (4), consideramos o modelo misto geral representado como:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{u} + \mathbf{e} \quad (4)$$

onde, \mathbf{y} é o vetor dos fenótipos, \mathbf{b} é o vetor dos efeitos fixos, \mathbf{u} é o vetor dos efeitos genéticos aditivos, \mathbf{X} e \mathbf{Z} são as matrizes de incidência para \mathbf{b} e \mathbf{u} , e \mathbf{e} é o vetor do resíduo. Considerando um modelo infinitesimal, $\text{var}(\mathbf{u}) = \mathbf{A}\sigma_u^2$, onde \mathbf{A} é a matriz de parentesco obtida pela informação do pedigree e σ_u^2 é a variância do efeito genético. E o sistema de equações seria (5):

$$\begin{pmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \mathbf{A}^{-1} \end{pmatrix} \begin{pmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{pmatrix} \quad (5)$$

Para demonstrar a inclusão de UPG na notação mais simples, se assumirá um modelo com UPG visível, ignorando todos os efeitos fixos, exceto dos grupos (FAMULA, 1985; MISZTAL et al., 2013b; QUAAS, 1988) (6):

$$y = \mathbf{ZQ}s + \mathbf{Za} + e \quad (6)$$

Onde s é o efeito de grupo, a é o efeito aditivo do animal, e \mathbf{Q} é a matriz dos animais atribuídos aos grupos. Mais especificamente, cada linha de \mathbf{Q} contém a composição genética do grupo de um animal. Por concordância, os componentes de variância σ_a^2 e σ_e^2 são definidos como unidade. O sistema de equações ficaria (7):

$$\begin{pmatrix} \mathbf{Z}'\mathbf{Z} + \mathbf{A}^{-1} & \mathbf{Z}'\mathbf{Z}\mathbf{Q} \\ \mathbf{Q}'\mathbf{Z}'\mathbf{Z} & \mathbf{Q}'\mathbf{Z}'\mathbf{Z}\mathbf{Q} \end{pmatrix} \begin{pmatrix} \hat{s} \\ \hat{a} \end{pmatrix} = \begin{pmatrix} \mathbf{Z}'\mathbf{y} \\ \mathbf{Q}'\mathbf{Z}'\mathbf{y} \end{pmatrix} \quad (7)$$

Neste modelo o vetor de valores genéticos \mathbf{u} é assumido como $\mathbf{u} = \mathbf{Q}s + \mathbf{a}$, valores computados na solução do sistema de equações, $s = (\mathbf{Q}'\mathbf{A}^{-1}\mathbf{Q})^{-1}\mathbf{Q}'\mathbf{A}^{-1}$ (FAMULA, 1985; QUAAS; POLLAK, 1981), os componentes que são originados dos grupos se ajustariam de acordo com a matriz de parentesco específica do grupo (MUFF et al., 2019) e estes efeitos do grupo dentro de $\hat{\mathbf{u}}$ significam que a incerteza do UPG está incluída dentro da previsão de $\hat{\mathbf{u}}$, aumentando a variância do erro de previsão de $\hat{\mathbf{u}}$ (WOLAK; REID, 2017). É importante levar em consideração que os efeitos do UPG *per se*, como outros efeitos fixos não são estimáveis, mas o que pode-se quantificar é a diferença entre eles, assim, os elementos no vetor \mathbf{u} não são previsíveis mas sim as diferenças (QUAAS, 1988a) (QUAAS, 1988b). Para esclarecer a absorção das equações dos UPG no modelo animal tomamos o mostrado por Westell et al. (1988a) (8):

$$\begin{pmatrix} \mathbf{A}^{00}\mathbf{k} & \mathbf{A}^{01}\mathbf{k} & \mathbf{0} \\ \mathbf{A}^{10}\mathbf{k} & \mathbf{Z}'\mathbf{Z} + \mathbf{A}^{-1} & \mathbf{Z}'\mathbf{Z}\mathbf{Q} \\ \mathbf{0} & \mathbf{Q}'\mathbf{Z}'\mathbf{Z} & \mathbf{Q}'\mathbf{Z}'\mathbf{Z}\mathbf{Q} \end{pmatrix} \begin{pmatrix} \hat{a}_0 \\ \hat{a}_1 \\ \hat{s} \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{Z}'\mathbf{y} \\ \mathbf{Q}'\mathbf{Z}'\mathbf{y} \end{pmatrix} \quad (8)$$

Mostrando-se ilustrativo, pois \mathbf{A}^{00} é o inverso da matriz de parentesco dos relacionamentos entre os UPG, \mathbf{A}^{01} é o inverso da matriz de parentesco dos relacionamentos entre os UPG e $\mathbf{k} = \sigma_e^2/\sigma_a^2$. Estas equações só foram usadas de exemplo, pois se realmente fossem ignorados os efeitos fixos reais as soluções para os efeitos fixos restantes e os efeitos aleatórios seriam tendenciosos (KENNEDY, 1981).

As variâncias e covariâncias dos UPG podem ser definidas como iguais aos parâmetros que correspondem aos pais conhecidos, os UPG podem ser inseridos no método tabular e na matriz inversa para contabilizar as variações modificadas, sendo para os descendentes desses grupos a variação genética especificada

automaticamente (VANRADEN, 1992) e ao que refere-se aos coeficientes de endogamia, a inclusão deles na matriz de relacionamento melhora as avaliações (MATILAINEN et al., 2018). As equações dos UPG podem ser integradas em A^{-1} , essa inserção foi denominado por Quaas e Pollak (1981) como transformação QP, baseados no algoritmo de VanRaden (1992) que contempla a endogamia de animais cujos pais são desconhecidos é igual à endogamia média de seus contemporâneos com pais conhecidos. A transformação QP ocorre na inversa da matriz de parentesco, como segue (9):

$$A_{UPG}^{-1} = \begin{pmatrix} A^{-1} & -A^{-1}Q \\ -Q'A^{-1} & -Q'A^{-1}Q \end{pmatrix} \quad (9)$$

Isto favorece a inclusão dos coeficientes de endogamia de pais desconhecidos (considerado iguais a média de pais conhecido no mesmo período) para explicar o aumento da endogamia e os níveis de relacionamento (VANRADEN, 1992). A matriz QP (10) para as transformações das equações é:

$$QP = \begin{pmatrix} I & 0 & Q & 0 \\ 0 & I & A_{10}Q & 0 \\ 0 & 0 & I & 0 \\ 0 & 0 & 0 & I \end{pmatrix} \quad (10)$$

As equações de forma simplificada do modelo (11) explicado anteriormente, seriam:

$$\begin{pmatrix} A^{00}k & A^{01}k & -Qk \\ A^{10}k & Z'Z + A^{-1}k & 0 \\ -Q'k & 0 & Q'Qk \end{pmatrix} \begin{pmatrix} \hat{a}_{0+Q\hat{s}} \\ \hat{a}_1 + A_{10}Q\hat{s} \\ \hat{s} \end{pmatrix} = \begin{pmatrix} 0 \\ Z'y \\ 0 \end{pmatrix} \quad (11)$$

Westell et al. (1988) apresentam as regras para a construção das diferentes submatrizes que incluem os UPG dentro do modelo com a informação do pedigree, aplicáveis para resolver as equações transformadas de QP, abrindo a porta para a compressão ao incluir a informação genômica.

Há possibilidade de que com o tempo o problema de paternidade incerta seja mitigado pela seleção genômica (GÓMEZ et al., 2021; MACEDO et al., 2020). Na avaliação genômica os UPG podem ser estimados sem fenótipos quando são atribuídos a animais jovens genotipados, pois \mathbf{G} não proporciona relacionamentos desconhecidos (TSURUTA et al., 2019). No ssGBLUP, a matriz A_{22}^{-1} corresponde a a inversa da matriz de parentesco do pedigree somente para os animais genotipados.

Neste modelo se consideram 3 tipos de animais: os ancestrais não genotipados, os indivíduos genotipados com valores genéticos e animais não genotipados com valores genéticos (LEGARRA; AGUILAR; MISZTAL, 2009). Ao incluir os UPG, a matriz H^{-1} será ampliada por meio da transformação QP (12), mas os UPG podem ser considerados só matriz A^{-1} dentro de H^{-1} , apresentado aqui como $H_{UPG_A}^{-1}$:

$$H_{UPG_A}^{-1} = \begin{pmatrix} A^{00} & A^{01} & A^{02} \\ A^{10} & A^{11} & A^{12} \\ A^{20} & A^{21} & A^{22} \end{pmatrix} + \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & G^{-1} - A_{22}^{-1} \end{pmatrix} \quad (12)$$

Ou, considerá-los em toda a matriz H^{-1} , representada como $H_{UPG_H}^{-1}$ (13):

$$H_{UPG_H}^{-1} = \begin{pmatrix} A^{00} & A^{01} & A^{02} \\ A^{10} & A^{11} & A^{12} \\ A^{20} & A^{21} & A^{22} \end{pmatrix} + \begin{pmatrix} Q_2'(G^{-1} - A_{22}^{-1})Q_2 & \mathbf{0} & -Q_2'(G^{-1} - A_{22}^{-1}) \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ -(G^{-1} - A_{22}^{-1})Q_2 & \mathbf{0} & G^{-1} - A_{22}^{-1} \end{pmatrix} \quad (13)$$

Onde, Q_2 é a matriz que contém os UPGs para os animais genotipados (MACEDO et al., 2020; MISZTAL; LOURENCO; LEGARRA, 2020). Alguns autores como Bradford et al. (2019) e Misztal et al. (2020) recomendaram omitir G^{-1} da equação, pois nas populações estudadas pelos pesquisadores, houve melhora na precisão e redução do viés, o que é esperado, pois extingue-se a necessidade de compatibilizar G^{-1} com A_{22}^{-1} . Entretanto, esta abordagem exige maior pesquisa, pois A_{22}^{-1} seria construída assumindo apenas uma população base, provocando à dificuldade de convergência e valores estranhos calculados, ou seja, valores genéticos preditos incorretamente (MATILAINEN et al., 2018).

A utilização de UPG deve ser avaliada considerando várias características da população, banco de dados, genótipos disponíveis, característica e estrutura almeçadas para os UPGs. A preocupação de compatibilidade entre essas as matrizes A e G pode ser reduzida com fatores de escala que alteram geralmente a G (TSURUTA et al., 2011, 2019), essas modificações consegue minimizar vieses nos GEBV que estão presentes também sem o uso de UPGs (MISZTAL et al., 2013b).

A escolha de estruturar os UPGs é um detalhe importante dentro da avaliação, e muitas vezes as definições podem ser arbitrárias (FAMULA, 1985; FAMULA; POLLAK; VAN VLECK, 1983; KENNEDY, 1981). Existem diferentes abordagens de agrupamento (em alguns casos contraditórios) que podem ser testados. Westell et al, (1988) propuseram regras para atribuir os grupos genéticos no modelo animal e sugeriram que o grupo de pais desconhecidos (fantasmas), que não tem fenótipo e são selecionados para serem pais, podem ser definidos como um UPG para mostrar

as diferenças genéticas entre grupos. As diferenças genéticas podem ser criadas através da intensidade da seleção que varia dependendo do sexo, da raça, a linha de seleção, a característica de interesse, o ano, a região e a quantidade disponível de pedigree e informação genômica. Uma dica simples é reduzir o número de gerações no pedigree, pois seria eliminado o efeito dos pais desconhecidos nas gerações mais velhas (MISZTAL et al., 2013b).

As diretrizes do INTERBULL (2001) recomendam o uso de UPG de acordo com a raça, o país de origem, a linha de seleção, a data de nascimento e os UPGs devem ter mínimo de 10 a 20 animais atingindo um tamanho razoável. Entretanto, grupos muito numerosos podem não conseguir captar as diferenças reais entre os indivíduos, (GÓMEZ et al., 2021; PETRINI et al., 2015). Como foi ressaltado previamente por Tsuruta et al. (2014), grupos de pais desconhecidos mal estruturados poderiam proporcionar valores genéticos de pouca confiabilidade e tendências genéticas enganosas. Nesse cenário, os UPG devem-se estruturar de forma cuidadosa para evitar a confusão com os efeitos fixos do modelo (SULLIVAN; SCHAEFFER, 1994).

As regras de agrupamento não são estritas, apenas sugestivas, e a abordagem mais aceita é defini-los por motivação biológica e testar o ajuste dos modelos com diferentes agrupamentos, procurando um equilíbrio entre o número de UPG e o número de ancestrais não identificados (WOLAK; REID, 2017). Um exemplo frequente é o uso da data de nascimento da progênie para modelar as tendências do tempo, podendo-se agregar outras características para estruturar o UPG. Existem diferentes vantagens identificadas com o uso de UPG (com e sem informação genômica) que conseguem apoiar o propósito dos modelos de avaliação genética. As mais discutidas são:

- Superar a incompatibilidade entre as matrizes de relacionamento A e G (BERMANN et al., 2021; MISZTAL et al., 2013a);
- Um efeito poligênico adicional incluído no UPG pode ser adicionado ao modelo (MISZTAL et al., 2013b);
- Para o ssGBLUP os UPG incluem contribuições genômicas e ignorá-las pode apontar previsões tendenciosas (MISZTAL et al., 2013b);
- Os grupos podem descrever subpopulações definíveis que são classificatórios como ano do primeiro parto, reprodutor de inseminação artificial, etc. (KENNEDY, 1981).

A inclusão de UPG apresenta algumas dificuldades, a maioria devido à má estrutura de agrupamento. Entre as desvantagens, podem ser citadas:

- Assume-se que os UPG não são correlacionados e não endogâmicos (QUAAS, 1988), mas, ao considerar coeficientes de consanguinidade e os UPGs, a possível inflação nas previsões genômicas pode ser reduzida (TSURUTA et al., 2019);
- Muitos autores mencionam que ignorar os UPG ou definir grupos poucos informativos adicionaria viés na avaliação genética (HENDERSON, 1974; MISZTAL et al., 2013b; TONUSSI et al., 2017), mas ao adicionar UPG desnecessariamente aumenta-se a variância do erro dos valores genéticos aditivos totais (FAMULA, 1981; WOLAK; REID, 2017);
- Os UPGs poderiam fornecer uma incorreta classificação dos indivíduos, o que ocasionaria uma decisão de seleção subestimada (PHOCAS; LALOË, 2004);
- O viés na solução de UPG aumenta quando o número de animais genotipados é maior (KUDINOV et al., 2020), pois quando o UPG estiver mal definido para animais jovens genotipados, sem progênie ou sem fenótipo, as soluções para o grupo não serão precisas ou não serão estimáveis, o que criaria viés nas previsões genômicas e nas tendências genéticas (TSURUTA et al., 2019);
- Os efeitos do UPG podem ser confundidos com os efeitos de grupo de contemporâneo no modelo, mais ainda, quando os descendentes de um RM estão exclusivamente em um grupo de contemporâneo (CARDOSO; TEMPELMAN, 2003; PEŠKOVIČOVÁ; GROENEVELD; WOLF, 2004);
- O modelo com UPG deve propagar as contribuições genéticas dos indivíduos que foram atribuídos ao UPG, obtendo-se uma matriz Q mais adequada e melhor estruturada (MUFF et al., 2019).

REFERENCIAS

ABCZ. **Apresentação PMGZ - ABCZ - 2019-2**. Disponível em: <https://www.abczstat.com.br/comunicacoes/sumario/apresentacao/Sumario-apresentacao.htm>. Acesso em: 7 nov. 2019.

AGRO-PECUÁRIA CFM. **Sumário de touros Nelore CFM 2020-2021**. São José do Rio Preto, 2020. Disponível em: <http://www.agrocfm.com.br/wp-content/uploads/2020/08/Sumário-Nelore-CFM-2020-2021-Digital.pdf>. Acesso em: 7 nov. 2019.

AGUILAR, I. *et al.* Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. **Journal of Dairy Science**, Lancaster, v. 93, n. 2, p. 743–752, 1 Feb. 2010.

AGUILAR, I. *et al.* Effects of ignoring inbreeding in model-based accuracy for BLUP and SSGBLUP. **Journal of Animal Breeding and Genetics**, Berlin, v. 137, n. 4, p. 356–364, 1 July. 2020.

ARAÚJO, E. P. de *et al.* Comparativo financeiro entre a inseminação artificial e a monta natural na bovinocultura de corte, na fazenda três corações, em Alta Floresta-MT. **Revista Eletrônica da Faculdade de Alta Floresta**, Alta Floresta, v. 1, n. 2238–5479, p. 0–22, 2012.

BARBOSA, P. F. Bovinos-raças puras, novas raças, cruzamentos e compostos de gado de corte. Estratégias para o uso adequado dos recursos genéticos na produção de carne bovina com qualidade. *In*: SEMINÁRIO NACIONAL SOBRE PRODUÇÃO DE CARNE BOVINA COM QUALIDADE, 3., 2003, Gurupi. **Anais [...]**. Gurupi, 2003. p. 124-129.

BERMANN, M. *et al.* Modeling genetic differences of combined broiler chicken populations in single-step GBLUP. **Journal of Animal Science**, Champaign, v. 99, n. 4, p. 1–7, 1 abr. 2021.

BOICHARD, D. *et al.* Genomic selection in domestic animals: principles, applications and perspectives. **Comptes Rendus Biologies**, Paris, v. 339, n. 7–8, p. 274–277, 1 July. 2016.

BOLIGON, A. A. *et al.* Herdabilidades e correlações entre pesos do nascimento à idade adulta em rebanhos da raça Nelore. **Revista Brasileira de Zootecnia**, Viçosa, v. 38, n. 12, p. 2320–2326, 2009.

BONIFAZI, R. *et al.* Impact of sub-setting the data of the main Limousin beef cattle population on the estimates of across-country genetic correlations. **Genetics Selection Evolution**, London, v. 52, n. 1, p. 1–16, 2020.

BRADFORD, H. L. L. *et al.* Modeling missing pedigree in single-step genomic BLUP. **Journal of Dairy Science**, Lancaster, v. 102, n. 3, p. 2336–2346, 2019.

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. **Recursos genéticos**. Disponível em: <https://www.gov.br/agricultura/pt-br/assuntos/sustentabilidade/recursos-geneticos-1>. Acesso em: 27 set. 2021.

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. **Registro de provas zootécnicas e aprovação de projetos de melhoramento genético animal**. Disponível em: <https://www.gov.br/agricultura/pt-br/assuntos/insumos-agropecuarios/insumos-pecuarios/registro-genealogico/aprovacao-do-projeto-para-emissao-do-certificado-especial-de-identificacao-e-producao-ceip>. Acesso em: 27 set. 2021.

CARDOSO, F. F.; TEMPELMAN, R. J. Bayesian inference on genetic merit under

uncertain paternity. **Genetics Selection Evolution**, London, v. 35, n. 5, p. 469–487, 2003.

CARDOSO, V. *et al.* Formando lotes uniformes de reprodutores múltiplos e usando-os em acasalamentos dirigidos, em populações nelore. **Revista Brasileira de Zootecnia**, Viçosa, v. 32, n. 4, p. 834–842, July. 2003.

CARNES, A. A. B.; DAS, I. E. de. Perfil da pecuária no Brasil. **BeefREPORT**, Brasília, p. 49, 2019.

CARVALHEIRO, R. Genomic selection in nelore cattle in Brazil R. *In*: WORLD CONGRESS OF GENETICS APPLIED TO LIVESTOCK PRODUCTION, 10., Vancouver, **Proceedings** [...]. Vancouver, 2014. Disponível em: https://www.asas.org/docs/default-source/wcgalp-posters/935_paper_9354_manuscript_613_0.pdf?sfvrsn=2. Acesso em: 25 set. 2021.

CHEN, C. Y. *et al.* Genome-wide marker-assisted selection combining all pedigree phenotypic information with genotypic data in one step: An example using broiler chickens. **Journal of Animal Science**, Champaign, v. 89, n. 1, p. 23–28, 2011.

CHRISTENSEN, O. F. *et al.* Single-step methods for genomic evaluation in pigs. **Animal**, Cambridge, v. 6, n. 10, p. 1565–1571, 2012.

CHRISTENSEN, O. F.; LUND, M. S. Genomic prediction when some animals are not genotyped. **Genetics Selection Evolution**, London, v. 42, n. 3, p. 2, 2010.

CORBEIL, R. R.; SEARLE, S. R. Restricted maximum likelihood (reml) estimation of variance components in the mixed model. **Technometrics**, Alexandria, v. 18, n. 1, p. 31–38, 1976.

DEKKERS, J. C. M. Application of genomics tools to animal breeding. **Current Genomics**, [s.l.], v. 13, n. 3, p. 207–212, 20 June. 2012.

ELER, J. P. **Teorias e métodos em melhoramento genético animal**: bases do melhoramento genético. Pirassununga: Faculdade de Zootecnia e Engenharia de Alimentos da Universidade de São Paulo, 2017.

ELER, J. P. **Informações pessoais preliminares sobre a avaliação genética do rebanho Nelore pertencente à Agropecuária**. [S.l.]: CFM, 2019.

FALCONER, D. S.; MACKAY, T. F. C. **Introduction to quantitative genetics**. 4. ed. London: ACRIBIA, 1996.

FAMULA, T. R. **Defining gentic groups in dairy sire evaluation**. [S.l.]: Cornell University, 1981.

FAMULA, T. R. An equivalence between models of restricted selection and genetic groups. **Theoretical and Applied Genetics**, Berlin, v. 71, n. 3, p. 413–416, 1985.

FAMULA, T. R.; POLLAK, E. J.; VAN VLECK, L. D. Genetic groups in dairy sire

evaluation under a selection model. **Journal of Dairy Science**, Lancaster, v. 66, n. 4, p. 927–934, 1 Apr. 1983.

FAMULA, T. R. R.; VAN VLECK, L. D. Monte Carlo study of genetic groups in sire evaluation. **Journal of Dairy Science**, Lancaster, v. 65, n. 7, p. 1286–1293, 1 July. 1982.

FOREIGN AGRICULTURAL SERVICE –FAS; UNITED STATES DEPARTMENT OF AGRICULTURE - USDA. **Foreign agricultural service**. Disponível em: <https://www.fas.usda.gov/>. Acesso em: 10 jan. 2022.

FERRAZ, J. B. S.; ELER, J. P. Melhoramento genético para aumento de produtividade em gado de corte no Brasil: a história, o presente e o futuro. *In*: PIRES, A. V. (ed.). **Bovinocultura de corte**. Piracicaba: FEALQ, 2010. p. 763–784.

FERRAZ, J. B. S.; FRIES, L. A. Beef cattle genetic evaluation programs in Brazil. *In*: SIMPÓSIO NACIONAL DA SOCIEDADE BRASILEIRA DE MELHORAMENTO ANIMAL, 5., Pirassununga. **Anais [...]**. Pirassununga, 2004.

FORNI, S.; AGUILAR, I.; MISZTAL, I. Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. **Genetics Selection Evolution**, London. v. 43, n. 1, p. 1–7, 5 Jan. 2011.

GÓMEZ, M. *et al.* Accounting for genetic differences among unknown parents in *Bubalus bubalis*: a case study from the Italian Mediterranean Buffalo. **Frontiers in Genetics**, Lausanne, v. 12, art. 625335, 4 Feb. 2021.

HALEY, C. S.; VISSCHER, P. M. Strategies to utilize marker-quantitative trait loci associations. **Journal of Dairy Science**, Lancaster, v. 81, n. suppl. 2, p. 85–97, 1998.

HENDERSON, C. R. Rapid method for computing the inverse of a relationship matrix. **Journal of Dairy Science**, Lancaster, v. 58, p. 1727–1730, 1975.

HENDERSON, C. R. *et al.* The estimation of environmental and genetic trends from records subject to culling. **Biometrics**, Washington, v. 15, n. 2, p. 192, 1959.

HENDERSON, C. R. General flexibility of linear model techniques for sire evaluation. **Journal of Dairy Science**, Lancaster, v. 57, n. 8, p. 963–972, 1974.

HENDERSON, C. R. Best linear unbiased estimation and prediction under a selection model. **Biometrics**, Washington, v. 31, n. 2, p. 423, 1975a.

HENDERSON, C. R. Use of relationships among sires to increase accuracy of sire evaluation. **Journal of Dairy Science**, Lancaster, v. 58, n. 11, p. 1731–1738, 1 Nov. 1975b.

HENDERSON, C. R. A Simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. **Biometrics**, Washington, v. 32, n. 1, p. 69, Mar. 1976.

HENDERSON, C. R. Use of an average numerator relationship matrix for multiple-sire joining. **Journal of Animal Science**, Champaign, v. 66, n. 7, p. 1614, 1988.

INTERBULL. **Interbull Guidelines for National & International Genetic Evaluation Systems in Dairy Cattle with Focus on Production Traits**. n. November, 2001.

KENNEDY, B. W. Bias and mean square error from ignoring genetic groups in mixed model sire evaluation. **Journal of Dairy Science**, Lancaster, v. 64, n. 4, p. 689–697, 1 Apr. 1981.

KENNEDY, B. W.; MOXLEY, J. E. Comparison of genetic group and relationship methods for mixed model sire evaluation. **Journal of Dairy Science**, Lancaster, v. 58, n. 10, p. 1507–1514, 1975.

KINGHORN, B.; WERF, J. VAN DER; RYAN, M. **Melhoramento animal: uso de novas tecnologias**. Piracicaba: FEALQ, 2016.

KNOTT, S. A. *et al.* Maximum likelihood estimation of genetic parameters in life-history studies using the 'Animal Model'. **Functional Ecology**, [s.l.], v. 9, n. 1, p. 122, 1995.

KUDINOV, A. A. *et al.* Metafounder approach for single-step genomic evaluations of Red Dairy cattle. **Journal of Dairy Science**, Lancaster, v. 103, n. 7, p. 6299–6310, 1 jul. 2020.

LEGARRA, A. *et al.* Multi-breed genetic evaluation in a Gelbvieh population. **Journal of Animal Breeding and Genetics**, Berlin, v. 124, n. 5, p. 286–295, out. 2007.

LEGARRA, A. *et al.* Ancestral relationships using metafounders: finite ancestral populations and across population relationships. **Genetics**, Austin, v. 200, n. 2, p. 455–468, 2015.

LEGARRA, A. Comparing estimates of genetic variance across different relationship models. **Theoretical Population Biology**, New York, v. 107, p. 26–30, 1 Feb. 2016.

LEGARRA, A.; AGUILAR, I.; MISZTAL, I. A relationship matrix including full pedigree and genomic information. **Journal of Dairy Science**, Lancaster, v. 92, n. 9, p. 4656–4663, 1 Sept. 2009.

LÔBO, R. B.; DE BITTNECOURT, T. C. B. dos S. C.; PINTO, L. F. B. Progresso científico em melhoramento animal no Brasil na primeira década do século XXI. **Revista Brasileira de Zootecnia**, Viçosa, v. 39, n. suppl. 1, p. 223–235, jul. 2010.

LOURENCO, D. *et al.* Single-step genomic evaluations from theory to practice: using snp chips and sequence data in blupf90. **Genes**, Basel, v. 11, n. 7, p. 1–32, 14 July. 2020.

LUTAAYA, E. *et al.* Inbreeding in populations with incomplete pedigrees. **Journal of Animal Breeding and Genetics**, Berlin, v. 116, n. 6, p. 475–480, 1 Dec. 1999.

MACEDO, F. L. *et al.* Bias and accuracy of dairy sheep evaluations using BLUP and

SSGBLUP with metafounders and unknown parent groups. **Genetics, Selection, Evolution**, London, v. 52, n. 1, p. 47, 12 ago. 2020.

MACNEIL, M. D. Value of genomics in breeding objectives for beef cattle. **Revista Brasileira de Zootecnia**, Viçosa, v. 45, n. 12, p. 794-801, 2016.

MALAFAIA, G. C. *et al.* A Sustentabilidade na cadeia produtiva da pecuária de corte brasileira. In: **Gestão Estratégica da Sustentabilidade**. [S.L.]: Embrapa, 2019. p. 63–81.

MARCELINO, S. D. DO R.; IEMMA, A. F. Métodos de estimação de componentes de variância em modelos mistos desbalanceados. **Scientia Agricola**, Piracicaba, v. 57, n. 4, p. 643–652, 2000.

MATILAINEN, K. *et al.* Single step genomic evaluation for female fertility in Nordic Red dairy cattle. **Journal of Animal Breeding and Genetics**, Berlin, v. 135, n. 5, p. 337–348, 1 out. 2018.

MEUWISSEN, T. H. E.; HAYES, B. J.; GODDARD, M. E. Prediction of total genetic value using genome-wide dense marker maps. **Genetics**, Austin, v. 157, n. 4, p. 1819–1829, 2001.

MEUWISSEN, T.; HAYES, B.; GODDARD, M. Genomic selection: A paradigm shift in animal breeding. **Animal Frontiers**, [s.l.], v. 6, n. 1, p. 6–14, 1 Jan. 2016.

MISZTAL, I. *et al.* Unknown-parent groups and incomplete pedigrees in single-step genomic evaluation. **Journal of Dairy Science**, Lancaster, v. 96, n. e-supp. 1, 2013a.

MISZTAL, I. *et al.* Unknown-parent groups in single-step genomic evaluation. **Journal of Animal Breeding and Genetics**, Berlin, v. 130, n. 4, p. 252–258, 1 Aug. 2013b.

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

MISZTAL, I.; LOURENCO, D.; LEGARRA, A. Current status of genomic evaluation. **Journal of Animal Science**, Champaign, v. 98, n. 4, p. 1–14, 1 abr. 2020.

MUFF, S. *et al.* Animal models with group-specific additive genetic variances: Extending genetic group models. **Genetics Selection Evolution**, London, v. 51, n. 1, art. 7, 2019.

OLIVEIRA JÚNIOR, G. A. de *et al.* Definição de grupos genéticos aditivos visando melhor predição de valores genéticos em bovinos de corte. **Revista Brasileira de Saúde e Produção Animal**, Salvador, v. 14, p. 277–286, 2013.

PATTERSON, H. D.; THOMPSON, R. Recovery of inter-block information when block sizes are unequal. **Biometrika**, [s.l.], v. 58, n. 3, p. 545–554, 1971.

PEREIRA, R. J. *et al.* Inbreeding depression in Zebu cattle traits. **Journal of Animal**

Breeding and Genetics, Hamburg, v. 133, n. 6, p. 523-533, 2016.

PEŠKOVIČOVÁ, D.; GROENEVELD, E.; WOLF, J. Effect of genetic groups on the efficiency of selection in pigs. **Livestock Production Science**, Amsterdam, v. 88, n. 3, p. 213–222, 1 July. 2004.

PETRINI, J. **Grupos genéticos na eficiência de seleção de bovinos de corte compostos (Bos taurus x Bos indicus)**. 2012. 157 p. Dissertação (Mestrado) - Universidade de São Paulo Escola Superior de Agricultura “Luiz de Queiroz”, Piracicaba, 2012.

PETRINI, J. *et al.* Genetic grouping strategies in selection efficiency of composite beef cattle (Bos taurus x Bos indicus). **Journal of Animal Science**, Champaign, v. 93, n. 2, p. 541–552, 1 Feb. 2015.

PHOCAS, F.; LALOË, D. Should genetic groups be fitted in BLUP evaluation? Practical answer for the French AI beef sire evaluation. **Genetics Selection Evolution**, London, v. 36, n. 3, p. 325–345, 2004.

PINHEROS, C. Melhoramento genético: o que a ciência está Fazendo pela pecuária de corte. *In: Casa da agricultura: bovinocultura de corte*. [s.l.: s.n.], 2011. p. 32–36.

QUAAS, R. L. Additive Genetic Model with Groups and Relationships. **Journal of Dairy Science**, Lancaster, v. 71, n. 5, p. 1338–1345, 1 June. 1988.

QUAAS, R. L.; POLLAK, E. J. Modified Equations for Sire Models with Groups. **Journal of Dairy Science**, Lancaster, v. 64, n. 9, p. 1868–1872, 1 Sept. 1981.

ROBINSON, G. K. That BLUP is a good thing: The estimation of random effects. **Statistical Science**, Hayward, v. 6, n. 1, p. 15–32, 1991.

SPEED, D.; BALDING, D. J. Relatedness in the post-genomic era: Is it still useful? **Nature Reviews Genetics**, London, v. 16, n. 1, p. 33–44, 18 Nov. 2015.

SULLIVAN, P. G.; SCHAEFFER, L. R. Fixed versus random genetic groups. *In: WORLD CONGRESS ON GENETICS APPLIED TO LIVESTOCK PRODUCTION*, 5., 1994, Canada. **Proceedings** [...]. Canada: University of Guelph, 1994.

THERON, H. E.; KANFER, F. H. J.; RAUTENBACH, L. The effect of phantom parent groups on genetic trend estimation. **South African Journal of Animal Sciences**, [s.l.], v. 32, n. 2, p. 130–135, 2002.

THOMPSON, R. Sire evaluation. **Biometrics**, Washington, v. 35, n. 1, p. 339, Mar. 1979.

TONG, A. K. W.; KENNEDY, B. W.; MOXLEY, J. E. Potential errors in sire evaluation from regional genetic differences. **Journal of Dairy Science**, Lancaster, v. 63, n. 4, p. 627–633, 1980.

TONUSSI, R. *et al.* Application of single step genomic BLUP under different uncertain

paternity scenarios using simulated data. **PLoS ONE**, San Francisco, v. 12, n. 9, art. E0181752, 2017.

TONUSSI, R. L. *et al.* Accuracy of genomic breeding values and predictive ability for postweaning liveweight and age at first calving in a Nellore cattle population with missing sire information. **Tropical Animal Health and Production**, Dordrecht, v. 53, n. 4, p. 1–9, 10 Aug. 2021.

TSURUTA, S. *et al.* Multiple-trait genomic evaluation of linear type traits using genomic and phenotypic data in US Holsteins. **Journal of Dairy Science**, Lancaster, v. 94, n. 8, p. 4198–4204, Aug. 2011.

TSURUTA, S. *et al.* Assigning unknown parent groups to reduce bias in genomic evaluations of final score in US Holsteins. **Journal of Dairy Science**, Lancaster, v. 97, n. 9, p. 5814–5821, 2014.

TSURUTA, S. *et al.* Controlling bias in genomic breeding values for young genotyped bulls. **Journal of Dairy Science**, Lancaster, v. 102, n. 11, p. 9956–9970, 1 Nov. 2019.

UNITED STATES DEPARTMENT OF AGRICULTURE - USDA. **FAS home**. Disponível em: <https://apps.fas.usda.gov/psdonline/psdHome.aspx>. Acesso em: 20 mar. 2018.

VALLE, E. R. D.; ANDREOTTI, R.; THIAGO, L. R. L. de S. **Estratégias para aumento da eficiência reprodutiva e produtiva em bovinos de corte**. Campo Grande: Embrapa, 1998.

VANRADEN, P. M. Accounting for inbreeding and crossbreeding in genetic evaluation of large populations. **Journal of Dairy Science**, Lancaster, v. 75, n. 11, p. 3136–3144, 1 Nov. 1992.

VANRADEN, P. M. Efficient methods to compute genomic predictions. **Journal of Dairy Science**, Lancaster, v. 91, n. 11, p. 4414–4423, 1 Nov. 2008.

WESTELL, R. A.; QUAAS, R. L.; VAN VLECK, L. D. Genetic groups in an animal model. **Journal of Dairy Science**, Lancaster, v. 71, n. 5, p. 1310–1318, 1988.

WOLAK, M. E.; REID, J. M. Accounting for genetic differences among unknown parents in microevolutionary studies: how to include genetic groups in quantitative genetic animal models. **Journal of Animal Ecology**, v. 86, n. 1, p. 7–20, 1 Jan. 2017.

WRIGHT, S. Coefficients of inbreeding and relationship. **The American Naturalist**, Chicago, v. 56, n. 645, p. 330–338, 1922.

3 AVALIAÇÃO DO USO DE GRUPOS DE PAIS DESCONHECIDOS EM CARACTERÍSTICAS SIMULADAS DE MODERADA E BAIXA HERDABILIDADE

A.S. Acero Valderrama¹, R. Espigolan¹, B.C. Abreu Silva², F.O. Bussiman³, F.S. dos Santos¹, J.B.S. Ferraz¹, E.C. Mattos¹, J.P. Eler¹

¹*Departamento de Medicina Veterinária, Faculdade de Zootecnia e Engenharia de Alimentos da Universidade de São Paulo (ZMV-FZEA/USP), Pirassununga, SP, Brasil.*

²*Departamento de Zootecnia, Cidade Universitária Prof. José Aloísio de Campos da Universidade Federal de Sergipe, São Cristóvão, SE, Brasil.*

³*Departamento de Nutrição e Produção Animal, Faculdade de Medicina Veterinária e Zootecnia da Universidade de São Paulo (VNP-FMVZ/USP), Pirassununga, SP, Brasil.*

*Autor correspondente: alisson.s.acero@gmail.com

Resumo

Objetivou-se avaliar o uso de grupos de pais desconhecidos (*unknown parent groups*, UPG) nas avaliações genéticas, utilizando dados fenotípicos e genotípicos simulados para características de baixa e moderada herdabilidade (0,10 e 0,35, respectivamente). A população foi simulada no software QMSIM versão 1.10 (SARGOLZAEI; SCHENKEL, 2009) com 5 réplicas, testando-se cenários com 30%, 50% e 70% de informação faltante do pai no pedigree. Os UPG foram definidos com informação de geração (UPGY, ssUGPY), e a combinação da geração e o sexo do animal (UPGS, ssUPGS). Eles foram adicionados somente na matriz de parentesco baseada nas informações de pedigree (**A**) ou na matriz de parentesco combinada (**H**). Os dados simulados foram analisados utilizando os métodos BLUP (*Best Linear Unbiased Prediction*) e ssGBLUP (*single-step Genomic Best Linear Unbiased Prediction*). Os valores genéticos genômicos estimados (GEBV) no modelo ssUPGY_A em todos os cenários e nas duas populações apresentaram os piores valores do coeficiente de regressão do GEBV sobre o TBV (β_1), diferença média absoluta (DMA) e coeficiente de correlação (r). Os outros modelos indicaram que o uso de UPG propicia menor deflação quando comparado aos GEBV estimados por meio do BLUP e do ssGBLUP. Os resultados encontrados sugerem que os UPG com informação genômica, corrobora com o aumento da acurácia e diminuição do viés dos GEBVs. Os ajustes na estrutura dos UPG provocaram pouco impacto nos GEBV. Os resultados sugerem considerar os UPG em toda a matriz H no modelo ssGBLUP.

Palavras chave: BLUP. Paternidade desconhecida. QMSim. Simulação. Single-step. UPG.

Abstract

This study aimed to evaluate the use of unknown parent groups (UPG) in genetic evaluations, using simulated phenotypic and genotypic data of low and moderate heritability traits (0.10 and 0.35, respectively). Population was simulated using QMSIM software version 1.10 (SARGOLZAEI; SCHENKEL, 2009) with 5 replicates, these tested in scenarios of 30%, 50% and 70% of missing information from the sires in the pedigree. The UPGs were defined with generation information (UPGY, ssUGPY), and the combination of generation and sex of the animal (UPGS, ssUPGS). They were added only in the relationship matrix from pedigree information (A) or in the combined relationship matrix H. The simulated data were analyzed using the BLUP (Best Linear Unbiased Prediction) and ssGBLUP (single-step Genomic Best Linear Unbiased Prediction) models. The predicted breeding values (GEBV) in the ssUPGY_A model in all scenarios and in both heritabilities showed the worst values of the regression coefficient of GEBV on TBV (β_1), mean absolute difference (DMA) and correlation coefficient (r). The other models indicated that the use of UPG has lower deflation compared to those GEBV estimated with BLUP and ssGBLUP. The results found suggest that the UPG, with genomic information, increases the accuracy and decreases the bias of GEBVs. The adjustment of genetic groups had a small impact on (G)EBV and our results show that considering the UPGs across the H matrix was the best approach.

Keyword: BLUP. QMSim. Simulation. Single-step. Unknown paternity. UPG.

3.1 Introdução

Uma população tem características especiais dependendo da espécie, tipo de reprodução, fisiologia, gerações, fenótipos, etc. por sua vez cada característica também tem um comportamento particular dentro da população, e no caso das espécies com objetivo produtivo procura-se animais com o melhor desempenho. Do ponto de vista genético, procura-se genes favoráveis nos pais para serem herdados pelos filhos, mas é possível que muitos genes afetem cada uma das características comercialmente importantes, sendo esses genes responsáveis por uma fração significativa da variação genética (KINGHORN; WERF; RYAN, 2016). A forma de medir essa fração devida a genética, é por meio da estimação de parâmetros genéticos, principalmente a variância genética aditiva (σ_A^2) baseada no grau de similaridade entre parentes, e a herdabilidade (h^2) que é a proporção da superioridade dos pais que é transmitida para a próxima geração, estatisticamente definida como o coeficiente entre a variância genética aditiva (σ_A^2) e a variância fenotípica (σ_P^2) (FALCONER; MACKAY, 1996; KINGHORN; WERF; RYAN, 2016). A herdabilidade é um parâmetro não estático e muda à medida que a população evolui, difere entre

características, espécies, populações dentro das espécies e ao longo do tempo. Herdabilidades de maior magnitude indicam maior diferenças observadas entre os fenótipos, explicada pelos parâmetros genéticos (HILL, 2001; KINGHORN; WERF; RYAN, 2016).

Nas avaliações genéticas, a paternidade desconhecida pode comprometer a precisão das estimativas dos valores genéticos, dificultando a seleção. A informação do relacionamento dos indivíduos uns com outros é parte fundamental do modelo animal, este parentesco é obtido a partir dos diferentes dados disponíveis: pedigree, dados genômicos ou a sua combinação (MISZTAL; LEGARRA; AGUILAR, 2009; MUFF et al., 2019). Na realidade, bancos completos não existem, e a falta de informações sobre linhagem parentesco entre os animais leva a uma série de dificuldades e consequências, tanto na avaliação dos animais como na escolha dos melhores exemplares para a reprodução e construir a nova geração melhorada com características desejáveis para a produção.

A genotipagem utilizando painéis de marcadores SNP de alta densidade pode resolver a maioria dos problemas ligados a paternidade desconhecida, mas só se aplica para animais vivos ou para amostras biológicas que estejam disponíveis (GÓMEZ et al., 2021). Nesse sentido, a inclusão dos dados de grupos de pais desconhecidos (*unknown parent groups*, UPG) consiste em uma alternativa que permite o estabelecimento das relações genéticas entre os indivíduos pertencentes ao mesmo grupo, proporcionando estimativas mais acuradas dos valores genéticos (QUAAS, 1988) tanto em modelos sem ou com informação genômica.

O déficit da informação do pedigree pode influenciar os parâmetros do modelo animal mesmo se UPG são incluídos (WOLAK; REID, 2017), e mais ainda, nenhuma abordagem foi aceita de forma geral, pois cada um deles tem implicações diferentes (SPEED; BALDING, 2015). Em estudos de simulação, a formação de grupos genéticos leva a melhora na acurácia de predição dos valores genéticos dos animais com paternidade desconhecida (OLIVEIRA et al., 2013). Por outro lado, Phocas e Laloë (2004) informaram que a inclusão de UPG pode apontar a uma incorreta classificação dos animais e a uma decisão de seleção subestimada. A contradição destes estudos e as condições específicas de estrutura condicionada da característica, despertam à necessidade de avaliar o uso de UPG na avaliação genética genômica em características com diferentes herdabilidades de diferentes magnitudes.

3.2 Material e métodos

Os dados foram simulados utilizando o programa QMSim versão 1.10 (SARGOLZAEI; SCHENKEL, 2009) de maneira que a população simulada tivesse uma extensão e um padrão de desequilíbrio de ligação (LD) conforme ao verificado em populações de Bovinos de corte *Bos Indicus*. As simulações foram realizadas utilizando parâmetros relacionados às gerações históricas, definidos de forma semelhante ao descrito por Brito et al. (2011).

3.2.1 População Simulada

A população histórica foi gerada com 1.000 gerações contendo 1.000 indivíduos como tamanho populacional efetivo. A partir daí, 1200 gerações históricas foram simuladas nas quais o número de animais foi gradualmente reduzido de 1.000 para 200, garantindo que o equilíbrio de mutação e deriva fosse estabelecido e que o desequilíbrio de ligação inicial fosse gerado. Nesses procedimentos, o número de indivíduos de cada sexo era o mesmo e o sistema de acasalamento era baseado na união aleatória dos gametas. A população foi expandida selecionando-se aleatoriamente 100 fundadores do sexo masculino e 100 fundadores do sexo feminino da última geração da população histórica. Posteriormente, foram selecionadas as últimas 8 gerações da população histórica, com cinco filhos por mãe e um crescimento exponencial do número de mães, também sob união aleatória de gametas e sem seleção para serem os fundadores da população recente (400 machos e 10.000 fêmeas).

A população recente foi composta por 15 gerações, com um tamanho populacional constante de 10.000 indivíduos, com acasalamentos que produziram progênie com chances iguais de serem machos ou fêmeas. Em cada geração, as taxas de substituição do touro e da vaca foram de 0,6 e 0,4, respectivamente. Para as análises do presente estudo, foram selecionados os indivíduos pertencentes às cinco últimas gerações, totalizando 50.000 animais com informações de parentesco, informação fenotípica e valor genético verdadeiro (TBV), considerando uma característica de moderada e baixa herdabilidade (0,10 e 0,35, respectivamente).

3.2.2 Genoma simulado

O genoma simulado consistiu em 735.293 marcadores e 1.000 QTLs distribuídos aleatoriamente em 29 cromossomos autossômicos com comprimentos semelhantes aos autossomos bovinos (genoma de referência ARS UCD 1.2) e SNPs uniformemente distribuídos. Foi assumido que os QTLs explicam 100% da variância genética. O número de marcadores e QTLs por cromossomo variou de 12.931 a 46.495 e de 121 a 438, respectivamente. Todos os marcadores eram bi alélicos e para QTLs a quantidade de alelos por *loci* variou aleatoriamente de 2 a 4, considerando que os efeitos dos alelos foram amostrados a partir de uma distribuição gama com um parâmetro de forma igual a 0,4 (MEUWISSEN; HAYES; GODDARD, 2001).

A fim de estabelecer o equilíbrio de mutação e deriva genética em gerações históricas, as taxas de erro de genotipagem de marcador e mutação recorrente (para marcadores e QTL) foram 0,005 e 10^{-4} , respectivamente. Considerando frequências de alelos menores (MAF) maiores ou iguais a 0,01, um total de 430.000 marcadores foram selecionados aleatoriamente da última geração da população histórica para gerar dados genotípicos para a população de seleção. O verdadeiro valor genético (TBV) de cada indivíduo correspondeu a soma dos efeitos de substituição do alelo QTL. Os fenótipos foram gerados pela adição de resíduos, sorteados aleatoriamente a partir de uma distribuição normal com média igual a zero, aos TBVs. Todo o processo de simulação foi replicado 5 vezes. Na tabela 1 se apresentam os parâmetros dos processos de simulação.

Tabela 1- Opções e parâmetros usados nos processos de simulação de dados

Opções para o processo de simulação	Parâmetros
População Histórica (HP)	
Fase 1	1000 gerações / 1000 animais
Fase 2	1020 gerações / 1000 a 200 animais
População expandida (EP)	
Número de fundadores machos para HP (Fase 2)	100
Número de fundadores fêmeas para HP (Fase 2)	100
Número de gerações	8
Número de filhos por cada fêmea	5
População Recente (RP)	
Número de fundadores machos da EP	400

Número de fêmeas da EP	10.000
Número de gerações	15
Número de filhos por cada fêmea	1
Proporção de machos na progênie	0,5
Acasalamento	Aleatório
Taxa de substituição do touro	0,6
Taxa de substituição da vaca	0,4
Seleção	EBV
Modelo de predição genética	Modelo animal - BLUP
Herdabilidade do fenótipo	0,10 e 0,35
Variância fenotípica	1.0

Genoma

Número de cromossomos	29
Cumprimento total	2.333cM
Número de marcadores	735.293
Distribuição dos marcadores	uniformemente distribuídos
Número de QTL	1.000
Distribuição dos QTL	Aleatório
MAF para os marcadores	0,01
MAF para os QTL	0,01
Efeitos alélicos aditivos para QTL	Distribuição Gama (0,4)
Taxa de genótipos de marcadores ausentes	0
Taxa de erro de genotipagem de marcador	0,005
Taxa de mutação recorrente	0,0001

Fonte: autoria própria

O número de marcadores do tipo SNPs simulados foi de 429.127, e um total de 349.989 passaram no controle de qualidade e permaneceram disponíveis para as análises. O coeficiente médio de endogamia foi de 0,04, com máximo de 0,46.

3.2.3 Modelos

A inclusão de UPG no BLUP é uma variação das equações do modelo misto e as soluções deste sistema modificado são funções lineares do grupo e os valores do indivíduo, mostrando uma relação entre modelo de agrupamento e o modelo de seleção (QUAAS; POLLAK, 1981). No modelo genético básico o EBV é formulado como a média dos seus pais mais o desvio causado pela amostragem mendeliana dos genótipos dos pais durante a gametogênese, e procurasse os ancestrais no arquivo de linhagem até os não identificados e esse pai é considerado um pai base

sendo a parada do ciclo (QUAAS, 1988). Os UPG são incluídos no modelo misto (1), consideramos o modelo misto geral representado como:

$$y = Xb + Zu + e \quad (1)$$

Onde, y é o vetor dos fenótipos, b é o vetor dos efeitos fixos de geração e sexo, u é o vetor dos efeitos genéticos aditivos, X e Z são as matrizes de incidência para b e u , e e é o vetor do resíduo. Considerando um modelo infinitesimal, $\text{var}(u) = A\sigma_u^2$, onde A é a matriz de parentesco obtida pela informação do pedigree e σ_u^2 é a variância do efeito genético. Para demonstrar a inclusão de UPG na notação mais simples, se assumirá um modelo com UPG visível, ignorando todos os efeitos fixos (2), exceto dos grupos (FAMULA, 1985; MISZTAL et al., 2013; QUAAS, 1988):

$$y = ZQs + Za + e \quad (2)$$

Onde s é o efeito de grupo explicado por $s = (Q'A^{-1}Q)^{-1}Q'A^{-1}$ (FAMULA, 1985; QUAAS; POLLAK, 1981), a é o efeito aditivo do animal, e Q é a matriz dos animais atribuídos aos grupos. Mais especificamente, cada linha de Q contém a composição genética do grupo de um animal. Por concordância, os componentes de variância σ_a^2 e σ_e^2 são definidos como unidade.

As equações dos UPG podem ser integradas em A^{-1} no modelo BLUP, essa inserção foi denominado por Quaas e Pollak (1981) como transformação QP, baseados no algoritmo de VanRaden (1992) que contempla a endogamia de animais cujos pais são desconhecidos é igual à endogamia média de seus contemporâneos com pais conhecidos. A transformação QP (3) ocorre na inversa da matriz de parentesco assim:

$$A_{UPG}^{-1} = \begin{pmatrix} A^{-1} & -A^{-1}Q \\ -Q'A^{-1} & -Q'A^{-1}Q \end{pmatrix} \quad (3)$$

No ssGBLUP, a inversa da matriz H é apresentada assim (4) (AGUILAR et al., 2010; CHRISTENSEN; LUND, 2010):

$$H^{-1} = A^{-1} + \begin{pmatrix} 0 & 0 \\ 0 & G^{-1} - A_{22}^{-1} \end{pmatrix} \quad (4)$$

Onde A_{22}^{-1} é a inversa da matriz de parentesco do pedigree somente para os animais genotipados. Ao incluir UPG a matriz H^{-1} será aumentada usando a transformação QP, mas os UPGs podem ser considerados só matriz A^{-1} dentro de H^{-1} (5), apresentado aqui como $H_{UPG,A}^{-1}$:

$$H_{UPG_A}^{-1} = \begin{pmatrix} A^{00} & A^{01} & A^{02} \\ A^{10} & A^{11} & A^{12} \\ A^{20} & A^{21} & A^{22} \end{pmatrix} + \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & G^{-1} - A_{22}^{-1} \end{pmatrix} \quad (5)$$

Ou, considerá-los em toda a matriz H^{-1} , representada como $H_{UPG_H}^{-1}$ assim

(6):

$$H_{UPG_H}^{-1} = \begin{pmatrix} A^{00} & A^{01} & A^{02} \\ A^{10} & A^{11} & A^{12} \\ A^{20} & A^{21} & A^{22} \end{pmatrix} + \begin{pmatrix} Q_2'(G^{-1} - A_{22}^{-1})Q_2 & \mathbf{0} & -Q_2'(G^{-1} - A_{22}^{-1}) \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ -(G^{-1} - A_{22}^{-1})Q_2 & \mathbf{0} & G^{-1} - A_{22}^{-1} \end{pmatrix} \quad (6)$$

Onde, Q_2 é a matriz que contém os UPGs para os animais genotipados (MACEDO et al., 2020; MISZTAL; LOURENCO; LEGARRA, 2020)

3.2.4 Cenários

Os cenários propostos consistiram em excluir aleatoriamente a informação do pai no pedigree simulado, nas seguintes proporções: 30%, 50% e 70%, nas mães não houve alteração na estrutura de pedigree, pois o comportamento reprodutivo e de manejo na população é mais frequente, havendo assim maiores informações descartando a necessidade de grupos genéticos para essa categoria (WOLAK; REID, 2017). As avaliações genéticas envolveram análises, sem informação genômica, por meio do BLUP tradicional com a estrutura de pais desconhecidos formados pela geração do animal (UPGY) e pela combinação do sexo e geração do animal (UPGS). Com genômica, as análises utilizaram o procedimento de passo único (ssGBLUP; *single-step Best Linear Unbiased Prediction*), com a estrutura de pais desconhecidos, grupos de pais desconhecidos formados pela geração do animal, inclusos na matriz **A** ou na matriz **H** (ssUPGY_A e ssUPGY_H) e grupos de pais desconhecidos formados por sexo e geração do animal inclusos na matriz **A** ou na matriz **H** (ssUPGS_A e ssUPGS_H), se apresentam na tabela 2. As avaliações genéticas foram realizadas com auxílio dos programas da família BLUPF90 (MISZTAL et al., 2015).

Tabela 2- Cenários para os diferentes grupos de pais desconhecidos (UPG) estruturados para a herdabilidade baixa (0,10) e moderada (0,35)

Modelo	Estrutura dos UPG	Abreviações
BLUP	Sem agrupamento	BLUP
	Geração	UPGY
	Geração e sexo	UPGS
ssGBLUP	Sem agrupamento	ssGBLUP
	Geração na matriz A	ssUPGY_A
	Geração e sexo na matriz A	ssUPGS_A
	Geração na matriz H	ssUPGY_H
	Geração e sexo na matriz H	ssUPGS_H

Abreviações: BLUP = *best linear unbiased prediction*; ssGBLUP = *single-step genomic best linear unbiased prediction*; UPG = Grupo de pas desconhecidos.

Fonte: autoria própria

3.2.5 Análise dos resultados

Para cada modelo e cenário os valores genéticos (genômicos) foram avaliados por meio do coeficiente de regressão do GEBV sobre o TBV (β_1), pela diferença média absoluta (DMA) e pelo coeficiente de correlação (7)

$$DMA = \frac{1}{n} \sum_{t=1}^n |\hat{y}_t - y_t| \quad (7)$$

E coeficiente de correlação (r) (8):

$$r = \frac{cov(\hat{y}_t, y_t)}{\sqrt{V(\hat{y}_t)V y_t}} \quad (8)$$

Em que, y_t é o valor do TBV, \hat{y}_t é o valor do GEBV e n é o número de dados.

3.3 Resultados e discussão

De maneira geral, os modelos que consideraram UPG e a informação genômica produziram GEBV mais acurados e menos viesados. Com o aumento da porcentagem de paternidade desconhecida, foi observado a correlação e aumento do viés (tabelas 3 e 4), apresentado também na observação da regressão dos GEBV com

o TBV nas figuras 1, 2, 3, 4, 5 e 6, que pode ser interpretada como uma medida da acurácia pois mostra a correlação entre os GEBVs estimados e o TBV. Resultados similares resultados similares aos reportados por Oliveira et al. (2013).

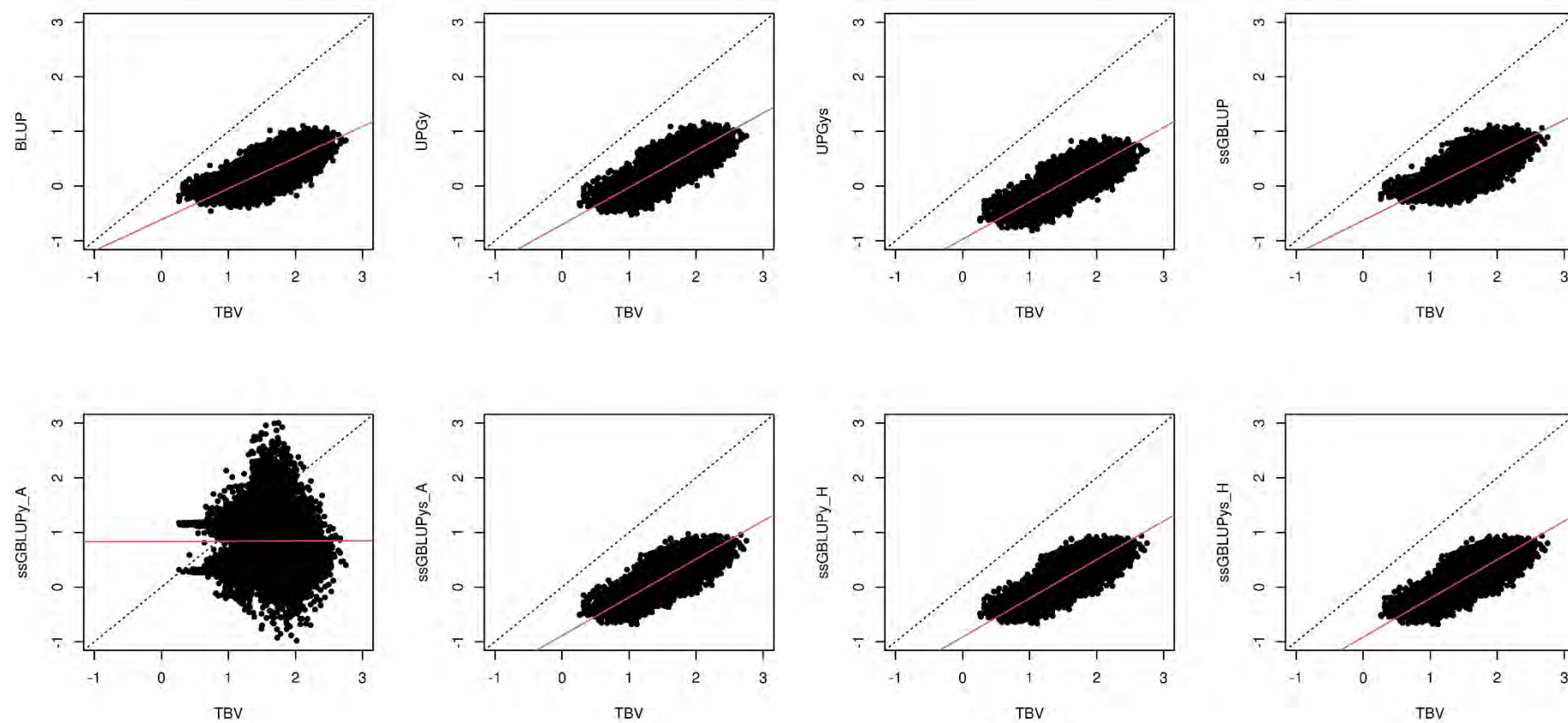
Tabela 3- Viés (β_1), diferença média absoluta (DMA) e acurácia (r) dos GEBV para herdabilidade baixa (0,1) nas diferentes porcentagens de pais desconhecidos

Modelos	Porcentagem de pais desconhecidos								
	30%			50%			70%		
	β_1	DMA	r	β_1	DMA	r	β_1	DMA	r
BLUP	0,57	1,26	0,77	0,55	1,31	0,74	0,47	1,37	0,68
UPGY	0,68	1,19	0,81	0,64	1,55	0,77	0,52	1,61	0,69
UPGS	0,68	1,45	0,81	0,64	1,55	0,77	0,53	1,23	0,70
ssGBLUP	0,61	1,21	0,78	0,58	1,28	0,76	0,50	1,36	0,70
ssUPGY_A	0,49	83,01	0,00	0,65	1,43	0,80	0,58	1,40	0,75
ssUPGS_A	0,70	1,34	0,82	0,65	1,43	0,80	0,59	1,41	0,75
ssUPGY_H	0,71	1,35	0,82	0,67	1,46	0,80	0,56	1,53	0,73
ssUPGS_H	0,71	1,35	0,82	0,67	1,46	0,80	0,57	1,54	0,74

Abreviações: BUP = pais desconhecidos na matriz **A**; UPGY = UPG considerando geração na matriz **A**; UPGS = UPG considerando sexo e geração na matriz **A**; ssGBLUP: ssGBLUP com pais desconhecidos; ssUPGY_A: ssGBLUP com UPG considerado a geração na matriz **A**. ssUPGS_A: ssGBLUP com UPG considerando sexo e geração na matriz **A**. ssUPGY_H: ssGBLUP com UPG considerando a geração na matriz **H**. ssUPGS_H: ssGBLUP com UPG considerando sexo e geração na matriz **H**.

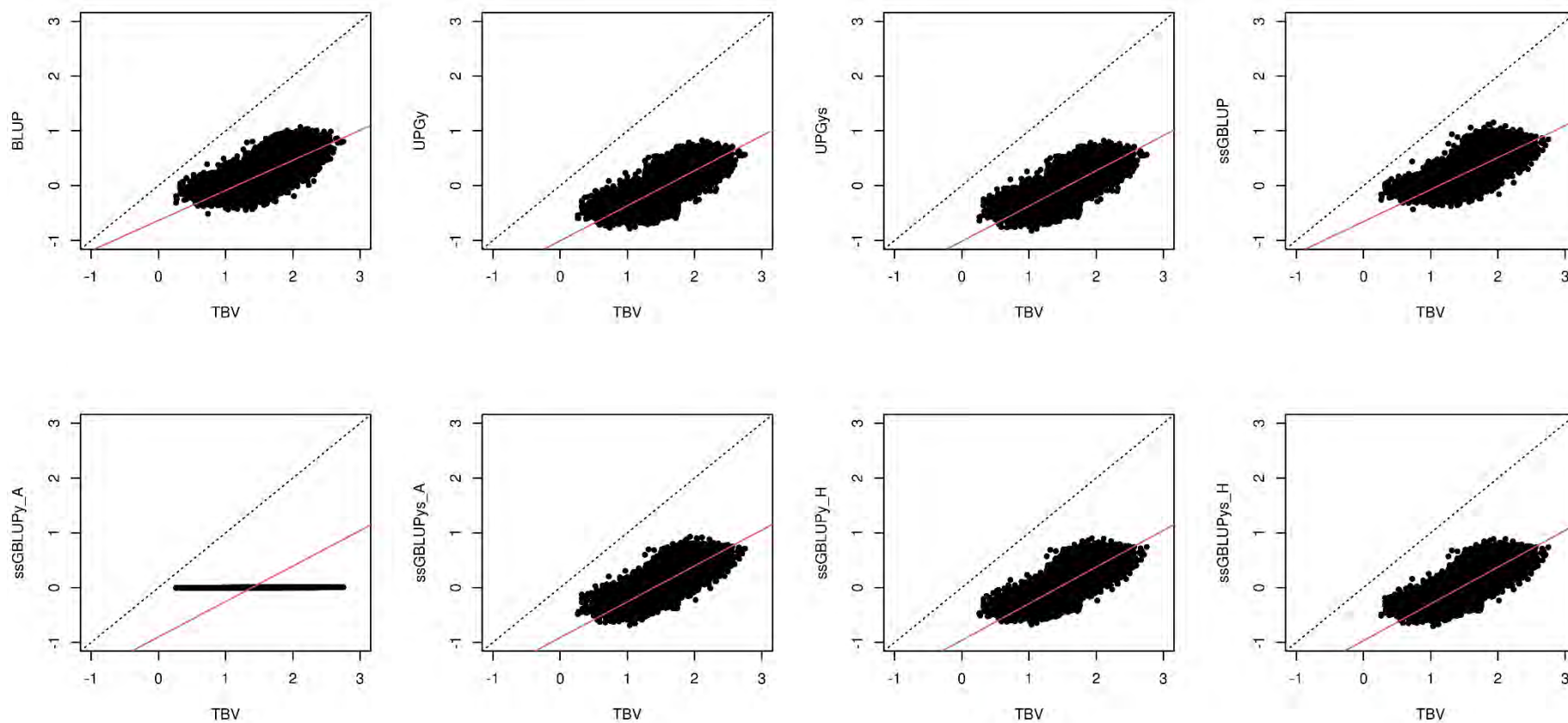
Fonte: autoria própria

Figura 1- Regressão dos valores genéticos preditos nas diferentes avaliações com o valor genético verdadeiro (TBV) para herdabilidade baixa (0,10) e com o 30% de pais desconhecidos



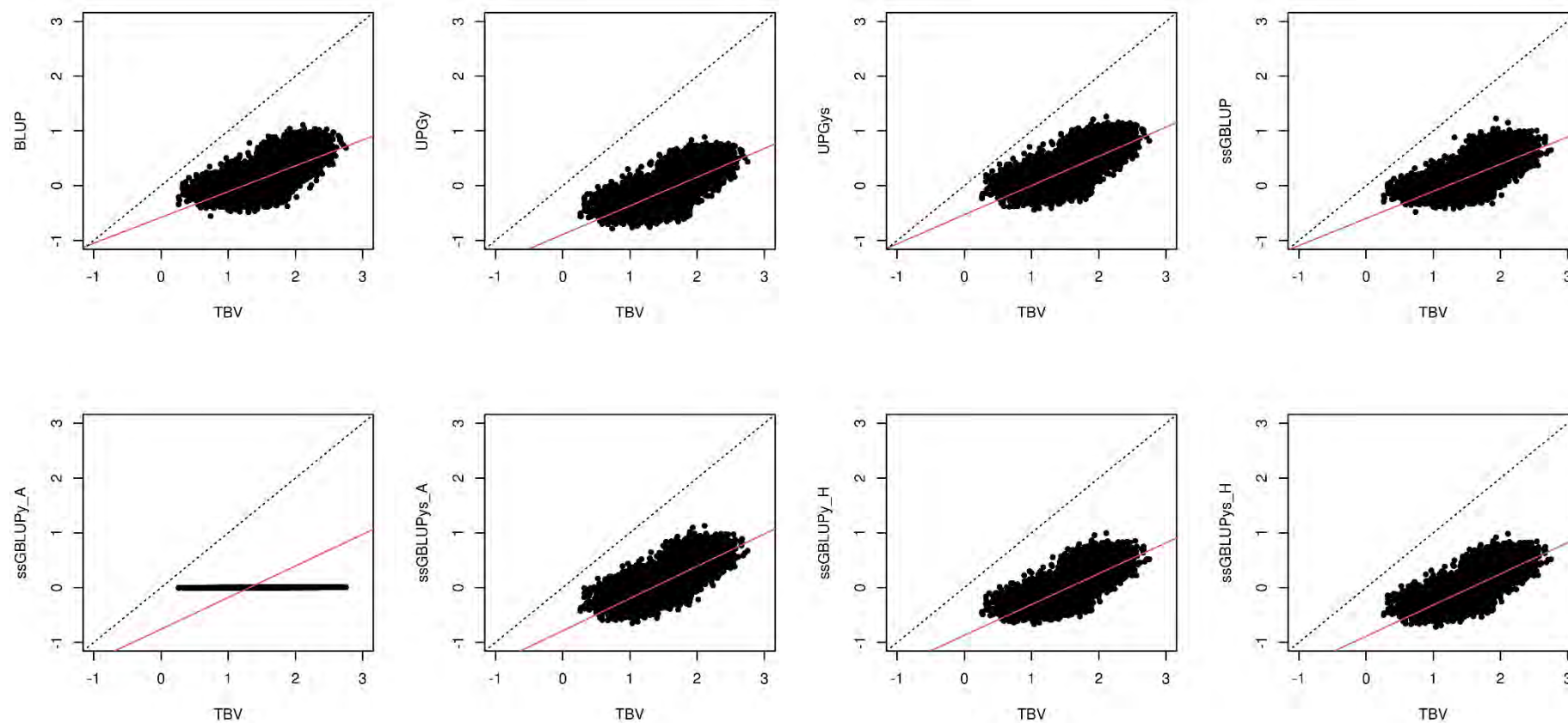
Fonte: autoria própria.

Figura 2- Regressão dos valores genéticos preditos nas diferentes avaliações com o valor genético verdadeiro (TBV) para herdabilidade baixa (0,10) com o 50% de pais desconhecidos



Fonte: autoria própria.

Figura 3- Regressão dos valores genéticos preditos nas diferentes avaliações com o valor genético verdadeiro (TBV) para herdabilidade baixa (0,10) com o 70% de pais desconhecidos



Fonte: autoria própria.

Para a característica de baixa herdabilidade o maior valor de DMA (83,01) foi encontrado no cenário de 30% no modelo de ssUPGY_A, caracterizando o pior modelo pois tanto r e β_1 foram baixos (0,00 e 0,49), demonstrado na dispersão dos valores na figura 1, com uma acurácia próxima a zero. Entre os modelos que consideraram a informação genômica, as diferenças foram sutis, exceto para o ssUPGY_A no cenário com 30% de pais desconhecidos, com problemas de convergência que afetaram os valores da diagonal da matriz **H**, gerando r igual a zero. Para os modelos genômicos, foram observados valores de r altos e β_1 próximo de 1, quando comparados com o ssGBLUP em cada cenário, no entanto, o DMA é menor para o ssGBLUP como foi relatado por Matilainen et al., (2018) para característica de fertilidade em gado leitero. Somente no cenário de 70% de pais desconhecidos, os modelos ssUPGY_A e ssUPGS_A apresentaram valores de β_1 , DMA e r melhores que os modelos ssUPGY_H e ssUPGS_H, visto que, nos cenários de 30% e 50% estes mostraram melhores β_1 , DMA e r . Como demonstrado nas figuras 2 e 3, a regressão do modelo ssUPGY_A descreve inferioridade quanto comparado ao TBV. Os resultados do presente trabalho evidenciaram que, para uma característica de baixa herdabilidade (0,1) não houve diferença relacionada à utilização de UPGY ou UPGS.

Tabela 4- Viés (β_1), diferença média absoluta (DMA) e acurácia (r) dos GEBV para herdabilidade moderada (0,35) nas diferentes porcentagens de pais desconhecidos

Modelos	Porcentagem de pais desconhecidos								
	30%			50%			70%		
	β_1	DMA	r	β_1	DMA	r	β_1	DMA	r
BLUP	0,63	4,54	0,85	0,59	4,68	0,81	0,54	4,82	0,76
UPGY	0,76	4,89	0,89	0,72	5,57	0,87	0,64	5,04	0,83
UPGS	0,76	4,91	0,89	0,72	5,39	0,87	0,64	5,14	0,82
ssGBLUP	0,63	4,49	0,86	0,58	4,66	0,82	0,53	4,81	0,77
ssUPGY_A	1,83	3,94	0,74	2,18	5,13	0,71	3,01	9,58	0,69
ssUPGS_A	0,79	4,72	0,90	0,72	4,96	0,89	1,19	2,63	0,66
ssUPGY_H	0,79	4,72	0,90	0,72	4,98	0,89	0,65	5,02	0,85
ssUPGS_H	0,79	4,72	0,90	0,72	4,98	0,89	0,65	5,04	0,85

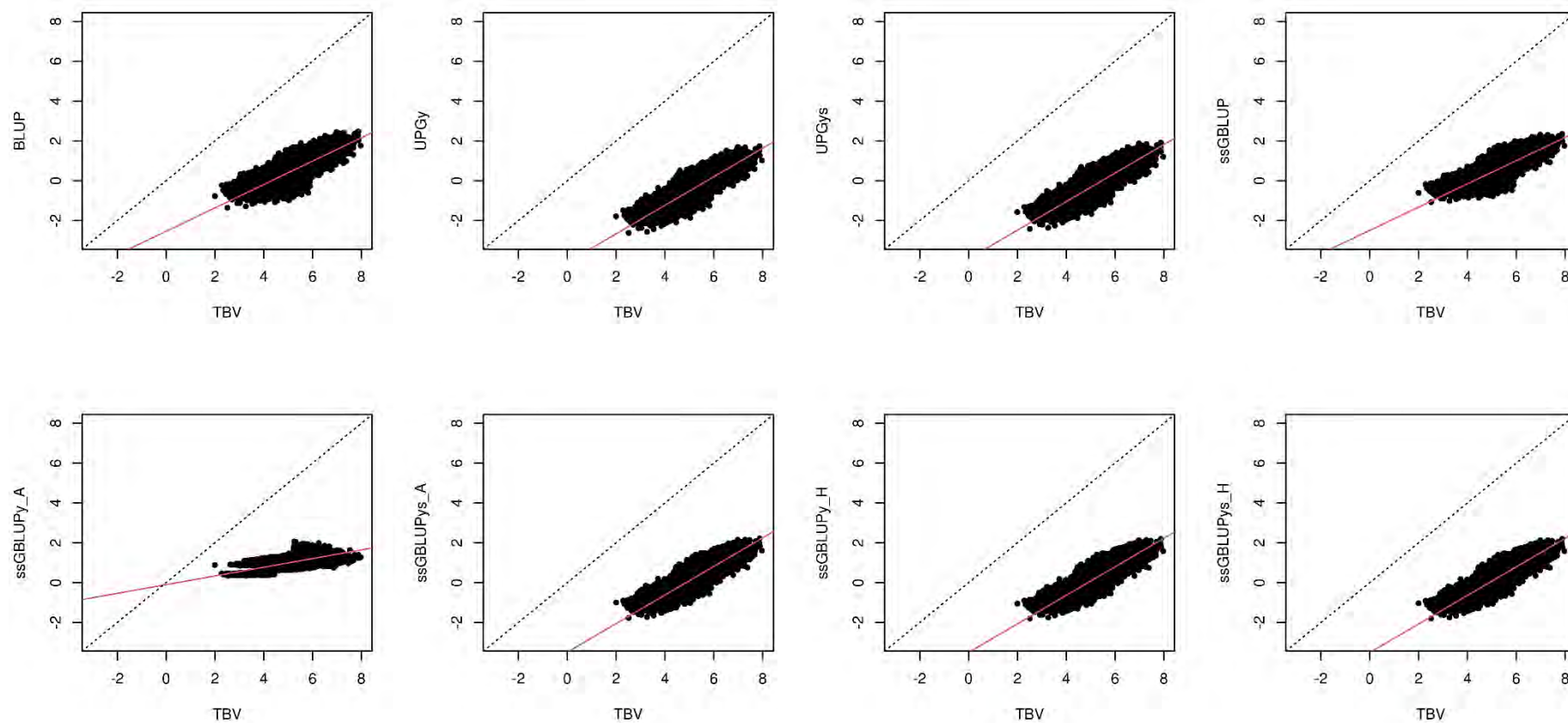
Abreviações: BLUP = pais desconhecidos na matriz **A**; UPGY = UPG considerando geração na matriz **A**; UPGS = UPG considerando sexo e geração na matriz **A**; ssGBLUP: ssGBLUP com pais desconhecidos; ssUPGY_A: ssGBLUP com UPG considerado a geração na matriz **A**. ssUPGS_A: ssGBLUP com UPG considerando sexo e geração na matriz **A**. ssUPGY_H: ssGBLUP com UPG considerando a geração na matriz **H**. ssUPGS_H: ssGBLUP com UPG considerando sexo e geração na matriz **H**.

Fonte: autoria própria

Para a característica simulada de herdabilidade moderada, os modelos sem informação genômica (UPG_Y e UPG_S) os valores de β^1 , DMA e r evidenciam que a utilização de UPG é mais vantajosa comparado ao BLUP sem UPG. Na comparação com o ssGBLUP os valores de β^1 mostram menor deflação e maior r, mas as DMA são maiores do que ssGBLUP em todos às porcentagens como se mostra na tabela 4, já que, ao ignorar os UPGs tira os possíveis laços genéticos o que resulta em viés e mais ainda em indivíduos não relacionados (QUAAS; POLLAK, 1981; TONG; KENNEDY; MOXLEY, 1980), onde os UPG conseguem ter parte da variância genética aditiva específica que pode ser próxima à (MUFF et al., 2019).

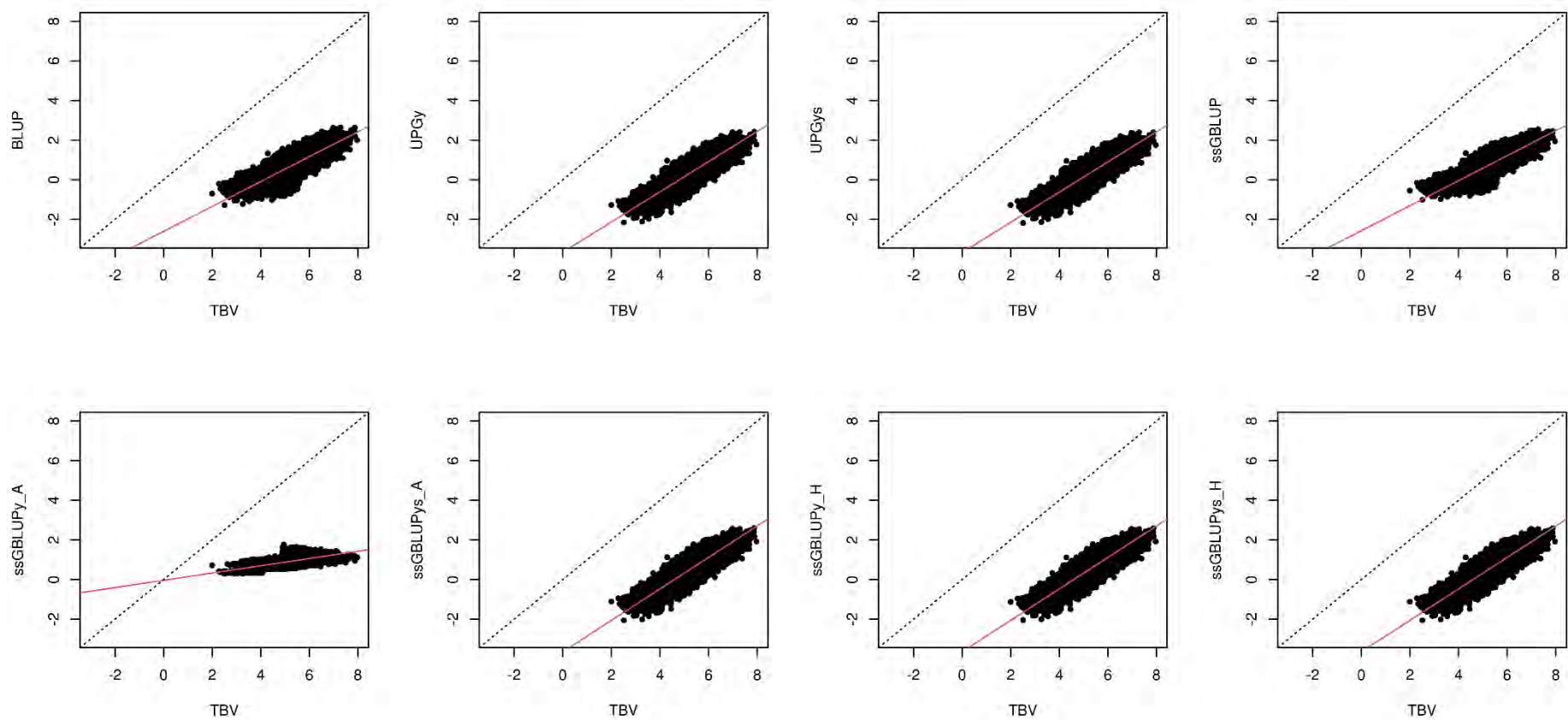
Nos modelos com informação genômica, os β^1 no ssUPGY_A em todas as porcentagens foram maiores que 1 (1,83, 2,18 e 3,01 para 30%, 50% e 70% respectivamente), produzindo maior viés no valores genéticos conforme aumentava-se a porcentagem de pais desconhecidos, mudanças também mostradas junto com os valores de DMA (3,94, 5,13 e 9,58 para 30%, 50% e 70% respectivamente) e a diminuição do r (0,74, 0,71 e 0,69 para 30%, 50% e 70% respectivamente) apresentado nas figuras 4, 5 e 6, o que expõe GEBV inflacionados quando a porcentagem de paternidade desconhecida é menor (30%) quando comparado com os outros modelos por exemplo, porem este modelo é o pior nas duas herdabilidades baixa (0,10) e moderada (0,35). A recorrência deste modelo pode ser explicada pelo fato do uso da matriz de relacionamento baseada só no pedigree o que pode ocasionar a subestimação de variâncias genéticas de animais não conectados com a população (LEGARRA, 2016) como no caso de animais com genótipo que não tem algum relacionamento identificado dentro da população de avaliação.

Figura 4- Regressão dos valores genéticos preditos nas diferentes avaliações com o valor genético verdadeiro (TBV) para herdabilidade moderada (0,35) com 30% de pais desconhecidos



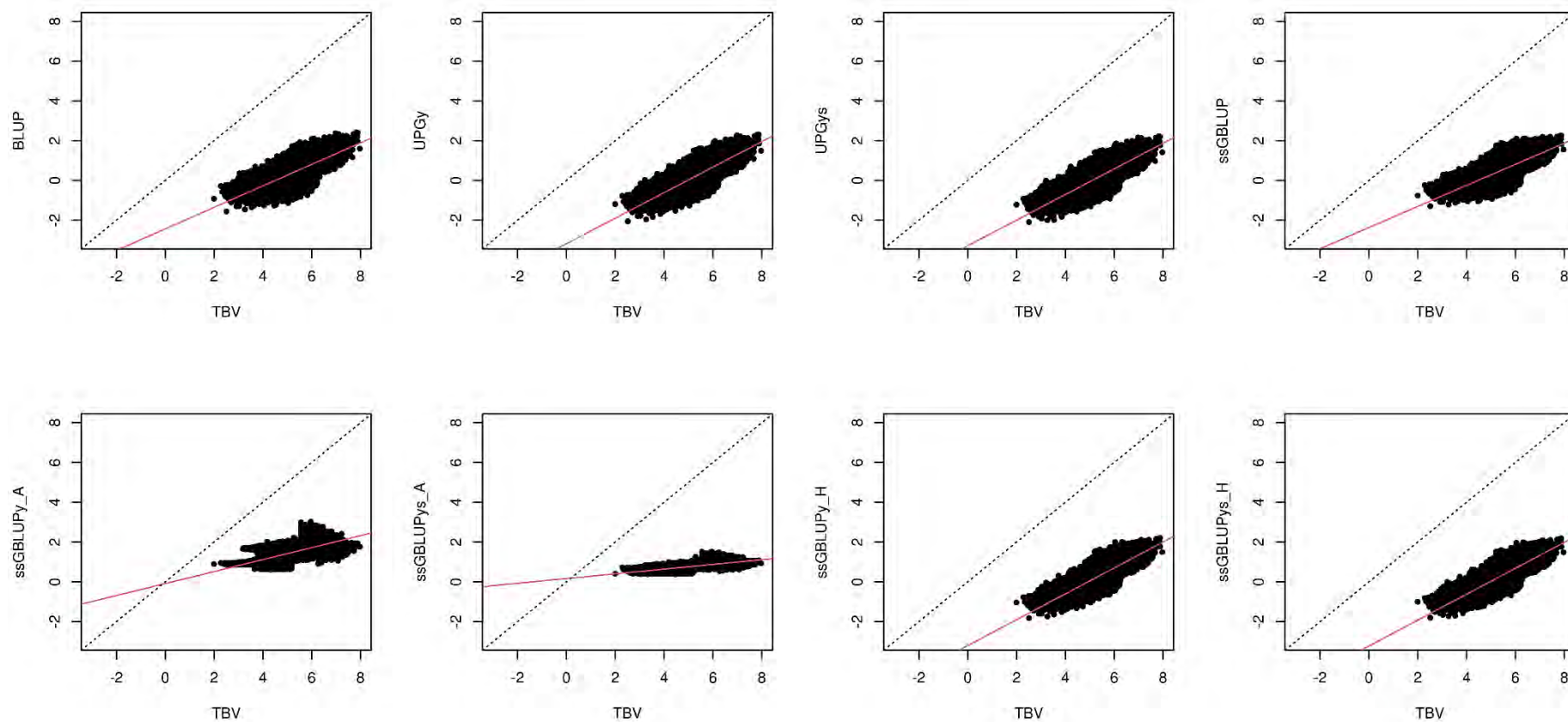
Fonte: autoria própria

Figura 5- Regressão dos valores genéticos preditos nas diferentes avaliações com o valor genético verdadeiro (TBV) para herdabilidade moderada (0,35) com o 50% de pais desconhecidos



Fonte: autoria própria

Figura 6- Regressão dos valores genéticos preditos nas diferentes avaliações com o valor genético verdadeiro (TBV) para herdabilidade moderada (0,35) com o 70% de pais desconhecidos



Fonte: autoria própria

Ressalta-se que a abordagem de estruturar os UPGs em toda a matriz H resulta em melhor convergência do modelo quando comparado aos UPG aplicados somente na matriz A. Foram encontradas diferenças sutis nos resultados obtidos com as avaliações genéticas com e sem informação genômica (BLUP e ssGBLUP), excetuando-se o modelo ssUPGS_A no cenário de 70%, que apresentou um β^1 inflacionado (1,19) e uma r menor (0,66) comparada com os outros modelos com genômica (Figura 6), apesar de um DMA baixo (2,63), os outros modelos apresentam melhores valores de β^1 , DMA e r em todas as porcentagens.

3.4 Conclusão

Os valores genéticos (genômicos) indicaram que os modelos que consideraram o uso de UPGs apresentaram menor deflação, quando comparado com os preditos com o BLUP e o ssGBLUP, a exceção do modelo ssGBLUPY_A que provoca a subestimação dos animais sem paternidade conhecida, fornecendo EBV não confiáveis tanto em características de herdabilidade baixa quanto em características de moderada herdabilidade.

A inclusão da informação genômica nas avaliações genéticas considerando grupos de pais desconhecidos consiste em uma abordagem que aumenta a estabilidade e diminui o viés dos GEBV.

A definição dos UPG no método de passo único deve ser analisada com cautela, mas neste estudo os resultados mostram que considerar os UPGs em toda a matriz H foi a melhor abordagem.

Referências

AGUILAR, I. *et al.* Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. **Journal of Dairy Science**, Lancaster, v. 93, n. 2, p. 743–752, 1 fev. 2010.

BRITO, F. V. *et al.* Accuracy of genomic selection in simulated populations mimicking the extent of linkage disequilibrium in beef cattle. **BMC Genetics**, London, v. 12, n. 1, p. 1–10, 20 set. 2011.

CHRISTENSEN, O. F.; LUND, M. S. Genomic prediction when some animals are not genotyped. **Genetics Selection Evolution**, London, v. 42, n. 3, p. 2, 2010.

FALCONER, D. S.; MACKAY, T. F. C. **Introduction to quantitative genetics**. 4. ed. London: ACRIBIA, 1996.

FAMULA, T. R. An equivalence between models of restricted selection and genetic groups. **Theoretical and Applied Genetics**, Berlin, v. 71, n. 3, p. 413–416, 1985.

GÓMEZ, M. *et al.* Accounting for genetic differences among unknown parents in *Bubalus bubalis*: a case study from the Italian Mediterranean Buffalo. **Frontiers in Genetics**, Lausanne, v. 12, art. 625335, 4 Feb. 2021.

HILL, W. G. Heritability. **Encyclopedia of Genetics**, [s.l.], p. 921–924, 1 Jan. 2001.

KINGHORN, B.; WERF, J. VAN DER; RYAN, M. **Melhoramento animal**: uso de novas tecnologias. Piracicaba: FEALQ, 2016.

LEGARRA, A. Comparing estimates of genetic variance across different relationship models. **Theoretical Population Biology**, New York, v. 107, p. 26–30, 1 Feb. 2016.

LEGARRA, A.; AGUILAR, I.; MISZTAL, I. A relationship matrix including full pedigree and genomic information. **Journal of Dairy Science**, Lancaster, v. 92, n. 9, p. 4656–4663, 1 set. 2009.

MACEDO, F. L. *et al.* Bias and accuracy of dairy sheep evaluations using BLUP and SSGBLUP with metafounders and unknown parent groups. **Genetics, Selection, Evolution**, London, v. 52, n. 1, p. 47, 12 Aug. 2020.

MATILAINEN, K. *et al.* Single step genomic evaluation for female fertility in Nordic Red dairy cattle. **Journal of Animal Breeding and Genetics**, Berlin, v. 135, n. 5, p. 337–348, 1 out. 2018.

MEUWISSEN, T. H. E.; HAYES, B. J.; GODDARD, M. E. Prediction of total genetic value using genome-wide dense marker maps. **Genetics**, Austin, v. 157, n. 4, p. 1819–1829, 2001.

MISZTAL, I. *et al.* Unknown-parent groups in single-step genomic evaluation. **Journal of Animal Breeding and Genetics**, Berlin, v. 130, n. 4, p. 252–258, 1 ago. 2013.

MISZTAL, I. *et al.* **Manual for BLUPF90 family of programs**. 2015. 125 p.

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

MISZTAL, I.; LOURENCO, D.; LEGARRA, A. Current status of genomic evaluation. **Journal of Animal Science**, Champaign, v. 98, n. 4, p. 1–14, 1 Apr. 2020.

MUFF, S. *et al.* Animal models with group-specific additive genetic variances: Extending genetic group models. **Genetics Selection Evolution**, London, v. 51, n. 1, 2019.

OLIVEIRA JÚNIOR, G. A. de *et al.* Definição de grupos genéticos aditivos visando melhor predição de valores genéticos em bovinos de corte. **Revista Brasileira de Saúde e Produção Animal**, Salvador, v. 14, p. 277–286, 2013.

QUAAS, R. L. Additive Genetic Model with Groups and Relationships. **Journal of Dairy Science**, Lancaster, v. 71, n. 5, p. 1338–1345, 1 June. 1988.

QUAAS, R. L.; POLLAK, E. J. Modified equations for sire models with groups. **Journal of Dairy Science**, Lancaster, v. 64, n. 9, p. 1868–1872, 1 Sept. 1981.

SARGOLZAEI, M.; SCHENKEL, F. S. QMSim: a large-scale genome simulator for livestock. **Bioinformatics**, Oxford, v. 25, n. 5, p. 680–681, 1 Mar. 2009.

SPEED, D.; BALDING, D. J. Relatedness in the post-genomic era: Is it still useful? **Nature Reviews Genetics**, London, v. 16, n. 1, p. 33–44, 18 Nov. 2015.

TONG, A. K. W.; KENNEDY, B. W.; MOXLEY, J. E. Potential Errors in Sire Evaluation from Regional Genetic Differences. **Journal of Dairy Science**, Lancaster, v. 63, n. 4, p. 627–633, 1980.

TONUSSI, R. *et al.* Application of single step genomic BLUP under different uncertain paternity scenarios using simulated data. **PLoS ONE**, San Francisco, v. 12, n. 9, art. E0181752, 2017.

VANRADEN, P. M. Accounting for Inbreeding and Crossbreeding in Genetic Evaluation of Large Populations. **Journal of Dairy Science**, Lancaster, v. 75, n. 11, p. 3136–3144, 1 nov. 1992.

WOLAK, M. E.; REID, J. M. Accounting for genetic differences among unknown parents in microevolutionary studies: how to include genetic groups in quantitative genetic animal models. **Journal of Animal Ecology**, London, v. 86, n. 1, p. 7–20, 1 jan. 2017.

4 INCLUSÃO DE GRUPOS DE PAIS DESCONHECIDOS NA AVALIAÇÃO GENÉTICA-GENÔMICA EM REBANHOS COMERCIAIS DA RAÇA NELORE

A.S. Acero Valderrama¹, R. Espigolan¹, B.C. Abreu Silva², F.O. Bussiman³, J.B.S. Ferraz¹, E.C. Mattos¹, J.P. Eler¹

¹*Departamento de Medicina Veterinária, Faculdade de Zootecnia e Engenharia de Alimentos da Universidade de São Paulo (ZMV-FZEA/USP), Pirassununga, SP, Brasil.*

²*Departamento de Zootecnia, Cidade Universitária Prof. José Aloísio de Campos da Universidade Federal de Sergipe, São Cristóvão, SE, Brasil.*

³*Departamento de Nutrição e Produção Animal, Faculdade de Medicina Veterinária e Zootecnia da Universidade de São Paulo (VNP-FMVZ/USP), Pirassununga, SP, Brasil.*

*Autor correspondente: alisson.s.acero@gmail.com

Resumo

Modelos com a inclusão de grupos de pais desconhecidos (*unknow parents groups-UPG*) foram idealizados para mitigar os efeitos da paternidade incerta (QUAAS, 1988) entretanto, nenhuma abordagem foi amplamente adotada, pois cada estrutura proposta tem implicações diferentes (SPEED; BALDING, 2015). Assim, o objetivo deste estudo foi avaliar a inclusão de UPG na avaliação genética-genômica em rebanhos comerciais da raça Nelore para características que compõe o índice de seleção da Agro-Pecuária CFM, como peso á desmama (PD), ganho de peso da desmama ao sobreano ajustado aos 550 dias de idade (GDPSOB), perímetro escrotal (PE), escore de musculosidade (MUSC) e peso aos 18 meses (PES18). No total, a informação fenotípica de 500.000 animais e um pedigree contendo 642.000 animais foram utilizados. No banco de dados foram encontrados 3.174 pais e 240.733 mães com 171.406 e 584.547 progênes, respectivamente. Um total de 243.907 indivíduos com progênie e 171.115 animais sem registros de pai e mãe e 344.564 animais com paternidade desconhecida. Nas análises, foram empregados os programas do pacote BLUPF90, utilizando os modelos: BLUP (*Best Linear Unbiased Prediction*) e ssGBLUP (*single-step Genomic Best Linear Unbiased Prediction*). Os UPG foram definidos com informação de ano de nascimento (UPGY, ssUGPY), ano de nascimento e sexo (UPGS, ssUPGS), e foram adicionados somente na matriz de parentesco baseada nas informações de pedigree (A) ou na matriz de parentesco combinada (H). Para PD os valores mais extremos foram apresentados nos modelos ssUPGY_H e ssUPGS_H, com uma correlação (r_1) de 0 (0, 0,01, respectivamente), os coeficientes de correlação de *rank* (Spearman- r_R) variando de -0,44 a 0,17 e elevado quadrado médio do erro de predição (QMEP) e porcentagem quadrática de viés (PQV), indicando mudanças maiores no ranking quando a seleção é praticada com esses métodos. Os modelos ssUPGY_A e ssUPGS_A foram os piores modelos nas características GDPSOB, MUSC e PES18, apresentando os valores mais baixos de r_1 : de 0,55 e 0,02 para GDPSOB, 0,75 e 0,03 para MUSC e 0,82 e 0,36 para PES18, da mesma forma que os altos valores de QMEP e PQV, quando comparado com os outros modelos. O r_R entre o ssGBLUP com ssUPGY_H e ssUPGS_H são: 0,82 (GDPSOB), 0,88 e 0,98 (PE), 0,89 e 0,88 (MUSC) e 0,83 e 0,84 (PES18). Os ganhos genéticos obtidos nas tendências genéticas para GDPSOB e MUSC foi maior no modelo ssUPGS_A de 7,96

kg/dia e 1,7 unidades, respectivamente, e, para PES18 e PE foi no modelo ssUPGY_A de 2,64 kg e 1,12 cm, respectivamente. Em conclusão, os coeficientes de correlação altos indicaram predições mais precisas nos modelos com informação genômica, com GEBVs menos viesados quando comparados aos modelos sem genômica. A inclusão de UPG nos modelos sem genômica não provocou impacto na classificação dos animais. Para as características pós desmama nesta população, os UPG considerados em toda a matriz H, propiciou melhores resultados.

Palavras-chave: BLUP. Correlação de *Rank*. Paternidade desconhecida. Pearson. Single-step. UPG.

Abstract

Models with the inclusion of unknown parents groups (UPG) were designed to minimize the effects of uncertain paternity (QUAAS, 1988). However, no approach was generally accepted as each one of them has differences (SPEED; BALDING, 2015). Thus, the aim of this study was to evaluate the inclusion of the UPG in the genetic-genomic evaluation of commercial Nellore herds for CFM index traits such as weaning weight (WW), post-weaning weight gain adjusted for 550 days (WWG), scrotal circumference (SC), muscle score (MUSC) and post-yearling weight (PW). In total, the phenotypic data was of 500,000 animals and a pedigree containing 642,000 animals. In the database were found 3,174 fathers and 240,733 mothers with 171,406 and 584,547 progenies, respectively. A total of 243,907 individuals with progeny and 171,115 animals with not records of sire and dam, and 344,564 animals with unknown paternity. The analyses were performed using the BLUPF90 family programs, using the BLUP (the Best Linear Unbiased Prediction) and ssGBLUP (single-step Genomic Best Linear Unbiased Prediction) models. The UPG were defined by the information of birth year (UPGY, ssUGPY), and combination of the birth year with sex (UPGS, ssUPGS). The UPG were added, either only in the relationship matrix from pedigree (A) or only in the genomic relationship matrix (H). For WW, the most extreme values were presented in the ssUPGY_H and ssUPGS_H models, with a correlation (r_1) of 0 (0.0.01, respectively); the rank correlation coefficients (Spearman- r_R) ranging from -0.44 to 0.17 and very high mean square of prediction error (QMEP) and squared percentage of bias (PQV), indicating greater changes in ranking when selection is practiced with these methods. The ssUPGY_A and ssUPGS_A models were the worst models in the characteristics WWG, MUSC and PES18 these are the models with the lowest values of r_1 : 0,55 and 0,02 for GP, 0,75 and 0,03 for MUSC and 0,82 and 0,36 for PES18, as well as the high values of QMEP and PQV, when compared with the other models. The r_R between ssGBLUP with ssUPGY_H and ssUPGS_H are: 0.82 (WWG), 0.88 and 0.98 (SC), 0.89 and 0.88 (MUSC) and 0.83 and 0.84 (PW). Regarding the genetic gain for GDP SOB and MUSC, it was higher in the ssUPGS_A model of 7.96 kg/day and 1.7 und, respectively, for PW and SC it was in the ssUPGY_A model of 2.64 kg and 1.12 cm, respectively. In conclusion, high correlation coefficients indicated more accurate predictions in models with genomic information, with less biased GEBVs when compared to models without genomic information. The inclusion of UPG in the models without genomics had no impact on the ranking of the animals. For the post-weaning characteristics in this population, the UPG considered throughout the H matrix provided better results.

Key words: BLUP. Rank correlation. Unknown paternity. Pearson's correlation. Single-step. UPG.

4.1 Introdução

Aumentar a produtividade nos rebanhos é o objetivo principal dos programas de melhoramento e sua principal ferramenta é a seleção dos animais geneticamente superiores, processo que altera as frequências alélicas mas beneficia o rendimento das características de interesse econômico na população, como por exemplo a resposta dos animais com maior taxa de crescimento ou precocidade sexual que favorecem a diminuição do ciclo produtivo da carne o que permite a sua vez o retorno econômico em menor tempo (BOLIGON et al., 2009) A obtenção de animais que produzam a maior quantidade de carne de ótima qualidade por unidade de produção combinado com uma conformação estética desejável, à pasto e no menor tempo e custo em cada geração é a base do melhoramento genético em gado de corte, otimizando a eficiência dos rebanhos (PETRINI, 2012).

Técnicas de acasalamento como o uso de reprodutores múltiplos (RM), na qual vários reprodutores são mantidos no mesmo piquete para se reproduzir com vacas durante a época de acasalamento (TONUSSI et al., 2017), são bastante utilizadas, pois a maioria da produção de bovinos de corte é de sistema extensivo, e a técnica demanda baixo custo de mão de obra. Embora os RM utilizados nos programas de melhoramento genético brasileiros sejam touros de alto valor genético para características de interesse econômico, pois geralmente são avaliados geneticamente, eles dificultam a identificação de paternidade o que provoca um pedigree incompleto com paternidades incertas. Há relatos na literatura de que 40% dos rebanhos comerciais brasileiros são provenientes de RM e não registram pai identificado, pois em muitas propriedades os touros são usados somente em uma estação de monta e, posteriormente, descartados, dificultando a seleção dos animais geneticamente superiores (CARDOSO et al., 2003; TONUSSI et al., 2017). É importante ressaltar que, com o aumento substancial da quantidade de animais genotipados nos programas de melhoramento genético, ainda existem os erros de genotipagem, que podem ser atribuídos à genótipos imputados, informações de linhagem ausentes e erros de identificação do animal ou da amostra de DNA, são fatores que impactam na compatibilidade entre as matrizes G e A nos procedimentos de avaliação genética utilizando o ssGBLUP, (MEUWISSEN; HAYES; GODDARD, 2016; TONUSSI et al., 2021) afetando diretamente as estimativas da avaliação (LEGARRA; AGUILAR; MISZTAL, 2009).

Nesse contexto, os grupos de pais desconhecidos (UPG) foram idealizados para solucionar problemas de falta de dados e pedigree no modelo animal. Os UPG podem ser estruturados baseados em critérios como ano de nascimento, geração, raça, sexo ou a combinação deles (QUAAS, 1988; TONUSSI et al., 2017), ajudando a superar algumas dificuldades na avaliação genético-genômica como é a incompatibilidade entre as matrizes A e G, aumento de precisão e eficiência nas respostas genéticas (FAMULA; POLLAK; VAN VLECK, 1983; MISZTAL et al., 2013; PEŠKOVIČOVÁ; GROENEVELD; WOLF, 2004; TSURUTA et al., 2014). Ainda assim, o déficit de informação do pedigree pode influenciar os parâmetros do modelo animal mesmo se UPG são incluídos. Assim, o objetivo deste estudo foi avaliar a inclusão de grupos de pais desconhecidos na avaliação genética- genômica em rebanhos comerciais da raça Nelore.

4.2 Material e Métodos

4.2.1 Dados genealógicos e fenotípicos

O banco de dados utilizado é de bovinos da raça Nelore pertencente a Agro-Pecuária CFM, composto de várias fazendas localizadas na região sudoeste e centro-oeste do Brasil, que participam no programa de avaliação genética realizada pelo Grupo de Melhoramento Animal e Biotecnologia (GMAB) da Faculdade de Zootecnia e Engenharia de Alimentos (FZEA/ USP), com animais nascidos entre 1950 a 2018. Os rebanhos são mantidos em pastagens tropicais, constituídas principalmente por solos ácidos com pasto *Brachiaria spp* e suplementação mineral.

No total, a informação fenotípica de 500.000 animais (253.000 machos e 247.000 fêmeas) e um pedigree contendo 642.000 animais foram utilizados. No banco de dados foram encontrados 3.174 pais e 240.733 mães com 171.406 e 584.547 progênes, respectivamente. Um total de 243.907 indivíduos com progênie e 171.115 animais sem registros de pai e mãe e 344.564 animais com paternidade desconhecida (Tabela 1). A população possui uma estrutura de 15 gerações estabelecidas, com 85.158 animais endogâmicos, dos quais 81.834 com um coeficiente de endogamia entre 0 e 0,05.

Tabela 1- Estrutura da população de estudo e coeficiente de endogamia

População	Número de Indivíduos
Número de animais no pedigree	642.670
Pais	3.217
Progênie dos pais	171.478
Mães	240.821
Progênie das mães	584.662
Animais com progênie	244.048
Animais com pai conhecido	171.478
Animais só com o pai desconhecido	412.442
Animais com mãe conhecida	584.662
Animais só com a mãe desconhecida	258
Animais com pai e mãe conhecidos	171.220
Animais com pai e mãe desconhecidos	57.750
Animais genotipados	5.261
Coeficiente de endogamia (F)	
Número endogamicos	85.163
Média da população	0,0015
Média entre endógamos	0,0118
Máximo	0,3139
Mínimo	3,81x10 ⁻⁶

Fonte: autoria própria

As características estudadas são: peso à desmama (PD), ganho de peso da desmama ao sobre ano (GDPSOB), perímetro escrotal (PE), escore de musculosidade (MUSC) e o peso aos 18 meses (PES18). As estatísticas descritivas de cada característica estão na tabela 2. As informações existentes no banco de dados consistiam na identificação dos animais, dos pais, grupos de contemporâneos (GC), fazenda e ano de nascimento, além das idades relacionadas com cada pesagem ou mensuração e a idade da mãe ao parto.

Tabela 2- Estatísticas descritivas das características de peso à desmama (PD), ganho de peso da desmama ao sobre ano (GDPSOB), perímetro escrotal (PE), escore de musculabilidade (MUSC) e peso aos 18 meses (PES18)

Características	Média	SD	Min	Max	N	GC
PD (Kg)	175,496	28,919	60	337,00	509.202	9.586
GDPSOB (Kg/dia)	109,508	36,762	-60	334,34	198.949	992
PE (cm)	26,511	3,566	10	44,00	102.719	502
MUSC (escore 1 a 6)	3,321	1,069	1	6	207.119	935
PES18 (Kg)	294,335	47,815	100	614,00	200.177	1.000

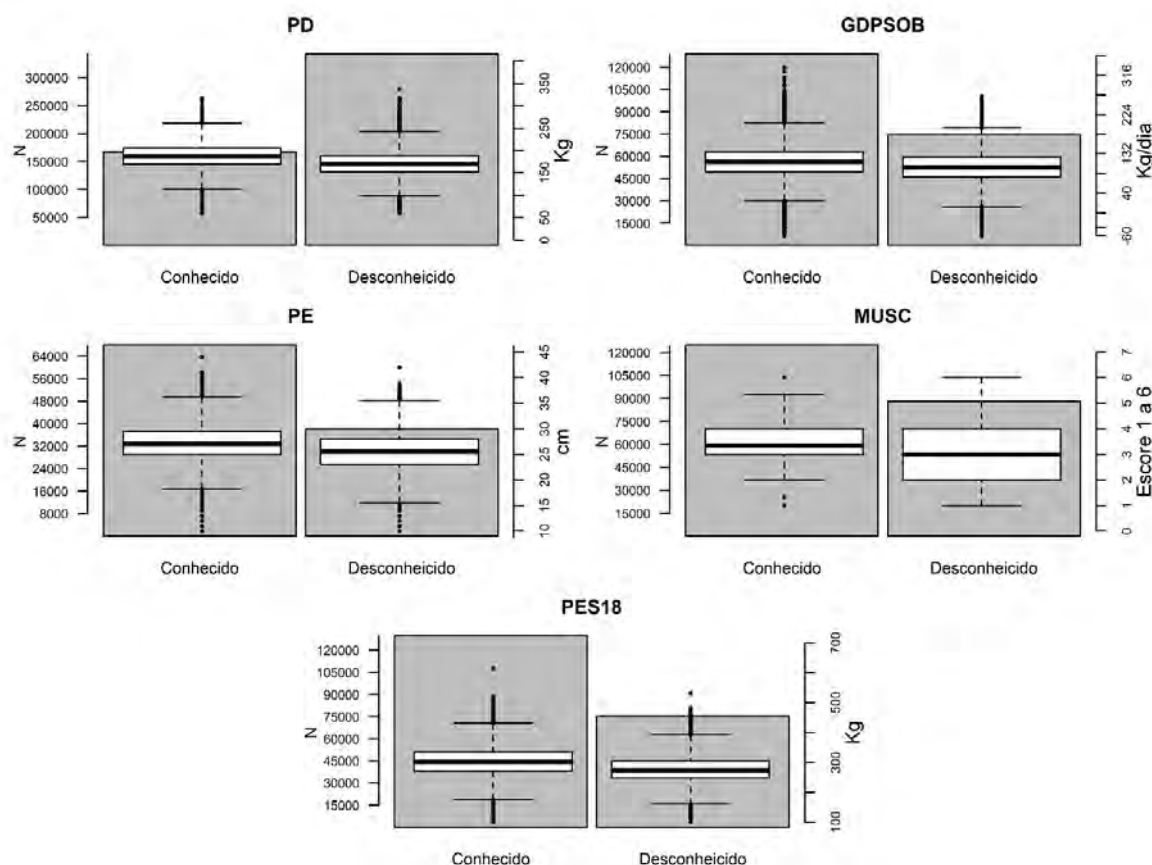
Abreviaturas: Média= média; SD= desvio padrão da característica; Min= valor mínimo encontrado para a característica; Max= valor máximo encontrado para a característica; N= número de animais de cada característica; GC= número de grupos de contemporâneo formados para as características.

Fonte: autoria própria

Os grupos de contemporâneos foram formados pela concatenação das informações de fazenda, safra do animal, sexo e grupo de manejo na respectiva fase de mensuração. O banco de dados foi editado por meio do programa Microsoft Visual FoxPro versão 9.0, excluindo aqueles animais com informações fora da amplitude de valores biologicamente aceitáveis, pertencentes a grupos de contemporâneos com menos de vinte animais, filhos de touros com apenas uma progênie e animais sem nenhum dado de produção mensurado.

Foi verificado o número de pais conhecidos para cada uma das características (Figura 1) observando-se maior número de pais desconhecidos para PD (342.590) seguido de musculabilidade (86.317). Para todas as características os animais sem genealogia têm médias menores se comparados aos animais com pai conhecido (PES18).

Figura 1- Número de pais conhecidos e desconhecidos (eixo da esquerda) e estatística descritiva no box-plot (eixo da direita) para peso a desmama (PD), ganho de peso da desmama ao sobre ano (GDPSOB), perímetro escrotal (PE), escore de musculatura (MUSC) e peso aos 18 meses



Abreviaturas: N= número de animais; *Conhecido = Pai conhecido; Desconhecido =Pai desconhecido.

Fonte: autoria própria

4.2.2 Dados genômicos

Um total de 4.000 animais com informação genômica para 54.791 marcadores do tipo SNP foi utilizado. A consistência dos dados genômicos foi realizada pelo software preGSf90 (MISZTAL et al., 2002). O controle de qualidade dos genótipos foi feito em um processo iterativo considerando os seguintes critérios (Wiggans et al. 2009) para exclusão de SNP: MAF < 0,03; Hardy-Weinberg Equilibrium p-value > 10⁻⁵ e Call Rate < 0,90. Foram considerados somente os 29 autossomos bovinos (BTA – *Bos taurus* autossome). Ainda, SNP foram excluídos do banco de dados quando apresentaram posição idêntica no genoma, assim como as amostras com Call Rate < 0,90. Excluiu-se também SNP altamente correlacionados com outros SNP e/ou SNP com desvio de heterozigosidade maior de 0,15 em relação à heterozigosidade

esperada de acordo com o equilíbrio de Hardy-Weinberg. Após o controle de qualidade, genótipos contendo 43.257 marcadores para 4.000 animais foram disponibilizados para as análises subsequentes.

4.2.3 Parâmetros genéticos

As análises preliminares para estimação dos parâmetros genéticos da população foram obtidas por meio de modelos uni característica. Na Tabela 3 é apresentada a estrutura dos efeitos utilizados em cada modelo para cada característica.

Tabela 3- Descrição do modelo de análise de peso à desmama (PD), ganho de peso da desmama ao sobre ano (GDPSOB), perímetro escrotal (PE), escore de musculosidade (MUSC) e o peso aos 18 meses (PES18)

Característica	D	M	GMAN	GC	IMP
PD	A	A	A	F	F
GDPSOB	A	-	A	F	-
PE	A	-	A	F	F
MUSC	A	-	A	F	F
PES18	A	-	A	F	F

Abreviaturas: D= efeito genético direto; M= efeito genético materno; GMAN= grupo de manejo; GC= Grupo contemporâneo; IMP= grupo de idade da mãe ao parto. A= efeito aleatório; F= efeito fixo; - = efeito não considerado.

Fonte: autoria própria

Na literatura, para a raça Nelore se reporta herdabilidades diretas de 0,15 a 0,17 e materna de 0,11 para PD (KOETZ JUNIOR et al., 2019; RAIMUNDO, 2018), 0,23 para peso ao sobreano e 0,25 para perímetro escrotal (RAIMUNDO, 2018). Para a característica de musculosidade, é reportado um intervalo de 0,27 a 0,45 herdabilidades (KOETZ et al., 2019; KOURY FILHO et al., 2010). Esses valores encontrados na literatura são similares aos encontrados no presente estudo, como está apresentado na Tabela 4:

Tabela 4- Componentes de (co)variância, herdabilidades diretas (h_a^2) e herdabilidade materna (h_m^2) para peso a desmama (PD), ganho de peso da desmama ao sobre ano (GDPSOB), perímetro escrotal (PE), escore de musculosidade (MUSC) e peso aos 18 meses (PES18) estimados com (ssGREML) e sem informação genômica (REML)

	REML						h_a^2	h_m^2
	σ_a^2	σ_e^2	σ_m^2	σ_{mpe}^2	σ_{am}	σ_{mg}^2		
PD	60,00	206,59	22,99	56,56	8,45	-	0,17	0,06
GDPSOB	78,91	329,31	-	-	-	96,30	0,16	-
PE	2,89	3,91	-	-	-	0,36	0,40	-
MUSC	0,14	0,56	-	-	-	0,03	0,19	-
PES18	204,53	382,51	-	-	-	67,32	0,31	-

	ssGREML							
	σ_a^2	σ_e^2	σ_m^2	σ_{mpe}^2	σ_{am}	σ_{mg}^2	h_a^2	h_m^2
PD	58,56	207,40	23,48	55,94	8,61	-	0,17	0,07
GDPSOB	79,00	329,05	-	-	-	96,10	0,16	-
PE	2,83	3,96	-	-	-	0,36	0,40	-
MUSC	0,14	0,56	-	-	-	0,04	0,18	-
PES18	204,07	382,26	-	-	-	67,17	0,31	-

Abreviaturas: σ_a^2 = Variância aditiva direta; σ_e^2 = Variância residual; σ_m^2 = Variância genética materna; σ_{mpe}^2 = Variância do ambiente permanente materno, σ_{am} = (Co)variância entre o efeito aditivo direto e o genético materno; σ_{mg}^2 = Variância do grupo de manejo considerado para cada característica; h_a^2 = Herdabilidade direta; h_m^2 = Herdabilidade materna. *REML= Máxima verossimilhança restrita; ssGREML = Máxima verossimilhança restrita com a inclusão de informação genômica.

Fonte: autoria própria

4.2.4 Modelos

As análises foram realizadas utilizando a família de programas BLUPF90 (MISZTAL et al., 2002, 2015) disponíveis em <http://nce.ads.uga.edu/wiki/doku.php>, e o R (R CORE TEAM, 2019). No método BLUP, foi processada a avaliação genética tradicional unicaracterística utilizando a informação do pedigree e dos fenótipos. O modelo (1) pode ser representado da seguinte maneira:

$$y = \mathbf{X}b + \mathbf{Z}u + e \quad (1)$$

Onde, y é o vetor dos fenótipos, b é o vetor dos efeitos fixos, u é o vetor dos efeitos genéticos aditivos, \mathbf{X} e \mathbf{Z} são as matrizes de incidência para b e u , e e é o vetor do resíduo. Considerando um modelo infinitesimal, $\text{var}(u) = \mathbf{A}\sigma_u^2$, onde \mathbf{A} é a matriz de parentesco obtida pela informação do pedigree e σ_u^2 é a variância do efeito genético. Para demonstrar a inclusão de UPG na notação mais simples, se assumirá um modelo com UPG visível, ignorando todos os efeitos fixos (2), exceto dos grupos (MISZTAL et al., 2013; QUAAS, 1988):

$$y = \mathbf{Z}Qs + \mathbf{Z}a + e \quad (2)$$

Onde s é o efeito de grupo, a é o efeito aditivo do animal, e \mathbf{Q} é a matriz dos animais atribuídos aos grupos. Mais especificamente, cada linha de \mathbf{Q} contém a composição genética do grupo de um animal. Por concordância, os componentes de variância σ_a^2 e σ_e^2 são definidos como unidade.

As equações dos UPG podem ser integradas em \mathbf{A}^{-1} no modelo BLUP, essa inserção foi denominado por Quaas e Pollak (1981) como transformação QP, baseados no algoritmo de VanRaden (1992) que contempla a endogamia de animais cujos pais são desconhecidos é igual à endogamia média de seus contemporâneos

com pais conhecidos. A transformação QP ocorre na inversa da matriz de parentesco, como segue (3):

$$A_{UPG}^{-1} = \begin{pmatrix} A^{-1} & -A^{-1}Q \\ -Q'A^{-1} & -Q'A^{-1}Q \end{pmatrix} \quad (3)$$

No BLUP genômico de “passo único” (ssGBLUP) proposto por Misztal *et al.* (2009), a matriz inversa de parentesco (A^{-1}) será substituída por H^{-1} que combina o pedigree e a informação genômica. A matriz H^{-1} (4) pode ser obtida da seguinte forma (AGUILAR *et al.*, 2010):

$$H^{-1} = A^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & G^{-1} - A_{22}^{-1} \end{bmatrix} \quad (4)$$

Onde G^{-1} é a inversa da matriz genômica de parentesco e A_{22}^{-1} é a inversa da matriz de parentesco do pedigree somente para os animais genotipados. Ao incluir os UPG, a matriz H^{-1} será ampliada por meio da transformação QP (VANRADEN, 1992), mas os UPG podem ser considerados só matriz A^{-1} dentro de H^{-1} , apresentado aqui como $H_{UPG_A}^{-1}$ (5):

$$H_{UPG_A}^{-1} = \begin{pmatrix} A^{00} & A^{01} & A^{02} \\ A^{10} & A^{11} & A^{12} \\ A^{20} & A^{21} & A^{22} \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & G^{-1} - A_{22}^{-1} \end{pmatrix} \quad (5)$$

Ou, considerá-los em toda a matriz H^{-1} , representada como $H_{UPG_H}^{-1}$ (6):

$$H_{UPG_H}^{-1} = \begin{pmatrix} A^{00} & A^{01} & A^{02} \\ A^{10} & A^{11} & A^{12} \\ A^{20} & A^{21} & A^{22} \end{pmatrix} + \begin{pmatrix} Q_2'(G^{-1} - A_{22}^{-1})Q_2 & 0 & -Q_2'(G^{-1} - A_{22}^{-1}) \\ 0 & 0 & 0 \\ -(G^{-1} - A_{22}^{-1})Q_2 & 0 & G^{-1} - A_{22}^{-1} \end{pmatrix} \quad (6)$$

Onde, Q_2 é a matriz que contém os UPGs para os animais genotipados (MACEDO *et al.*, 2020; MISZTAL; LOURENCO; LEGARRA, 2020).

4.2.5 Cenários testados

Para avaliar a aplicação dos métodos BLUP e ssGBLUP, incluindo grupos de pais desconhecidos (UPG), considerando ou não as informações genômicas nas avaliações genéticas, quatro cenários serão testados (Tabela 5):

Tabela 5- Cenários para os diferentes de grupos de pais desconhecidos (UPG) estruturados para cada característica de estudo

Modelo	Estrutura dos UPG	Abreviações
BLUP	Sem agrupamento	BLUP
	Ano de nascimento	UPGY
	Ano de nascimento e sexo	UPGS

	Sem agrupamento	ssGBLUP
	Ano de nascimento na matriz A	ssUPGY_A
ssGBLUP	Ano de nascimento e sexo na matriz A	ssUPGS_A
	Ano de nascimento na matriz H	ssUPGY_H
	Ano de nascimento e sexo na matriz H	ssUPGS_H

Abreviaturas: BLUP = best linear unbiased prediction; ssGBLUP = single-step genomic best linear unbiased prediction; UPG = Grupo de pai desconhecido.

Fonte: autoria própria

A estrutura de agrupamento foi utilizada de acordo com o critério de contabilizar as tendências genéticas e as diferenças genéticas entre as gerações, evitando ignorar as diferenças reais na população (TONG; KENNEDY; MOXLEY, 1980) fornecido pela informação própria do animal. Dentro da estrutura dos grupos foi considerada um mínimo de animais com fenótipo, e grupos pequenos foram juntados. Os cenários serão avaliados pelos valores genéticos (genômicos) estimados, a acurácia dos valores genéticos e classificação dos animais considerando o índice.

4.2.6 Análise dos resultados

Para cada modelo e cenário os GEBV foram avaliados de acordo a indicadores de qualidade como o coeficiente de correlação de Pearson (r) (8):

$$r = \frac{cov(\hat{y}_t, y_t)}{\sqrt{V(\hat{y}_t)V y_t}} \quad (8)$$

O quadrado médio do erro de predição (QMEP) (9):

$$QEMP = \frac{1}{n} \sum_{t=1}^n (\hat{y}_t - y_t)^2 \quad (9)$$

A porcentagem quadrática de viés (PQV) (10):

$$PQV = \left(\frac{\sum_{t=1}^n (\hat{y}_t - y_t)^2}{\sum_{t=1}^n (y_t)^2} \right) * 100 \quad (10)$$

Em que, y_t corresponde aos GEBV preditos pelo primeiro modelo, \hat{y}_j são os GEBV preditos pelo segundo modelo e n é o número de dados, no caso da correlação de Pearson também foi calculada entre os GEBV de todos os modelos.

O coeficiente de correlação de Rank (Spearman) entre os GEBV dos modelos (11):

$$r_R = 1 - \frac{6 \sum_i d_i^2}{n(n^2-1)} \quad (11)$$

Onde d_i é a diferença entre os dois postos de cada GEBV predito.

4.3 Resultados e discussão

Os indicadores de qualidade de ajuste utilizados para comparar os modelos estão na tabela 6, onde correlação simples (r_1) entre os GEBV predito pelos modelos propostos e os GEBV estimados ajustando os efeitos fixos.

Tabela 6- Quadrado médio do erro de predição (QMEP), porcentagem quadrática de viés (PQV) e correlação (r_1) dos valores genéticos de peso á desmama (PD), ganho de peso da desmama ao sobre ano (GDPSOB), perímetro escrotal (PE), escore de musculosidade (MUSC) e peso aos 18 meses (PES18) (Continua)

CARAC	Análise	QMEP	PQV(%)	r_1
PD	BLUP	143,71	0,45	0,91
	UPGY	143,22	0,45	0,91
	UPGS	143,07	0,45	0,91
	ssGBLUP	145,02	0,46	0,91
	ssUPGY_A	145,01	0,46	0,91
	ssUPGS_A	640,83	2,03	0,67
	ssUPGY_H	871998692,00	2756409,00	0,00
	ssUPGS_H	26727798,00	84487,23	0,01
GDPSOB	BLUP	274,67	2,07	0,89
	UPGY	273,43	2,06	0,89
	UPGS	273,06	2,05	0,89
	ssGBLUP	274,67	2,07	0,89
	ssUPGY_A	1873,86	14,09	0,55
	ssUPGS_A	1988432,00	14951,53	0,02
	ssUPGY_H	273,30	2,05	0,89
	ssUPGS_H	272,96	2,05	0,89
PE	BLUP	2,46	0,34	0,91
	UPGY	2,45	0,34	0,91
	UPGS	2,46	0,34	0,91
	ssGBLUP	2,53	0,35	0,91
	ssUPGY_A	103,29	14,44	0,25
	ssUPGS_A	2,50	0,35	0,91
	ssUPGY_H	2,52	0,35	0,91
	ssUPGS_H	2,53	0,35	0,91
MUSC	BLUP	0,47	3,83	0,78
	UPGY	0,46	3,82	0,78
	UPGS	0,46	3,82	0,78
	ssGBLUP	0,47	3,83	0,78
	ssUPGY_A	0,50	4,13	0,75
	ssUPGS_A	682,93	5611,62	0,03
	ssUPGY_H	0,46	3,81	0,78
	ssUPGS_H	0,46	3,81	0,78

Tabela 6- Quadrado médio do erro de predição (QMEP), porcentagem quadrática de viés (PQV) e correlação (r_1) dos valores genéticos de peso á desmama (PD), ganho de peso da desmama ao sobre ano (GDPSOB), perímetro escrotal (PE), escore de musculosidade (MUSC) e peso aos 18 meses (PES18) ()

CARAC	Análise	QMEP	PQV(%)	r_1
PES18	BLUP	273,99	0,31	0,94
	UPGY	270,42	0,31	0,94
	UPGS	270,07	0,31	0,94
	ssGBLUP	274,51	0,31	0,94
	ssUPGY_A	566,38	0,64	0,87
	ssUPGS_A	11238,74	12,68	0,36
	ssUPGY_H	270,95	0,31	0,94
	ssUPGS_H	270,59	0,31	0,94

Abreviaturas: BLUP= avaliação baseada na matriz A; UPGY = avaliação com UPG estruturado considerando geração; UPGY = avaliação com UPG estruturado considerando geração e sexo; ssGBLUP = avaliação de passo único; ssUPGY_A = avaliação de passo único baseado com UPG estruturado com a geração considerado na matriz A. ssUPGS_A = avaliação de passo único com UPG estruturado com a geração e sexo considerado na matriz A. ssUPGY_H = avaliação de passo único com UPG estruturado com a geração considerado em toda matriz H. ssUPGS_H = avaliação de passo único com UPG estruturado com a geração e sexo considerado em toda matriz H.

Fonte: autoria própria

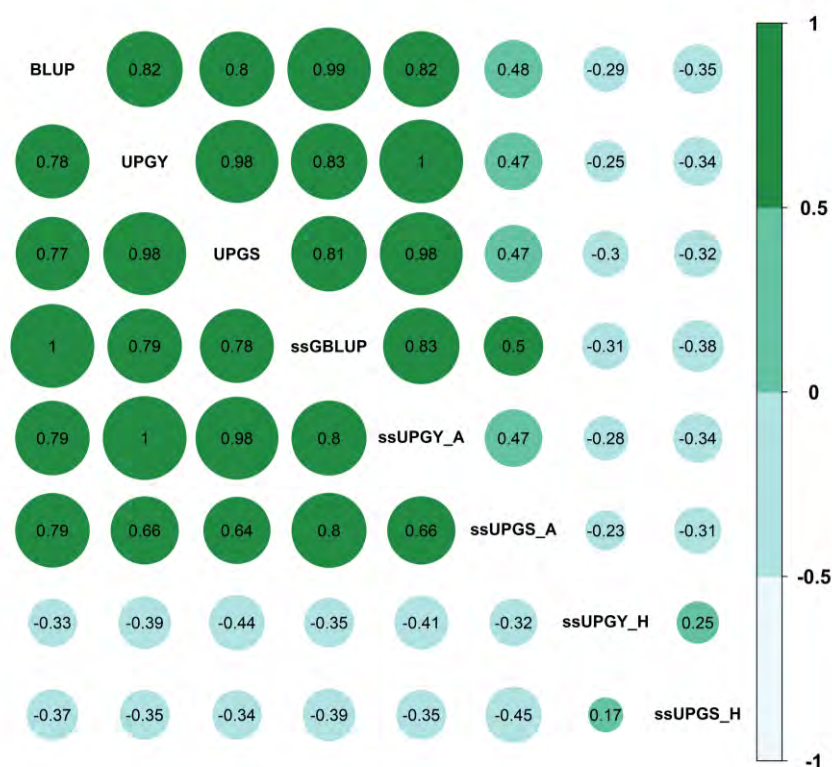
Para PD os valores mais extremos foram apresentados nos modelos ssUPGY_H e ssUPGS_H, com uma (r_1) de 0 (0, 0,01, respectivamente) e elevados QMEP e PQV (tabela 6), resultados que demonstram as dificuldades de convergência, valores não usuais e viés na aplicação desses modelos, o que, em parte, pode ser explicado pela complexidade do efeito materno contido dentro da característica, pois a falta de conexão entre os parentes maternos com e sem informação genômica dificulta a estimação deste efeito e ao grande número de pais desconhecidos para os animais dentro desta característica (WILLHAM, 1972). Para os outros modelos com a informação genômica, a melhor abordagem é dada pelo modelo ssUPGY_A, pois quando comparado com o ssGBLUP, as diferenças nos valores de qualidade (QMEP, PQV e r_1) são similares, evidenciando que, ao se considerar os UPG de estrutura simples (ano de nascimento do animal) na matriz A, os resultados são mais informativos, do que uma abordagem mais complexa em toda a matriz H, mas não mostra alguma ganho de precisão como relatam Tonussi et al. (2021) e menos grupos poderiam ser necessários (QUAAS; POLLAK, 1981), considerando r_1 como uma medida de precisão.

Os modelos ssUPGY_A e ssUPGS_A foram as abordagens menos interessantes para as características GP, MUSC e PES18, apresentando valores mais

baixos de r_1 : de 0,55 e 0,02 para GP, 0,75 e 0,03 para MUSC e 0,82 e 0,36 para PES18, da mesma forma que os altos valores de QMEP e PQV, quando comparado com os outros modelos. É importante ressaltar que para MUSC o modelo ssUPGY_A não apresenta valores exacerbados de QMEP (0,50) e PQV (4,13) como acontece para as outras características. Para CE, poderia dizer-se que o melhor modelo é ssUPGS_A porque por uma pequena diferença o QMEP (2,50) é menor (2,52 e 2,53 para ssUPGY_H e ssUPGS_H, respectivamente) e os outros critérios permanecem iguais (0,35 e 0,91 para PQV e r_1 respectivamente). Excetuando-se o modelo ssUPGY_A, o qual gerou valores altos de viés e baixa r_1 para esta característica (tabela 6).

Os coeficientes de correlação de *rank* (Spearman- r_R) variaram em todas as características. Para PD, os valores mais baixos foram entre os modelos ssUPGY_H e ssUPGS_H, variando de -0,44 a 0,17, indicando mudanças maiores no ranking quando a seleção é praticada com esses métodos (Figura 2), no entanto, os valores entre o ssGBLUP e os outros dois modelos com genômica ssUPGY_A e ssUPGY_A, apresentam o mesmo coeficiente de r_R de 0,8, mas o coeficiente de correlação de Pearson (r_2) entre ssGBLUP e ssUPGS_A foi de 0,5, menor do que entre ssGBLUP e ssUPGY_A (0,83).

Figura 2- Correlação de $rank$ r_R (embaixo da diagonal) e correlação de Pearson- r_2 (acima da diagonal) dos valores genéticos para peso á desmama (PD) nos diferentes modelos

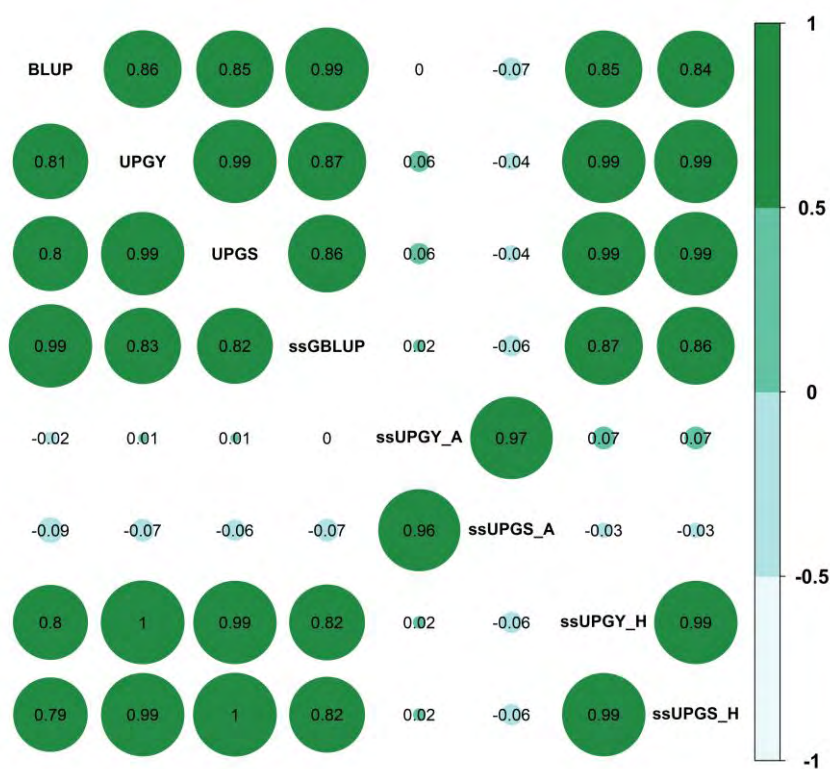


Fonte: autoria própria

Os modelos ssUPGY_H e ssUPGS_H contra todas as outras abordagens, exceto os modelos com a estrutura de UPG na matriz A, geraram valores altos em GDPSOB, PE, MUSC e PES18 (Figuras 3, 4, 5 e 6). O coeficiente de correlação de $rank$ entre o ssGBLUP com ssUPGY_H e ssUPGS_H são: 0,82 (GDPSOB), 0,88 e 0,98 (PE), 0,89 e 0,88 (MUSC) e 0,83 e 0,84 (PES18), refletindo em GEBVs mais consistentes e estáveis, garantindo a expectativa correta dos EBV dos animais genotipados, de maneira similar ao reportado por Matilainen et al. (2018) para características de fertilidade em gado nórdico atribuído os UPG na matriz H e na pesquisa de Misztal et al., (2017) com gado Holstein. Em PE o r_R e o r_2 entre os modelos ssGBLUP e ssUPGY_A foi de 0,14, e com ssUPGS_A foram maiores de 0,9 e 0,92 (r_R e r_2 , respectivamente). De maneira antagônica ao PE, para MUSC, a r_R e o r_2 foram menores entre o ssGBLUP e ssUPGS_A (0,04 e 0,05, respectivamente) que comparados com os valores obtidos entre ssGBLUP e ssUPGY_A (0,52 e 0,5, respectivamente). Os resultados antagônicos para PE e MUSC podem estar relacionados com a estrutura dos UPG considerando o sexo, pois PE só é mensurado em machos o que faz com que as estimativas em fêmeas sejam baixas, coisa diferente

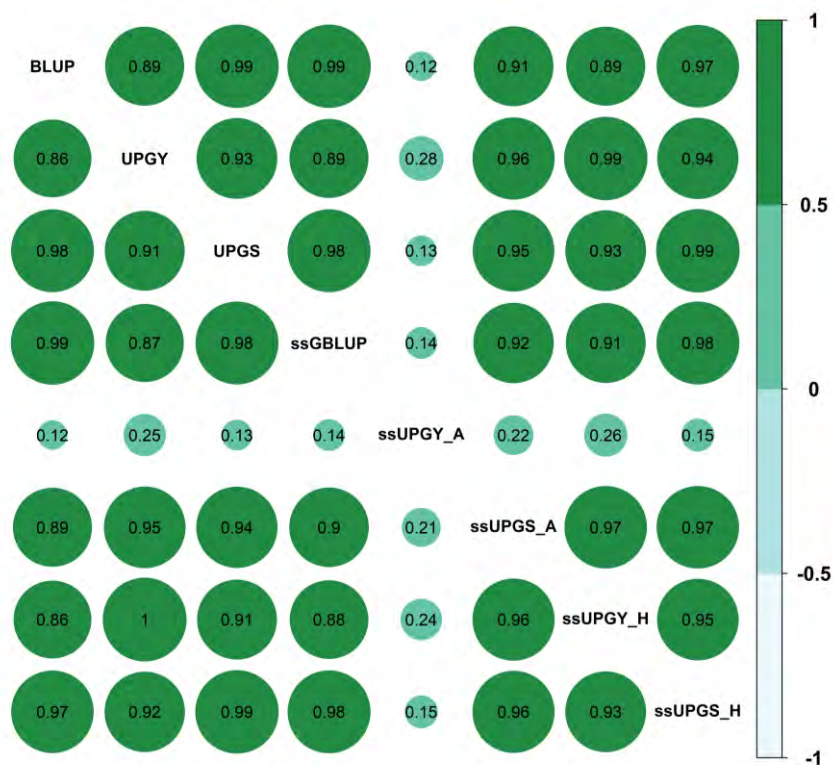
que acontece com MUSC similar a como o descrevem Gómez et al. (2021) em seus resultados. Todos esses valores baixos mostram que a pressuposição de assumir que os pais “fantasmas” são não-correlacionados e não endógamos (QUAAS, 1988) pode apontar a uma incorreta classificação dos animais e uma decisão de seleção subestimada.

Figura 3- Correlação de $rank\ r_R$ (embaixo da diagonal) e correlação de Pearson- r_2 (acima da diagonal) dos valores genéticos para ganho de peso da desmama ao sobre ano (GDPSOB) nos diferentes modelos



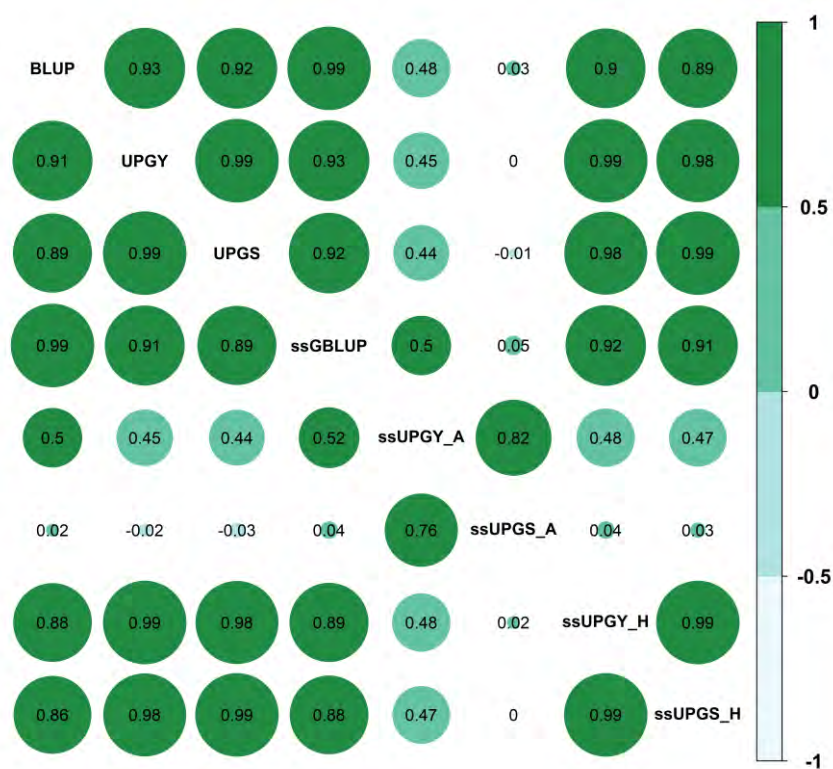
Fonte: autoria própria

Figura 4- Correlação de $rank-r_R$ (embaixo da diagonal) e correlação de Pearson- r_2 (acima da diagonal) dos valores genéticos para perímetro escrotal (PE) nos diferentes modelos



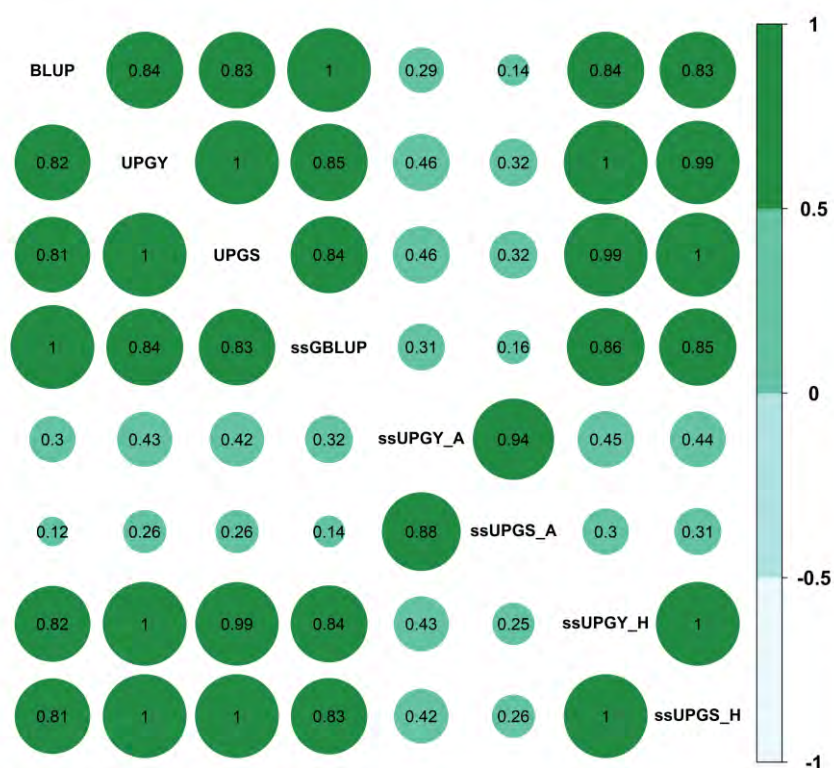
Fonte: autoria própria

Figura 5- Correlação de $rank-r_R$ (embaixo da diagonal) e correlação de Pearson- r_2 (acima da diagonal) dos valores genéticos para score de musculabilidade (MUSC) nos diferentes modelos



Fonte: autoria própria

Figura 6- Correlação de r_R (embaixo da diagonal) e correlação de Pearson- r_2 (acima da diagonal) dos valores genéticos para peso aos 18 meses (PES18) nos diferentes modelos

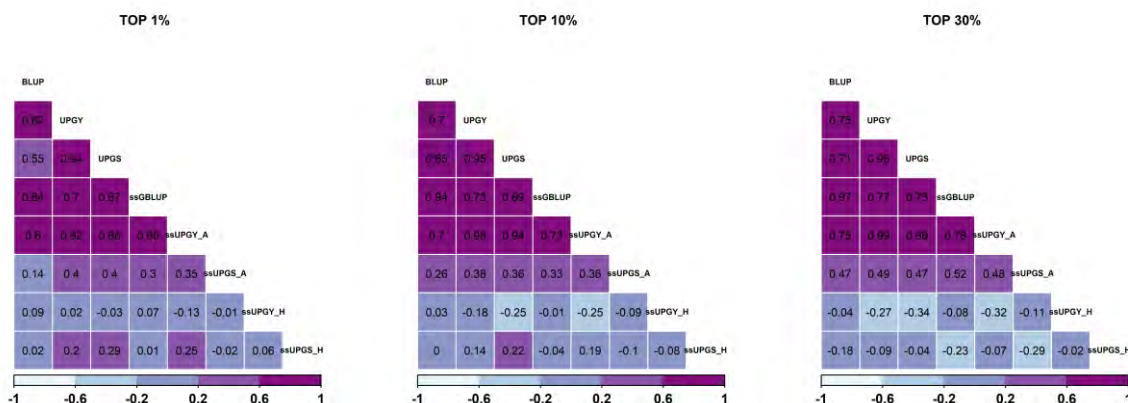


Fonte: autoria própria

Nos modelos sem informação genômica e para todas as características estudadas, os r_R e r_2 os valores foram $>0,77$, por exemplo, o r_R foi de 0,78 entre BLUP com UPGY e 0,77 entre BLUP com UPGS para PD, mostrando-se da mesma maneira com r_2 com valores de 0,82 e 0,8 (BLUP com UPGY e BLUP com UPGS, respectivamente). Ou seja, que a melhor abordagem segundo estes critérios de qualidade e coeficientes de correlação (r_R e r_2) para o uso de UPG e informação genômica é considerá-los em toda a matriz H, pois foram evidenciadas diferenças relevantes entre eles, excetuando PD. Os modelos sem informação genômica (UPGY e UPGS) apresentam os mesmos valores do BLUP para os critérios de ajuste de qualidade em todas as características de estudo, com mínima diferença. Podemos inferir que a contabilização dos relacionamentos foi particularmente eficaz ao reduzir as variações nestes modelos (KENNEDY; MOXLEY, 1975), mas as pequenas diferenças de QMEP entre eles pode mostrar o viés e possíveis EBV tendenciosos ao ignorar o efeito do UPG na avaliação tradicional (KENNEDY, 1981).

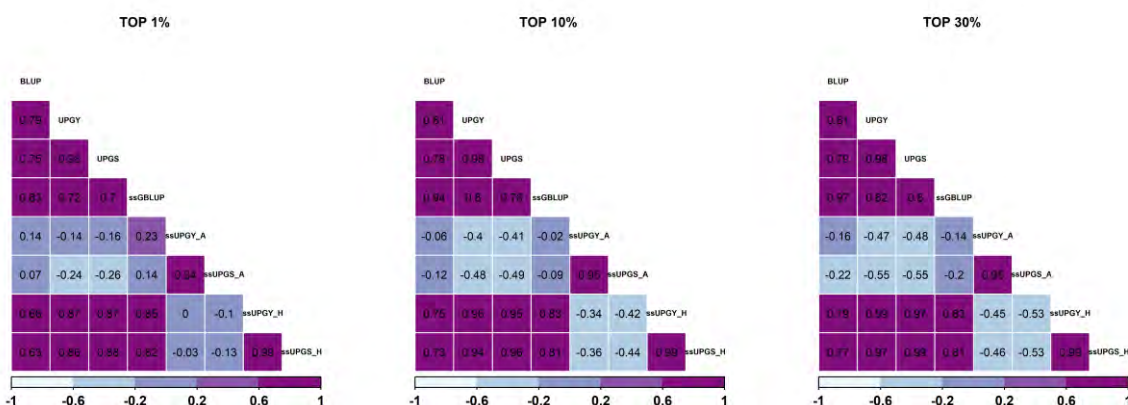
O mesmo padrão de r_R para PD, GDPSOB, PE, MUSC e PES18 foi observado ao selecionar 1%, 10% e 30% dos melhores animais (figuras 7,8,9, 10 e 11) que com toda a população.

Figura 7- Correlação de $rank-r_R$ dos valores genéticos ao selecionar o 1%, 10% e 30% dos melhores animais (TOP) para peso á desmama (PD) nos diferentes modelos



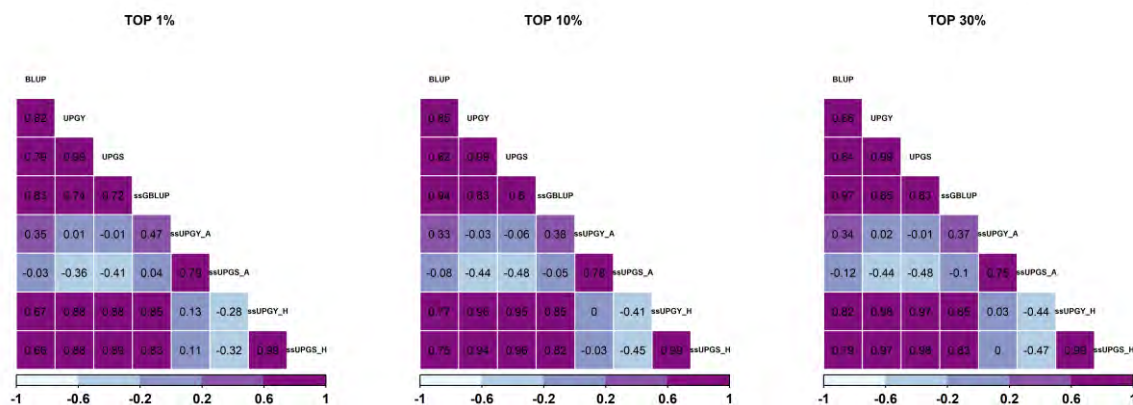
Fonte: autoria própria

Figura 8- Correlação de $rank-r_R$ dos valores genéticos ao selecionar o 1%, 10% e 30% dos melhores animais (TOP) para ganho de peso da desmama ao sobre ano (GDPSOB) nos diferentes modelos



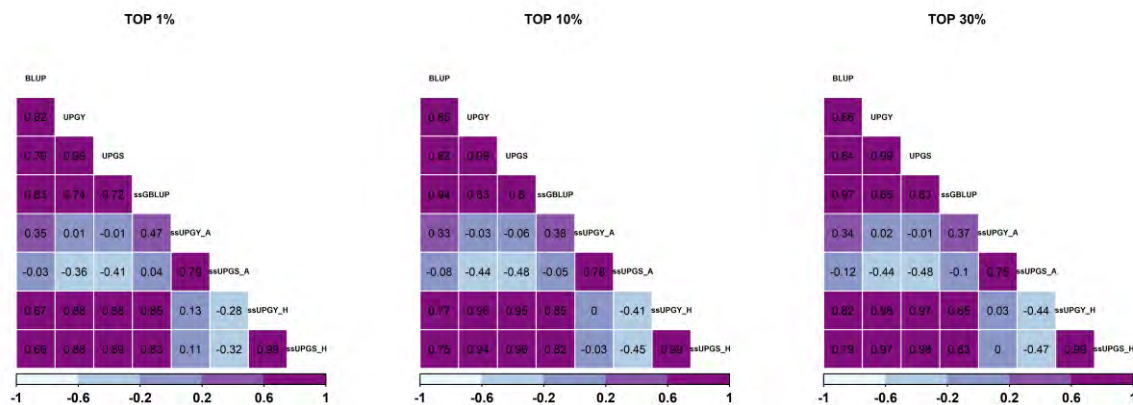
Fonte: autoria própria

Figura 9- Correlação de $rank-r_R$ dos valores genéticos ao selecionar o 1%, 10% e 30% dos melhores animais (TOP) para perímetro escrotal (PE) nos diferentes modelos



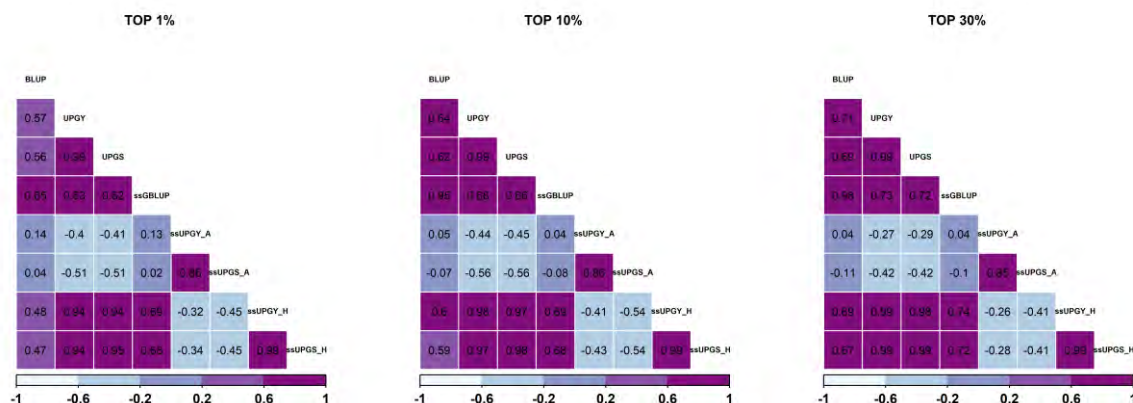
Fonte: autoria própria

Figura 10- Correlação de $rank-r_R$ dos valores genéticos ao selecionar o 1%, 10% e 30% dos melhores animais (TOP) para score de musculabilidade (MUSC) nos diferentes modelos



Fonte: autoria própria

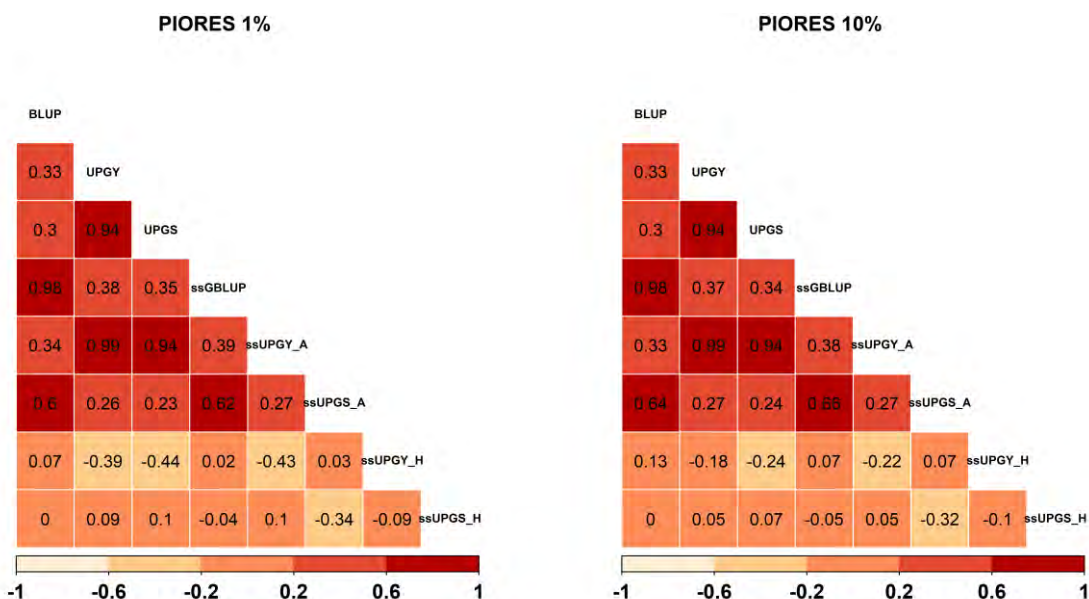
Figura 11- Correlação de r_R dos valores genéticos ao selecionar o 1%, 10% e 30% dos melhores animais (TOP) para peso aos 18 meses (PES18) nos diferentes modelos



Fonte: autoria própria

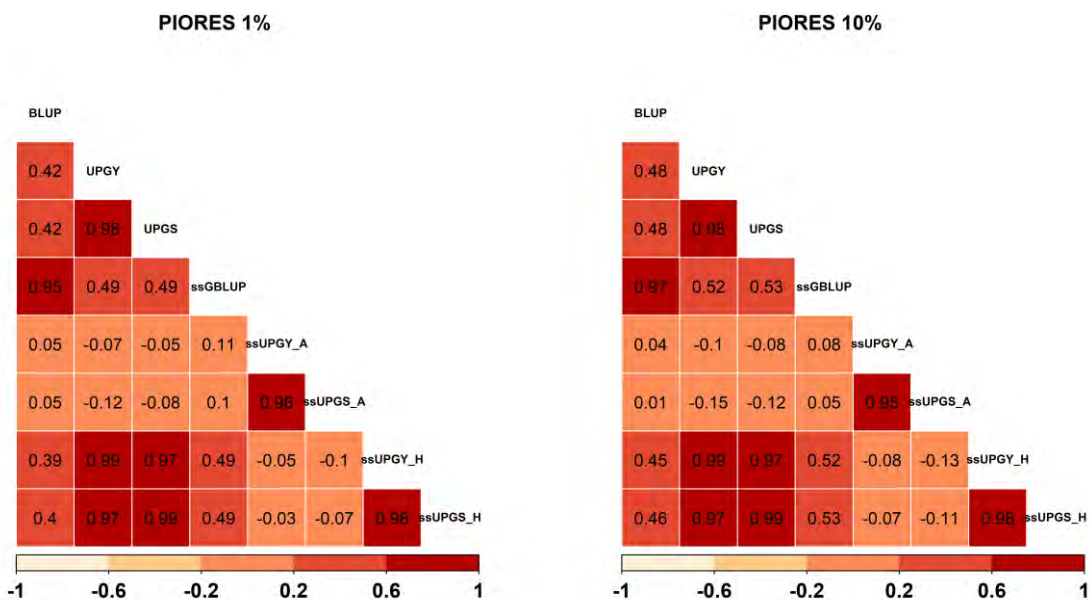
No entanto, ao observar r_R ao selecionar 1% e 10% dos piores animais, onde se encontram a maioria dos animais com paternidade indeterminada, apresenta valores entre -0,15 a 0,98 para todas as características (figuras 12,13, 14,15 e 16), o que significa uma reclassificação dos animais graças à atribuição de UPG. Para esta 1% dos piores animais nos modelos sem informação genômica o r_R entre BLUP com UPGY e UPGS foi de 0,3 e 0,98 para PD, 0,42 e 0,95 para GDPSOB, 0,88 e 0,96 par PE, 0,51 e 0,91 para MUSC e 0,55 e 0,99 para PES18, com diferenças sutis com o 10% dos piores animais, dando menores mudanças na classificação de animais no UPGS do que usando UPGY, em todas as características. Já nos modelos com informação genômica estes animais (1% e 10% PIORES) seguiram o mesmo padrão de r_R do que comparado com os valores da população total e na seleção dos melhores animais (1%, 10% e 30% TOP) ainda que os valores som menores, isto significa que a informação genômica proporciona uma certa estabilidade na classificação dos animais, assim como mostraram Tonussi et al. (2021) ao identificar não seu trabalho que as precisões foram sempre maiores quando as informações genômicas foram incluídas, isto é devido em parte a captura de alguma contribuição poligênica pelos grupos (SPEED; BALDING, 2015).

Figura 12- Correlação de $rank-r_R$ dos valores genéticos ao selecionar o 1% e 10% dos piores animais para peso á desmama (PD) nos diferentes modelos



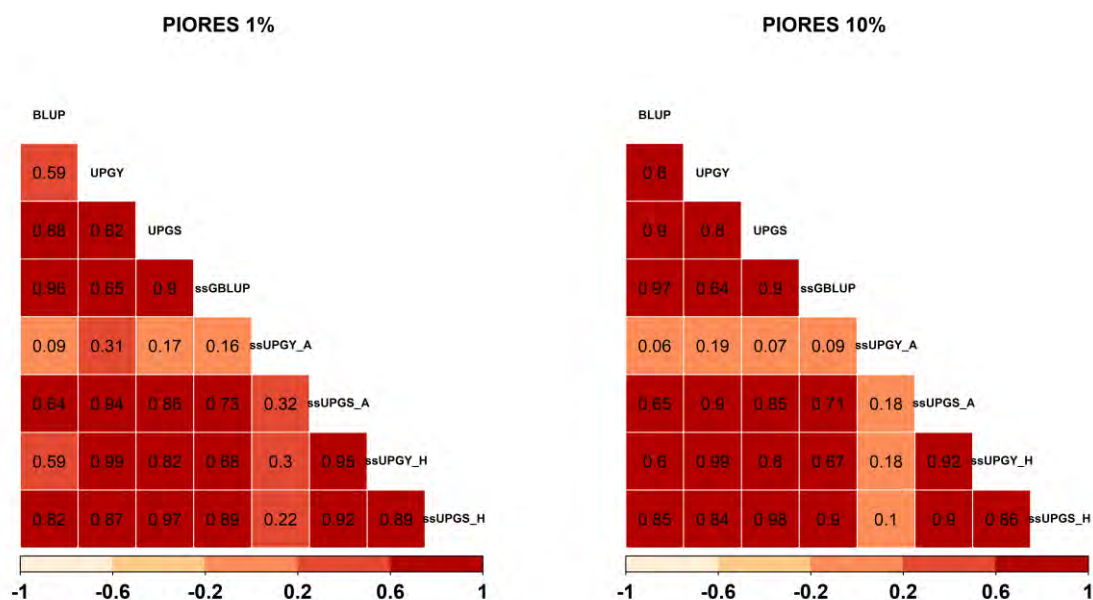
Fonte: autoria própria

Figura 13- Correlação de $rank-r_R$ dos valores genéticos ao selecionar o 1% e 10% dos piores animais para ganho de peso ao sobre ano (GDPSOB) nos diferentes modelos



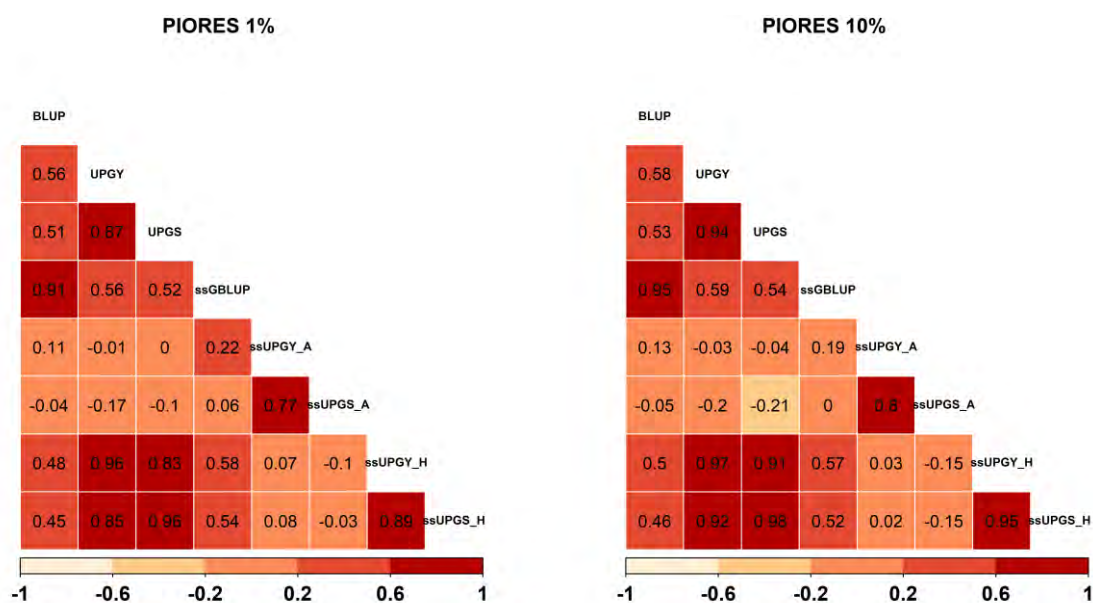
Fonte: autoria própria

Figura 14- Correlação de $rank-r_R$ dos valores genéticos ao selecionar o 1% e 10% dos piores animais para perímetro escrotal (PE) nos diferentes modelos



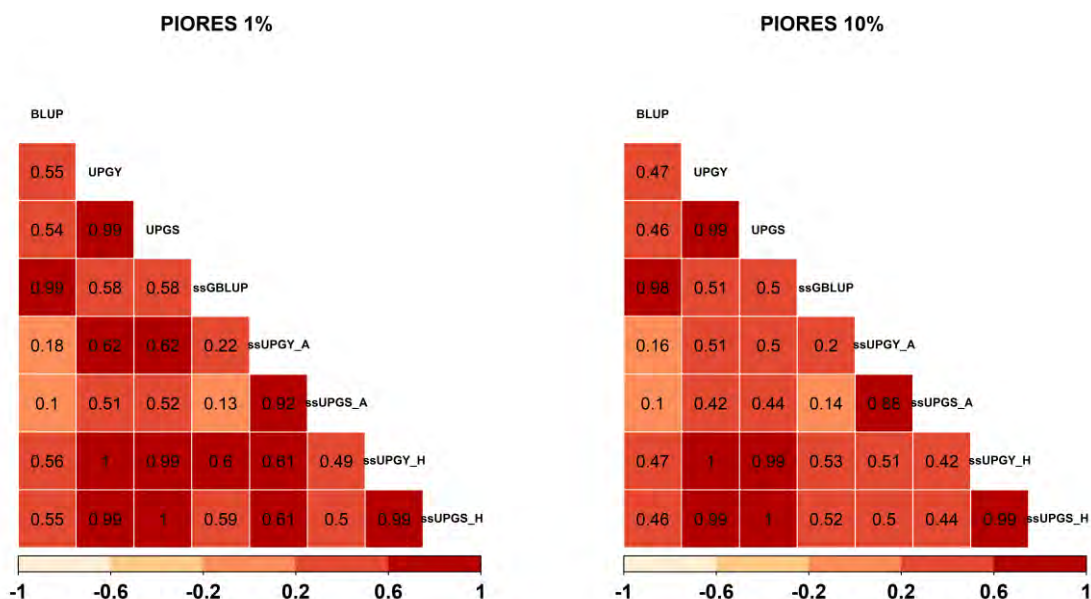
Fonte: autoria própria

Figura 15- Correlação de $rank-r_R$ dos valores genéticos ao selecionar o 1% e 10% dos piores animais score de musculabilidade (MUSC) nos diferentes modelos



Fonte: autoria própria

Figura 16- Correlação de $rank-r_R$ dos valores genéticos ao selecionar o 1% e 10% dos piores animais peso aos 18 meses (PES 18) nos diferentes modelos



Fonte: autoria própria

O comportamento dos coeficientes de correlação tanto de Pearson e de $rank$ das estimativas individuais dos GEBV diferem entre modelos mesmo se a tendência é semelhante ao BLUP ou ssGBLUP (MATILAINEN et al., 2018). As tendências genéticas da população foram obtidas para as características de estudo, mostrando a importância do desenvolvimento do programa de melhoramento nos últimos anos. O ganho genético anual foi definido como coeficiente de regressão linear da média do valor genético da população em função do ano de nascimento do animal (safra) nos últimos 12 anos. Os resultados mostram a eficiência do programa de seleção genética que vem sendo implementado. Todas as características mostram progresso genético, o PES18 apresentou, por exemplo um ganho genético anual de 0,2 Kg/ano (Figura 21) tanto no modelo tradicional quanto para o ssGBLUP, reflexo da herdabilidade média a alta (0,35) estimada para esta característica. As tendências genéticas não são apresentadas pois os EVBs apresentaram valores extrapolados com as características e em alguns modelos não se consegue uma gráfica de tendência.

O ganho médio do PD foi de +0,1 kg, +0,18 kg, +0,21 kg nos modelos BLUP, UPGY e UPGS, respectivamente, seguindo uma tendência similar entre eles. Mas, nos modelos ssUPGY_H e ssUPGS_H o ganho médio foi de -94090 e 92,8, ou seja estão deflacionados os (G)EBV, sugerindo que o modelo não levou em conta totalmente o desenvolvimento genético ao longo do tempo (KENNEDY; MOXLEY,

1975). No entanto, para as outras características não houve valores negativos no ganho anual, para GDP SOB e MUSC o ganho genético maior foi calculado no modelo ssUPGS_A de 7,96 kg/dia e 1,7 und, respectivamente, para PES18 e PE foi no modelo ssUPGY_A de 2,64 kg e 1,12 cm, respectivamente. As tendências observadas tiveram diferenças com a inclusão dos UPG como inflação ou deflação dos EBV, isso pode ser resultado pela própria estrutura dos UPG para cada característica (PETRINI et al., 2015), o ganho genético foi maior nos modelos que incluíram o UPG) esse aumento das tendências ao incluir os UPG também foi exposto por Theron et al. (2002), exceto para os descritos para PD, mas é explicado pela forte seleção das características o que amplifica a inflação nas previsões (TSURUTA et al., 2019), em cada característica e modelo as tendências ascendentes podem indicar que o tipo de agrupamento dá indícios da melhor estratégia de acordo á característica e considerando as mudanças genéticas ao longo do tempo (GÓMEZ et al., 2021).

4.4 Conclusão

Os coeficientes de correlação altos indicaram previsões mais precisas nos modelos com informação genômica, com GEBVs menos viesados quando comparados aos modelos sem genômica.

A inclusão de UPG nos modelos sem genômica não provocou impacto na classificação dos animais. Para as características pós desmama nesta população, os UPG considerados em toda a matriz H, propiciou melhores resultados. Recomenda-se conhecimento prévio das particularidades de cada população e cada característica e cautela na utilização da informação genômica nos UPG.

Referências

- AGUILAR, I. *et al.* Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. **Journal of Dairy Science**, Lancaster, v. 93, n. 2, p. 743–752, 1 fev. 2010.
- BARBOSA, P. F. Bovinos-raças puras, novas raças, cruzamentos e compostos de gado de corte. *In*: SEMINÁRIO NACIONAL SOBRE PRODUÇÃO DE CARNE BOVINA COM QUALIDADE, 3., 2003, Gurupi. **Anais [...]**. Gurupi, 2003. p. 124-129.
- BOLIGON, A. A. *et al.* Revista Brasileira de Zootecnia Herdabilidades e correlações entre pesos do nascimento à idade adulta em rebanhos da raça Nelore 1. **Revista Brasileira de Zootecnia**, Viçosa, v. 38, n. 12, p. 2320–2326, 2009.
- BONIFAZI, R. *et al.* Impact of sub-setting the data of the main Limousin beef cattle population on the estimates of across-country genetic correlations. **Genetics**

Selection Evolution, London, v. 52, n. 1, p. 1–16, 2020.

CARDOSO, V. et al. Formando lotes uniformes de reprodutores múltiplos e usando-os em acasalamentos dirigidos, em populações nelore. **Revista Brasileira de Zootecnia**, Viçosa, v. 32, n. 4, p. 834–842, jul. 2003.

ELER, J. P. **Informações pessoais preliminares sobre a avaliação genética do rebanho Nelore pertencente à Agropecuária**. CFM, 2019.

FAMULA, T. R.; POLLAK, E. J.; VAN VLECK, L. D. Genetic Groups in Dairy Sire Evaluation Under a Selection Model. **Journal of Dairy Science**, Lancaster, v. 66, n. 4, p. 927–934, 1Apr. 1983.

GÓMEZ, M. *et al.* Accounting for genetic differences among unknown parents in bubalus bubalis: a case study from the Italian Mediterranean Buffalo. **Frontiers in Genetics**, Lausanne, v. 12, art. 625335, 4 Feb. 2021.

KENNEDY, B. W. Bias and mean square error from ignoring genetic groups in mixed model sire evaluation. **Journal of Dairy Science**, Lancaster, v. 64, n. 4, p. 689–697, 1 abr. 1981.

KENNEDY, B. W.; MOXLEY, J. E. Comparison of Genetic Group and Relationship Methods for Mixed Model Sire Evaluation. **Journal of Dairy Science**, Lancaster, v. 58, n. 10, p. 1507–1514, 1975.

KOETZ, C. *et al.* Heritability estimation and genetic correlations for mature weight, visual scores, and growth traits in Nelore cattle. **Revista Brasileira de Zootecnia**, Viçosa, v. 48, art. e20170246, 2019.

KOURY FILHO, W. *et al.* Estimativas de parâmetros genéticos para os escores visuais e suas associações com peso corporal em bovinos de corte. **Revista Brasileira de Zootecnia**, Viçosa, v. 39, n. 5, p. 1015–1022, maio. 2010.

LEGARRA, A.; AGUILAR, I.; MISZTAL, I. A relationship matrix including full pedigree and genomic information. **Journal of Dairy Science**, Lancaster, v. 92, n. 9, p. 4656–4663, 1 set. 2009.

MACEDO, F. L. *et al.* Bias and accuracy of dairy sheep evaluations using BLUP and SSGBLUP with metafounders and unknown parent groups. **Genetics, selection, evolution : GSE**, v. 52, n. 1, p. 47, 12 ago. 2020.

MATILAINEN, K. *et al.* Single step genomic evaluation for female fertility in Nordic Red dairy cattle. **Journal of Animal Breeding and Genetics**, Berlin, v. 135, n. 5, p. 337–348, 1 Oct. 2018.

MEUWISSEN, T.; HAYES, B.; GODDARD, M. Genomic selection: A paradigm shift in animal breeding. **Animal Frontiers**, [s.l.], v. 6, n. 1, p. 6–14, 1 Jan. 2016.

MISZTAL, I. *et al.* BLUPF90 and related programs. *In: WORLD CONGRESS ON GENETICS APPLIED TO LIVESTOCK PRODUCTION, 7., 2002, Montpellier. Proceedings [...].* Montpellier, 2002. session 28.

MISZTAL, I. *et al.* Unknown-parent groups in single-step genomic evaluation. **Journal of Animal Breeding and Genetics**, Berlin, v. 130, n. 4, p. 252–258, 1 Aug. 2013.

MISZTAL, I. *et al.* **Manual for BLUPF90 family of programs**. 2015. 125 p.

MISZTAL, I. *et al.* Studies on Inflation of GEBV in Single-Step GBLUP for Type. **Interbull Buletin**, [s.l.], n. 51, p. 38–42, 2017.

MISZTAL, I.; LOURENCO, D.; LEGARRA, A. Current status of genomic evaluation. **Journal of Animal Science**, Champaign, v. 98, n. 4, p. 1–14, 1 Apr. 2020.

PEŠKOVIČOVÁ, D.; GROENEVELD, E.; WOLF, J. Effect of genetic groups on the efficiency of selection in pigs. **Livestock Production Science**, Amsterdam, v. 88, n. 3, p. 213–222, 1 July. 2004.

PETRINI, J. **Grupos genéticos na eficiência de seleção de bovinos de corte compostos (Bos taurus x Bos indicus)**. 2012. 157 p. Dissertação (Mestrado) - Universidade de São Paulo Escola Superior de Agricultura “Luiz de Queiroz”, Piracicaba, 2012.

PETRINI, J. *et al.* Genetic grouping strategies in selection efficiency of composite beef cattle (*Bos taurus* x *Bos indicus*). **Journal of Animal Science**, Champaign, v. 93, n. 2, p. 541–552, 1 Feb. 2015.

QUAAS, R. L. Additive genetic model with groups and relationships. **Journal of Dairy Science**, Lancaster, v. 71, n. 5, p. 1338–1345, 1 June. 1988.

QUAAS, R. L.; POLLAK, E. J. Modified equations for sire models with groups. **Journal of Dairy Science**, Lancaster, v. 64, n. 9, p. 1868–1872, 1 Sept. 1981.

R CORE TEAM. **R: A language and environment for statistical computing**. Vienna, Austria: R Foundation for Statistical Computing, 2019. Disponível em: <https://www.r-project.org/>. Acesso em: 10 fev. 2022.

RAIMUNDO MARTINS, T. **Parâmetros genéticos para características de importância zootécnica em bovinos da raça senepol**. 2018. Dissertação (Mestrado) - Universidade Federal de Goiás, Goiânia, 2018.

SARGOLZAEI, M.; CHESNAIS, J. P.; SCHENKEL, F. S. A new approach for efficient genotype imputation using information from relatives. **BMC Genomics**, Heidelberg, v. 15, n. 1, 2014.

SPEED, D.; BALDING, D. J. Relatedness in the post-genomic era: Is it still useful? **Nature Reviews Genetics**, London, v. 16, n. 1, p. 33–44, 18 Nov. 2015.

THERON, H. E.; KANFER, F. H. J.; RAUTENBACH, L. The effect of phantom parent groups on genetic trend estimation. **South African Journal of Animal Sciences**, Champaign, v. 32, n. 2, p. 130–135, 2002.

TONG, A. K. W.; KENNEDY, B. W.; MOXLEY, J. E. Potential Errors in Sire Evaluation from Regional Genetic Differences. **Journal of Dairy Science**, Lancaster, v. 63, n. 4, p. 627–633, 1980.

TONUSSI, R. *et al.* Application of single step genomic BLUP under different uncertain paternity scenarios using simulated data. **PLoS ONE**, San Francisco, v. 12, n. 9, art. e0181752, 2017.

TONUSSI, R. L. *et al.* Accuracy of genomic breeding values and predictive ability for postweaning liveweight and age at first calving in a Nellore cattle population with missing sire information. **Tropical Animal Health and Production**, Dordrecht, v. 53, n. 4, p. 1–9, 10 Aug. 2021.

TSURUTA, S. *et al.* Assigning unknown parent groups to reduce bias in genomic evaluations of final score in US Holsteins. **Journal of Dairy Science**, Lancaster, v. 97, n. 9, p. 5814–5821, 2014.

TSURUTA, S. *et al.* Controlling bias in genomic breeding values for young genotyped bulls. **Journal of Dairy Science**, Lancaster, v. 102, n. 11, p. 9956–9970, 1 Nov. 2019.

VANRADEN, P. M. Accounting for Inbreeding and Crossbreeding in Genetic Evaluation of Large Populations. **Journal of Dairy Science**, Lancaster, v. 75, n. 11, p. 3136–3144, 1 Nov. 1992.

VANRADEN, P. M. Efficient methods to compute genomic predictions. **Journal of Dairy Science**, Lancaster, v. 91, n. 11, p. 4414–4423, 1 Nov. 2008.

WIGGANS, G. R. *et al.* Selection of single-nucleotide polymorphisms and quality of genotypes used in genomic evaluation of dairy cattle in the United States and Canada. **Journal of Dairy Science**, Lancaster, v. 92, n. 7, p. 3431–3436, 2009.

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

WOLAK, M. E.; REID, J. M. Accounting for genetic differences among unknown parents in microevolutionary studies: how to include genetic groups in quantitative genetic animal models. **Journal of Animal Ecology**, London, v. 86, n. 1, p. 7–20, 1 jan. 2017.