

ANA PAULA SOARES E SILVA MATIAS

**ESTRATÉGIAS PARA IMPLEMENTAÇÃO DA SELEÇÃO GENÔMICA EM
OVINOS DE CORTE: UM ESTUDO DE SIMULAÇÃO**

TERESINA – PIAUÍ

2024

ANA PAULA SOARES E SILVA MATIAS

**ESTRATÉGIAS PARA IMPLEMENTAÇÃO DA SELEÇÃO GENÔMICA EM
OVINOS DE CORTE: UM ESTUDO DE SIMULAÇÃO**

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

Área de concentração: Produção animal.

Orientador: Prof. Dr. José Lindenberg Rocha Sarmiento

Coorientador: Dr. André Campêlo Araújo

TERESINA – PIAUÍ

2024

FICHA CATALOGRÁFICA
Universidade Federal do Piauí
Biblioteca Setorial CCA
Serviço de Representação da Informação

M433e Matias, Ana Paula Soares e Silva.
Estratégias para implementação da seleção genômica em ovinos
de corte: um estudo de simulação. / Ana Paula Soares e Silva Matias.
-- 2024.
96 f.

Tese (Doutorado) – Universidade Federal do Piauí, Centro de
Ciências Agrárias, Programa de Pós-Graduação em Zootecnia
Tropical, 2025.

“Orientador: Prof. Dr. José Lindenberg Rocha Sarmento.”

1. Herdabilidade. 2. SNP chip. 3. Predição genética. 4. Ovino . I.
Sarmento, José Lindenberg Rocha. II. Título.

CDD 636.3

Bibliotecário: Rafael Gomes de Sousa - CRB3/1163

ESTRATÉGIAS PARA IMPLEMENTAÇÃO DA SELEÇÃO GENÔMICA EM OVINOS DE CORTE: UM ESTUDO DE SIMULAÇÃO

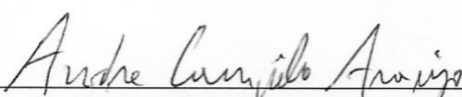
ANA PAULA SOARES E SILVA MATIAS

Tese aprovada em 14 de Agosto de 2024.

Banca Examinadora:

Documento assinado digitalmente
gov.br JOSE LINDENBERG ROCHA SARMENTO
Data: 24/10/2024 15:05:05-0300
Verifique em <https://validar.iti.gov.br>

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


Prof. Dr. André Campêlo Araújo (Externo) AcuFast/Canadá

Documento assinado digitalmente
gov.br NATANAEL PEREIRA DA SILVA SANTOS
Data: 24/10/2024 17:04:43-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. Natanael Pereira da Silva Santos (Interno) DZO/CCA/UFPI

Documento assinado digitalmente
gov.br FABIO BARROS BRITTO
Data: 24/10/2024 21:08:01-0300
Verifique em <https://validar.iti.gov.br>

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

Documento assinado digitalmente
gov.br GLEYSON VIEIRA DOS SANTOS
Data: 25/10/2024 11:06:31-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. Gleyson Vieira Dos Santos (Externo) UESPI

Documento assinado digitalmente
gov.br JOHNNY IGLESIAS MENDES ARAUJO
Data: 31/10/2024 09:21:51-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. Johnny Iglesias Mendes Araujo (Externo) /SEDUC/GO

AGRADECIMENTOS

À Deus, pois sem Ele eu não teria conseguido concluir essa jornada. Somente Ele e eu sabemos o quanto foi difícil e árduo, mas jamais estive desamparada.

Ao professor Lindenberg Rocha Sarmento, pela orientação e paciência. Reconheço todo o cuidado e incentivo nos momentos de dificuldade. Tenho profunda admiração e respeito pelo trabalho que você desenvolve no Grupo de Estudos em Genética e Melhoramento (GEMA).

Ao Artur Oliveira Rocha, pelas assistências, sempre muito prestativo e atencioso. Sua ajuda foi de fundamental importância, e sou muito grata por isso.

À Bruna Lima Barbosa, por ter me acompanhado durante toda a trajetória e, por muitas vezes ter dividido a carga, sempre incentivando e dando suporte. Agradeço também pela amizade que construímos e por ser esse ser humano incrível.

Ao meu marido, Jefferson Lucas Matias Sousa, e aos meus filhos, Clara Silva Matias e João Gabriel Silva Matias, por serem a razão de tudo, meu alicerce, e por fazerem todo o esforço valer a pena. Após um dia cansativo, quando as forças parecem ter se esgotado, lembrar de vocês me renova.

Às minhas amigas de vida, Dayane Pereira de Castro e Tamires de Sousa Silva. Com vocês tudo vira motivo para risada, sempre nos fazemos presentes na vida uma da outra, nem que seja só para perturbar (rsr).

Ao Luís Fernando Brito, por ter disponibilizado os meios necessários para a realização das análises deste trabalho.

Ao meu coorientador, André Campêlo Araújo, sem dúvida, uma das pessoas mais inspiradoras que conheci nos últimos anos. Nenhuma palavra pode expressar toda a minha admiração e gratidão a você. Uma pessoa simples, que sempre encontra uma maneira de descomplicar as coisas mais complexas (essa é a melhor descrição que posso fazer). Apesar da correria que é a sua vida, você nunca deixou de me apoiar e segurar minha mão quando precisei. O mundo e a Academia precisam de mais pessoas como você.

À minha mãe, Dorides Soares e Silva, e ao meu pai, Martins Alves e Silva. Apesar da pouca instrução de minha mãe, sempre incentivou e fez de tudo para que seus filhos pudessem estudar, mesmo em uma realidade cheia de dificuldades. Aos meus 10 irmãos, especialmente à minha irmã, Magda Soares e Silva (*in memoriam*), essa conquista também é sua. Penso em você todos os dias.

À Universidade Federal do Piauí e à Fundação Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)."

SUMÁRIO

LISTA DE FIGURAS	vii
LISTA DE TABELAS	viii
RESUMO	ix
ABSTRACT	xi
1 INTRODUÇÃO GERAL	13
2 REFERENCIAL TEÓRICO	15
2.1 Aspectos da ovinocultura no Brasil e no mundo	15
2.2 Contribuições da seleção genômica no melhoramento animal.....	17
2.3 Desafios para a implementação da seleção genômica	19
CAPÍTULO I	33
Imputação de genótipos em população de ovinos de corte simuladas com diferentes estruturas de referência e validação	33
1 Introdução.....	37
2 Material e Métodos.....	37
3 Resultados.....	42
4 Discussão	49
5 Conclusão	52
CAPÍTULO II.....	56
Predição genômica com dados simulados de ovinos em programa de melhoramento usando diferentes estratégias de imputação	56
1 Introdução.....	59
2 Material e Métodos.....	60
3 Resultados.....	64
4 Discussão	67
5 Conclusão	69

LISTA DE FIGURAS

REFERENCIAL TEÓRICO

Figura 1. Processo de imputação para um haplótipo de estudo (SG) genotipado em 6 marcadores usando um painel de referência de haplótipos sequenciados em 21 marcadores. Adaptada de: Das; Abecasis; Browning, 2018.....18

CAPÍTULO I

Figura 1. Esquema de simulação para obtenção de populações sem troca de reprodutores e com troca de reprodutores na espécie ovina. O mesmo *design* foi aplicado para obtenção de 10 gerações com troca de reprodutores.....36

Figura 2. Cenários avaliados para imputação de genótipos em dados simulados de ovinos, considerando 2 herdabilidades (0,30 e 0,10); duas estratégias de genotipagem (100% e 5%), um painel de SNPS (50K) e 4 populações, sendo duas sem troca de reprodutores e com N_e distintos (300 e 80) e duas com troca de reprodutores (TR: 50% e TR: 5%) e mesmo N_e (300) 39

Figura 3. Acurácia de imputação de SNPs para a característica simulada com herdabilidade 0,30. **a:** POP1_ST20K100%80. **b:** POP1_CT20K100%300, TR: 50%. **c:** POP1_ST1K5%80. **d:** POP1_CT1K5%300, TR: 50%. **e:** POP2_ST20K100%300. **f:** POP2_CT20K100%300, TR: 5%. **g:** POP2_ST1K5%300. **h:** POP2_CT1K5%300, TR: 5%.....42

Figura 4. Acurácia de imputação de SNPs para a característica simulada com herdabilidade 0,10. **a:** POP1_ST20K100%80. **b:** POP1_CT20K100%300, TR: 50%. **c:** POP1_ST1K5%80. **d:** POP1_CT1K5%300, TR: 50%. **e:** POP2_ST20K100%300. **f:** POP2_CT20K100%300, TR: 5%. **g:** POP2_ST1K5%300. **h:** POP2_CT1K5%300, TR: 5%.....43

Figura 5. Distribuição da frequência de alelo menos frequente (MAF). **a:** população com menor número efetivo (N_e =80). **b:** população com maior número efetivo (N_e =300) 44

Figura 6. Acurácia de imputação de indivíduos para a característica simulada com herdabilidade 0,30. **a:** POP1_ST20K100%80. **b:** POP1_CT20K100%300, TR: 50%. **c:** POP1_ST1K5%80. **d:** POP1_CT1K5%300, TR: 50%. **e:** POP2_ST20K100%300. **f:** POP2_CT20K100%300, TR: 5%. **g:** POP2_ST1K5%300. **h:** POP2_CT1K5%300, TR: 5%.....46

Figura 7. Acurácia de imputação de indivíduos para a característica simulada com herdabilidade 0,10. **a:** POP1_ST20K100%80. **b:** POP1_CT20K100%300, TR: 50%. **c:** POP1_ST1K5%80. **d:** POP1_CT1K5%300, TR: 50%. **e:** POP2_ST20K100%300. **f:** POP2_CT20K100%300, TR: 5%. **g:** POP2_ST1K5%300. **h:** POP2_CT1K5%300, TR: 5%.....47

LISTA DE TABELAS

CAPÍTULO I

Tabela 1. Distribuição de marcadores SNPs por cromossomo em chip de moderada (50K) e baixa (6K) densidade.....40

Tabela 2. Média de acurácia de imputação de SNPs para cada cenário avaliado contendo duas herdabilidades (0,30 e 0,10), valores mínimos (Mín), Máximos (Máx) e desvios padrão (SD).....41

Tabela 3. Média de acurácia de indivíduos para cada cenário avaliado contendo duas herdabilidades (0,30 e 0,10), valores mínimos (Mín), Máximos (Máx) e desvios padrão (SD).....45

CAPÍTULO II

Tabela 1. Medias e desvios-padrão (SD) da acurácia e viés de predição e dispersão dos GEBVs utilizando dados reais e imputados de ovinos obtidos a partir de simulação.....63

Tabela 2. Comparações dos cenários de predição em estudo com valores imputados utilizando acurácia, viés, dispersão e os desvios-padrão (SD) para cada métrica.....64

RESUMO

MATIAS, Ana Paula Soares e Silva. **Estratégias para implementação da seleção genômica em ovinos de corte: Um estudo de simulação.** 2024. Tese (Doutorado em Ciência Animal) – Universidade Federal do Piauí, Teresina, Piauí.

A genômica revolucionou os programas de melhoramento genético animal e garantiu importantes ganhos produtivos em características de interesse econômico. Nesse aspecto, a simulação de dados otimizou o desenvolvimento de métodos genômicos. Isso devido proporcionar avaliação de cenários que imitam populações reais a relativamente baixo custo e menos tempo. Neste estudo, objetivou-se investigar estratégias de implementação da seleção genômica em ovinos por meio do uso da simulação de dados com foco em programas de melhoramento de ovinos em regiões/países em desenvolvimento, como o Nordeste do Brasil. Cenários de imputação de genótipos e predição/seleção genômica foram investigados. No capítulo 1, através de dados gerados pelo *software* QMSim foram simulados 16 cenários utilizando duas herdabilidades (0,30 e 0,10), números de animais genotipados (20K e 1K), estruturas de rebanho (com e sem troca de reprodutores). A imputação foi feita de um SNP chip de baixa densidade (6K) para um chip média densidade (50K). Dentro de cada cenário gerado foram realizados dois métodos de imputação, um populacional e outro que considera a informação de família (pedigree), totalizando 32 cenários. Os resultados foram comparados usando a acurácia de imputação de SNPs e indivíduos, através da correlação entre as matrizes padronizada dos genótipos imputados e reais. Verificou-se que as acurácias de imputação de SNPs variaram de 0,922 a 0,96 nos cenários com tamanho efetivo maior e de 0,927 a 0,944 nos tamanhos efetivos menores. Nos cenários com maior número de animais genotipados as médias de acurácia variaram de 0,943 a 0,96 e de 0,922 a 0,934 nos cenários com menos animais genotipados. O uso de diferentes estruturas de rebanhos resultou em acurácias de imputação semelhantes. Os resultados de acurácia de imputação de indivíduos foram semelhantes aos de acurácia de imputação de SNPs descritos anteriormente. Observou-se menor dispersão da acurácia de imputação de SNPs nos cenários com maior N_e , assim como também nos cenários em que houve maior quantidade de animais genotipados. Os cenários de imputação avaliados foram eficientes e permitiram obter acurácias de imputação maiores que 0,90 em todos os cenários avaliados. No capítulo 2, todos os cenários criados anteriormente com a herdabilidade 0,10 foram utilizados para análises de predição genômica, no intuito de avaliar o efeito das estratégias de imputação sobre as acurácias de predição. O método single-step GBLUP foi utilizado para a predição dos GEBVs usando ambos os genótipos imputados e reais. Para as medidas de validação, a acurácia de predição foi obtida pela correlação de Pearson entre o TBV e o GEBV na população de validação. O viés de predição foi avaliado pela diferença entre a média do GEBV e TBV e a dispersão dos GEBVs foi medida como 1 menos o coeficiente de regressão do TBV sobre o GEBV. As médias dos cenários foram comparadas pelo teste t e tukey a 5%. Os resultados mostraram que não houve diferenças entre os cenários reais e imputados para todas as métricas avaliadas pelo teste t entre pares de média (cenários com genótipos imputados e reais) para todas as medidas de validação utilizadas nesse estudo. No entanto, a comparação entre diferentes cenários imputados mostrou diferenças nas três métricas pelo teste tukey a 5%. Maiores médias de acurácias foram observadas nas populações com maior porcentagem de troca de reprodutores (0,8500) e menores médias ocorreram nas populações sem troca (0,4140). Houve tendência de aumento do viés nas populações sem troca de reprodutores (-18,5096) e menores vieses foram observados nas populações com troca (-13,2610). As dispersões foram relativamente baixas em todos os cenários avaliados. Os cenários de predição avaliados tiveram melhores valores de acurácia, viés e dispersão quando

se tinha maior proporção de troca de reprodutores. A imputação dos genótipos não afetou a predição dos valores genéticos.

Palavras-chave: Herdabilidade, SNP chip, Predição genômica, Ovino, Simulação.

ABSTRACT

MATIAS, Ana Paula Soares e Silva. **Strategies for implementing genomic selection in meat sheep: A simulation study.** 2024. Tese (Doutorado em Ciência Animal) – Universidade Federal do Piauí, Teresina, Piauí.

Genomics has revolutionized animal genetic improvement programs and guaranteed important productive gains in traits of economic interest. In this respect, data simulation has optimized the development of genomic methods. This is because it provides assessment of scenarios that imitate real populations at relatively low cost and less time. The objective of this study was to investigate strategies for implementing genomic selection in sheep through the use of data simulation with a focus on sheep improvement programs in developing regions/countries, such as the Northeast of Brazil. Genotype imputation and genomic prediction/selection scenarios were investigated. In chapter 1, using data generated by QMSim software, 16 scenarios were simulated using two heritabilities (0.30 and 0.10), numbers of animals (20K and 1K), herd structures (with and without changing breeders). Imputation was done from a low density SNP chip (6K) to a medium-density SNP chip (50K). Within each scenario generated, two imputation methods were carried out, one population-based and the other that considers family relationships, totaling 32 scenarios. The results were compared using the imputation accuracy of SNPs and individuals, through the correlation between the standardized matrices of the imputed and real genotypes. It was found that SNP imputation accuracies varied from 0.922 to 0.96 in scenarios with larger effective sizes and from 0.927 to 0.944 in smaller effective sizes. In scenarios with a greater number of genotyped animals, average accuracy ranged from 0.943 to 0.96 and from 0.922 to 0.934 in scenarios with fewer genotyped animals. Herd structures (with or without sire exchange) resulted in similar accuracy values, between 0.922 and 0.96. The same minimums and maximums were observed for the accuracy values of individuals within the scenarios described previously. A better dispersion of SNPs and accuracy values was observed in scenarios with higher N_e , as well as in those in which there were a greater number of genotyped animals. Population size also influenced the increase in accuracy. The assessed imputation scenarios were efficient and allowed imputation accuracies greater than 0.90 to be obtained in all evaluated scenarios. Chapter 2 utilized all previously created scenarios for genomic selection/prediction, in order to evaluate the effect of imputation strategies on prediction accuracies. In chapter 2, all previously created scenarios with heritability 0.10 were used for genomic prediction analyses, in order to evaluate the effect of imputation strategies on prediction accuracies. The single-step GBLUP method was used for the prediction of GEBVs using both imputed and real genotypes. For validation measurements, prediction accuracy was obtained by Pearson's correlation between TBV and GEBV in the validation population. Prediction bias was assessed by the difference between the mean of GEBV and TBV and the dispersion of GEBVs was measured as 1 minus the regression coefficient of TBV on GEBV. The means of the scenarios were compared using the t and Tukey test at 5%. The results indicated no significant differences between the real and imputed scenarios for all metrics evaluated by the t test between pairs of means (scenarios with imputed and real genotypes) for all validation measures used in this study. However, the comparison between different imputed scenarios showed differences in the three metrics using the 5% Tukey test. Higher average accuracies were observed in populations with a higher percentage of reproductive exchange (0.8500) and lower averages occurred in populations without exchange (0.4140). There was a tendency for bias to increase in populations without breeder exchange (-18.5096) and smaller biases were observed in populations with exchange (-13.2610). Dispersions were relatively low in all scenarios evaluated. The prediction scenarios evaluated had better accuracy, bias and

dispersion values when there was a higher proportion of reproductive exchange. Genotype imputation did not affect the prediction of genetic values.

Keywords: Heritability, SNP chip, Genomic prediction, Sheep, Simulation.

1 INTRODUÇÃO GERAL

Melhorar características de interesse econômico em animais de corte é um esforço constante na produção animal. Além de aprimorar as condições ambientais de criação, é essencial o desenvolvimento de estratégias para a seleção de indivíduos que possuam alelos favoráveis para melhorar características de interesse e aumentar a produtividade animal. O melhoramento genético clássico, baseado no fenótipo e pedigree (método tradicional), por muito tempo foi a base para avaliação de indivíduos superiores, porém, recentemente, a seleção genômica tem revolucionado a produção animal (Navajas, 2014; Weiler; Fernando, 2020).

Aumentar a eficiência de animais de corte tornou-se uma tarefa desafiadora para pesquisadores e criadores, quando se considera a lacuna existente entre a produção animal e a demanda da população mundial por proteína de origem animal. Nesse sentido, a seleção genômica, comparada ao método tradicional, proporciona a predição de valores genéticos mais acurados, o que é benéfico para os programas de melhoramento (Meuwissen; Hayes; Goddard, 2001). O princípio dessa técnica é o uso de marcadores genéticos, principalmente os polimorfismos de nucleotídeos únicos (*single nucleotide polymorphism* - SNPs). Os SNPs são marcadores altamente repetitivos no genoma (Reshma; Das, 2023) e são utilizados para cobrir o genoma em alta densidade para estabelecer desequilíbrio de ligação (LD) com *loci* de características quantitativas (*Quantitative Trait loci* - QTL). A seleção genômica proporciona a redução no intervalo de gerações e aumento na acurácia preditiva de valores genéticos (EBV) (Meuwissen Hayes; Goddard, 2001; Shumbusho *et al.*, 2016), e tem alavancado o progresso nos programas de melhoramento tanto animal como vegetal (Lourenco *et al.*, 2020).

A implementação de programas de seleção genômica é um caminho promissor na busca de melhorar a produtividade animal em menor tempo. Essa metodologia baseia-se na formação de uma população de treinamento (fenótipo e genótipo conhecidos) e validação (apenas genótipos conhecidos), que proporciona a possibilidade de predizer a valor genético de indivíduos jovens apenas com genótipo de forma acurada (Meuwissen; Hayes; Goddard, 2001; 2016). No entanto, antes da devida implementação, é necessário validar a seleção genômica através do uso de populações de validação e de medidas de acurácia, viés, dispersão, entre outras (Legarra; Reverter, 2018). O planejamento adequado dessa etapa e a comparação dos custos e benefícios associados são fundamentais (VanRaden, 2020). Adicionalmente, são necessários um número de indivíduos e uma densidade de marcadores apropriados, para assegurar a confiabilidade na predição de valores genéticos genômicos estimados (GEBV) (Neves *et al.*, 2014).

No Brasil, há poucos programas de melhoramento genético para ovinos. Devido à desestruturação da cadeia produtiva, o baixo investimento tecnológico e o tamanho dos rebanhos, geralmente criados por pequenos criadores (Morris *et al.*, 2017). Aliado a esses fatores, o aumento dos custos na manutenção dos rebanhos e de genotipagem de animais são também desafios para implantação da seleção genômica nessa espécie (Sarmiento; Sena, 2017). Estratégias para minimizar esses custos têm sido consideradas no esforço dos pesquisadores da área de Genética e Melhoramento Animal, a fim de viabilizar a implementação de metodologias eficazes de avaliação genética.

O uso da simulação de dados tem sido uma proposta relevante por fornecer subsídios para a realização de estudos a partir da imitação de cenários reais (Daetwyler *et al.*, 2013, Piccoli *et al.* 2018). A imputação de genótipos também contribui como estratégia para viabilizar o progresso genético. Essa técnica explora o compartilhamento de haplótipos entre os indivíduos para preencher os genótipos faltantes, em que é possível obter painéis de SNPs de alta densidade a partir de painéis de menor densidade (Li *et al.*, 2009; 2022). Além de reduzir gastos com a genotipagem de animais para o criador, por meio do uso de painéis menos densos, essa técnica também pode ser utilizada para aumentar a intensidade da seleção. No entanto, para melhor eficiência da seleção genômica, a etapa de verificação da acurácia de imputação como pré-requisito é crucial (Calus *et al.*, 2014). Na espécie bovina alguns trabalhos já foram publicados com esse intuito (Carvalho *et al.*, 2014; Chud *et al.*, 2015; Piccoli *et al.*, 2014), porém trabalhos com ovinos ainda existem poucos.

Diversos estudos já foram realizados na espécie ovina a partir de dados simulados (Latifi *et al.*, 2020; Martin *et al.*, 2023; Piccoli *et al.*, 2018). As vantagens dessa ferramenta incluem a possibilidade de reproduzir cenários reais com o uso de parâmetros que estejam de acordo com a população alvo para estudar características de interesse e permitir testes rápidos e replicáveis para diversas hipóteses a baixo custo. Como exemplo, a viabilidade inicial da seleção genômica e o impacto do tamanho populacional de referência pode ser feito antes de realizar a genotipagem dos animais, podendo ainda serem investigados efeitos de longo prazo da seleção (Daetwyler *et al.*, 2013). A imputação de genótipos também é uma técnica que tem sido aplicada em diversos estudos nessa espécie (Araújo *et al.*, 2021; Marina *et al.*, 2021; 2022). Uma das principais vantagens da imputação é minimizar o custo sobre a genotipagem (Butty *et al.*, 2019; Hayes *et al.*, 2012), tornando-se uma tecnologia importante para países em desenvolvimento.

O uso da simulação permite ainda estudar diversos aspectos que envolvem a implantação da seleção genômica, como analisar a escolha da densidade de painéis de SNPs, a quantidade ótima de animais que devem ser genotipados e qual o melhor critério de escolha.

Outro aspecto estudado é o tamanho populacional de treinamento e validação, além de possibilitar a avaliação de estratégias de redução de custo com genotipagem, como é o caso da imputação de genótipos (Mudadu *et al.*, 2016).

Neste trabalho, buscou-se investigar estratégias de implementação da seleção genômica em ovinos por meio do uso da simulação de dados com foco em programas de melhoramento de ovinos em países em desenvolvimento, como o Brasil, especialmente na região Nordeste.

2 REFERENCIAL TEÓRICO

2.1 Aspectos da ovinocultura no Brasil e no mundo

A criação de ovinos é uma das principais fontes de segurança alimentar e renda para famílias no Nordeste brasileiro (Silveira *et al.*, 2021). Nessa região, os sistemas de produção de ovinos são geralmente caracterizados como extensivos, onde os animais se alimentam de pastagem natural e vegetação nativa durante todo o ano (Arandas *et al.*, 2020; Teixeira *et al.*, 2015). Trata-se de uma atividade que tem sido estimulada, especialmente pela demanda de consumo de sua proteína, aumentando a demanda pelo melhoramento de características importantes para o mercado consumidor, como a qualidade de carcaça por exemplo (Santos *et al.*, 2016). Segundo a ARCO – Associação Brasileira de Criadores de Ovinos (2019), 31 raças de ovinos são formalmente registradas, apresentando grande variabilidade genética e morfológica em território nacional.

No mundo, estima-se que a produção de carne ovina em 2022 alcançou 16,4 milhões de toneladas, sendo a Ásia, Oceania, União Europeia e África destaques na categoria tanto de produção como de exportação (FAO, 2022). Atualmente a China e o Oriente Médio são os maiores importadores de carne ovina e a Austrália e Nova Zelândia, os maiores exportadores (FAO, 2022). Como exportadores, esses países são referência no mercado mundial de carne ovina, determinando preços e padrão de qualidade no mercado internacional (Sorio, 2017). A Nova Zelândia se destacou nos últimos trinta anos pelo seu investimento em melhoramento genético em ovinos de corte e vem trabalhando no aprimoramento de tecnologias para seleção e monitoramento desses animais (Brito *et al.*, 2017). Na América Latina, o Brasil se destaca na produção, mas boa parte dessa produção é consumida pela população brasileira (FAO, 2021).

Em 2017 a produção de ovinos no Brasil tinha maior concentração na região Nordeste (64,2%), seguido pela região Sul (23,7%), Centro-Oeste (4,9%), Norte (3,7%) e Sudeste (3,4%) (IBGE, 2017). Em 2018, o número efetivo do rebanho estava em torno de 18 milhões em todo o país (IBGE, 2018), em 2019; subiu para 19,7 milhões, a região Nordeste foi responsável por

68,5% do número total, os estados da Bahia, Rio Grande do Sul e Pernambuco lideraram o *ranking* de criação com 22,8%, 15,5% e 13,7%, respectivamente (IBGE, 2019). Em 2020 houve aumento no efetivo de animais em 3,3%, totalizando 20,6 milhões de cabeças, novamente a região Nordeste manteve sua posição com 70,6% do total de ovinos. O estado da Bahia continuou como principal estado produtor, com 22,8% do efetivo nacional (IBGE, 2020), assim como no último censo, com 21,7% do rebanho nacional. O crescimento em 2022 foi de 4,7% no efetivo nacional resultando num total de 21,5 milhões de ovinos em todo o país (IBGE, 2022).

Com relação a produção de carne ovina, no cenário mundial, a produção aproximada foi de 10 milhões de toneladas em 2019. As perspectivas agrícolas com relação a produção e o comércio de carne ovina em 2023 mostraram que, no mundo, a China é o principal país importador de carne ovina e, na América Latina, o Brasil é o país de maior destaque nessa categoria. As projeções de crescimento na produção de carne ovina para o Brasil entre 2023-2032 diminuíram (0,48) em relação a projeção feita para 2013-2022 (2,69) (FAO, 2023). A Austrália e a Nova Zelândia continuam sendo os países que mais exportam o produto. No Brasil, foram produzidas 136 mil toneladas e a taxa de importação passou de 3.000 toneladas em 2021 (FAO, 2021). Em 2022, a produção estimada subiu para 143 mil toneladas e a taxa de importação com cerca de 4 mil toneladas (FAO, 2022). Espera-se que até 2032 a produção chegue em 150 mil toneladas (FAO, 2023). Contudo, há uma incipiente oferta do produto no mercado nacional por conta da desestruturação da cadeia produtiva, o que implica na necessidade de importação (Lobo, 2019; Viana; Moraes; Dorneles, 2015).

Apesar do consumo nacional de carne ovina ser considerado baixo, em torno de 0,6 kg per capita por ano (EMBRAPA, 2022), ainda assim a produção é insuficiente para atender o mercado nacional. Isso implica em uma necessidade de expansão da cadeia produtiva e espaço para o seu crescimento, em que o uso de estratégias e tecnologias pode fazer com que o país avance e ocupe uma posição mais expressiva no mercado mundial. O desempenho pouco expressivo da ovinocultura no Brasil está intrinsecamente associado a falta de organização e tecnificação desta cadeia produtiva (Sorio; Rasi, 2010; Viana, 2008), assim como a subutilização do melhoramento genético que, de acordo com Lobo (2019), é um dos principais entraves.

No Brasil, sobretudo no Nordeste, há escassez de programas de melhoramento genético de pequenos ruminantes. A criação desses animais em sua maioria é feita por pequenos produtores, onde grande parte desconhece os benefícios do melhoramento genético, fator que

contribui para que a cadeia produtiva ovina no Nordeste seja informal e desorganizada (Lobo, 2019).

Ressalta-se a necessidade de implantação de estratégias que busquem aumentar a produtividade desses animais, assim como oferecer uma melhor qualidade do produto para os consumidores. Nesse cenário, a seleção genômica trata-se de uma importante alternativa que visa atender a essas demandas a partir da identificação de animais geneticamente superiores.

2.2 Contribuições da seleção genômica no melhoramento animal

Por muito tempo a seleção de indivíduos superiores nos programas de melhoramento foi feita com base em informações de fenótipo e pedigree. O uso dessas informações via metodologia de modelos mistos para obtenção do BLUP (*Best Linear Unbiased Prediction*) para a predição de valores genéticos caracteriza o melhoramento clássico. No entanto, essa metodologia vem sendo complementada por apresentar limitações. Alguns exemplos são à capacidade de registrar dados fenotípicos de candidatos a seleção em tempo hábil, características de difícil mensuração, baixa herdabilidade ou que se manifestam tardiamente na vida dos animais.

A seleção genômica ganhou lugar de destaque por oferecer a vantagem do aumento da acurácia dos valores genéticos com a inclusão de informações presentes no DNA. Isso permitiu melhorar a eficiência da seleção e diminuir o intervalo de geração para o ganho genético quando comparada ao melhoramento clássico (Ibtisham *et al.*, 2017; Misztal *et al.*, 2021; VanRaden *et al.*, 2009).

As descobertas de Gregor Mendel sobre o padrão de herança de características mensuráveis foram essenciais para o sucesso da seleção baseada em informações genômicas. Somado aos avanços da ciência na área da biologia molecular, da estatística e da bioinformática, possibilitaram uma melhor predição da aptidão genética dos indivíduos (Tsairidou *et al.*, 2020).

A implementação da tecnologia de genotipagem junto ao uso de polimorfismos de nucleotídeos únicos (*single nucleotide polymorphism*- SNPs) para cobrir todo o genoma e rastrear características de interesse resultou em inúmeras vantagens para o melhoramento animal. Possibilitou a detecção e o mapeamento de *loci* de características quantitativas (QTLs) e a inclusão dessas informações para predição de EBVs (Sarmiento; Sena, 2017).

A genômica também provocou uma redução de custo em testes de progênie, com a simplificação do processo de seleção e redução no intervalo de gerações (Boichard *et al.*, 2016). Com a implantação dessa metodologia, maiores números de reprodutores podem ser selecionados e comercializados (Henryon *et al.*, 2014). A genotipagem de rebanhos permitiu a

predição mais acurada inclusive para animais jovens que ainda não dispõem de informações fenotípicas, otimizando a seleção ou o descarte de animais (Meuwissen *et al.*, 2016). Essa ferramenta também tem sido importante para seleção de características reprodutivas, que normalmente apresentam baixas acurácias para os valores genéticos preditos por técnicas tradicionais e são de difícil mensuração (Lillehammer *et al.*, 2020).

Com essa nova metodologia houve um acréscimo significativo na produtividade animal pelo aumento do ganho genético em comparação com o melhoramento tradicional em diversas espécies. O ganho anual com o uso da seleção genômica em ovinos leiteiros chegou a 51,7% e 17,9% para ovinos de corte (Shumbusho *et al.*, 2013). A Nova Zelândia conseguiu resultados consideráveis com a seleção genômica, chegando a um aumento na ordem de 83% em kg de cordeiro produzido por ovelha e 28% (+4,1 kg) de aumento no peso de carcaça (Brito *et al.*, 2017). Outros estudos utilizando dados genômicos com a mesma espécie também mostraram resultados promissores, em que o ganho médio da acurácia populacional foi de 0,20 (Daetwyler *et al.*, 2012). Baloché *et al.* (2014) obtiveram um ganho de em acurácia entre 0,10 e 0,20 em característica de produção de leite.

Em outro estudo com ovinos na Noruega, Lillehammer *et al.* (2020) mostraram que a implementação da seleção genômica pode aumentar significativamente o ganho genético para características reprodutivas em 65-77% por ano. Sena *et al.* (2020) ao avaliarem características de carcaça, obtiveram valores de GEBVs maiores em todos os cenários de inclusão de informação genômica, em que os ganhos na acurácia variaram de 17% a 68%. Da mesma forma, Torres *et al.* (2020), ao estudarem sobre infecção intestinal por nematoide e tamanho corporal nessa mesma espécie, concluíram que a inclusão de informações genômicas para as características em estudo foi benéfica, pois os ganhos na acurácia dos valores genéticos estimados variaram de 2,2% a 75%.

A utilização da genômica também contribuiu no âmbito do manejo da taxa de endogamia e diversidade genética. É notório que o número restrito de machos e o acasalamento entre indivíduos aparentados numa população ao longo de gerações resulta no aumento da endogamia e consequente diminuição da diversidade genética (Deniskova *et al.*, 2019). O método genômico possibilita reduzir a endogamia e aumentar a diversidade genética quando é mais assertiva no direcionamento de acasalamentos dentro de uma população, diminuindo a frequência de alelos recessivos em homozigose (Curik *et al.* 2014). Utilizar o método genômico para garantir a diversidade genética é especialmente importante quando os registros de pedigree não são confiáveis ou estão indisponíveis (Al-Mamun *et al.*, 2015; Li *et al.*, 2011). O desenvolvimento de chips comerciais tornou a genotipagem de SNPs a técnica mais útil e

robusta para inferir sobre níveis de endogamia (Deniskova *et al.*, 2019). Atualmente, além dos cálculos sobre a heterozigosidade observada e esperada e os índices estatísticos F de *Wright* (F_{is}), que são métodos clássicos, o uso do tamanho efetivo da população (N_e) com base no desequilíbrio de ligação (LD) (Corbin *et al.*, 2010; Kijas *et al.*, 2012) e corridas de homozigose (ROH) são metodologias relevantes para estimar os níveis de endogamia (Ojeda-Marín *et al.*, 2024; Yin *et al.*, 2023).

Contudo, em se tratando da implantação de programas de melhoramento genético para ovinos no Brasil, Lobo (2019), ao falar sobre as tentativas de estruturação para pequenos ruminantes, refletiu sobre algumas questões importantes e os motivos pelos quais essas tentativas foram malsucedidas. O autor avaliou que a distribuição de genética exótica, às fontes de financiamento e às prioridades de aplicação dos recursos são questões a serem avaliadas com maior atenção. Segundo ele, os custos envolvidos em um programa de melhoramento, devem ser viáveis e promoverem retornos na melhoria de características animais e consequentemente financeiro.

Dessa maneira, é inegável que a genômica promoveu melhoria acentuada no processo de seleção viabilizando estimativas mais acuradas de características de interesse econômico em diversas espécies. Nesse sentido, os programas de melhoramento animal têm buscado cada vez mais implementar a seleção genômica como meio de maximizar os ganhos genéticos e o retorno econômico. No entanto, cabe ressaltar que a implementação dessa metodologia esbarra em inúmeros obstáculos, sendo ainda um desafio para produtores em algumas espécies, como nos ovinos.

2.3 Desafios para a implementação da seleção genômica

Apesar das vantagens trazidas pelo uso da genômica, a implementação dessa técnica nos programas de melhoramento animal não é uma tarefa fácil. Misztal *et al.* (2013) resumiram os desafios da implementação da seleção genômica na pecuária enfatizando que o sucesso desse processo em escala comercial depende de atenção para detalhes, dependendo também do conhecimento das questões de seleção intrínsecos para cada população.

Na seleção genômica o valor genético total dos candidatos à seleção é obtido com base na estimativa dos efeitos de SNPs. Sabendo disso, a primeira etapa é a formação de uma população de referência, que têm fenótipo e genótipos conhecidos. Essa população precisa ser representativa o suficiente e relevante para a população selecionada (Andonov *et al.*, 2016; Calus *et al.*, 2013). A formação de uma população de validação também é necessária para testar se a população de referência é adequada. Na validação, apesar de conhecidos, os fenótipos são

mascarados a fim de testar a capacidade preditiva do modelo, sendo utilizados apenas o genótipo e o pedigree. Esses requisitos têm sido um dos principais desafios para ovinocultura, que diferentemente de outras espécies, como os bovinos, os tamanhos dos rebanhos são pequenos, existe carência de registros fenotípicos confiáveis e principalmente limitações em investimentos para genotipagem (Meuwissen; Hayes; Goddard, 2016). Isso ocorre por que normalmente esses animais são criados em sua maioria por pequenos produtores, que tem conhecimento limitado em relação aos benefícios do melhoramento genético (Lobo, 2012).

Apesar da diminuição dos custos em genotipagem ao longo dos anos, ainda é um dos principais desafios em espécies que o valor agregado do indivíduo é relativamente baixo e em países em desenvolvimento. Responder a questionamentos sobre quantos e quais animais precisam ser genotipados é relevante nesse processo (Wolc *et al.*, 2016), a fim de reduzir gastos extras. Uma alternativa interessante baseia-se na utilização de painéis de baixa densidade para posterior uso de técnicas de imputação (Marina *et al.*, 2021; Vanraden *et al.*, 2015). A Figura 1 traz uma ilustração do processo de imputação de genótipos para um haplótipo em estudo.

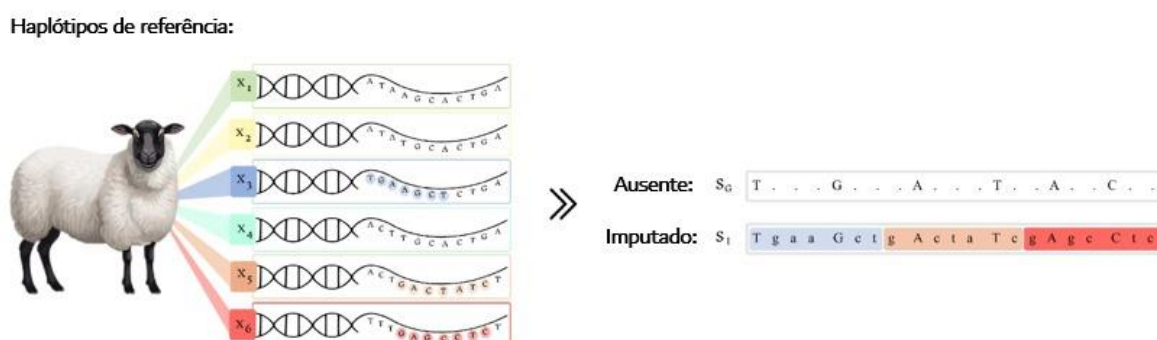


Figura 1. Processo de imputação para um haplótipo de estudo (SG) genotipado em 6 marcadores usando um painel de referência de haplótipos sequenciados em 21 marcadores. Adaptada de: Das; Abecasis; Browning, 2018.

Dentro da seleção genômica, a imputação de genótipos é uma ferramenta robusta para minimizar os custos de genotipagem. Trata-se de uma técnica que se baseia no princípio de que marcadores genéticos são herdados em blocos entre indivíduos próximos. Esses haplótipos compartilhados evidenciam relações de parentesco (Hayes *et al.*, 2012; Marina *et al.*, 2021; Ventura *et al.*, 2014; 2016), portanto, refere-se a imputação de genótipos ausentes que não foram genotipados diretamente dos indivíduos amostrados.

Para o uso dessa técnica, alguns animais são genotipados com chips mais densos, posteriormente outros animais são genotipados com chips de densidade mais baixa. A partir desses, é possível imputar (prever) genótipos ausentes no painel de densidade mais baixa para a mais alta antes de calcular o GEBV (Carvalho *et al.*, 2014). Avaliar a acurácia da imputação

do número ótimo de SNPs é crucial para definir uma estratégia de genotipagem apropriada para minimizar os custos de genotipagem (Yoshida *et al.*, 2018) na espécie ovina brasileira.

Em bovinos Nelore foi possível aumentar a informação genômica pela imputação de grandes painéis de SNPs (>700.000) a partir de painéis de 15.000 SNPs (Carvalho *et al.*, 2014) de forma acurada. Nessa mesma raça, Sharma *et al.* (2018) encontraram uma concordância genotípica de 0,40 e correlação de 0,31 entre microssatélites reais e imputados. Em ovinos essa estratégia foi utilizada para imputar marcadores SNPs a partir de microssatélites como meio de verificação de parentesco na raça Assaf (Marina *et al.*, 2021). A imputação de genótipo foi um elemento chave no processo de implantação da seleção genômica em ovinos na Nova Zelândia (Ventura *et al.*, 2016), que também já foi implementada na Austrália e em ovinos leiteiros na França (Rupp *et al.*, 2016).

Somado a ferramenta de imputação, a simulação de dados também tem sido importante para o processo de implantação da seleção genômica, pois permite gerar dados para pesquisa e realização de testes rápidos e replicáveis. Através desse método é possível gerar regiões genômicas conhecidas e avaliar sob diferentes metodologias, como comparar procedimentos múltiplos e de etapa única para prever o GEBV (Vitezica *et al.*, 2011). Além disso, é possível incluir níveis diferentes de informações para observar a acurácia preditiva (Wang *et al.*, 2012) e estudar diferentes cenários numa mesma população (Van den berg *et al.*, 2013). Esse é um método que tem sido utilizado nos programas de melhoramento animal a partir de parâmetros baseados numa população alvo para simular cenários reais (Piccoli *et al.* 2018).

Diversos trabalhos fizeram uso dessa ferramenta para comparar os métodos GBLUP em um e dois passos em bovinos de corte (Piccoli *et al.*, 2018) e para comparar a seleção genômica com a clássica em gado leiteiro (Gaspa *et al.*, 2015). Em ovinos, foi utilizada para avaliar a eficiência do método ssBLUP sob diferentes cenários (Pértille *et al.*, 2016) e também para comparar métodos de seleção e estudar características reprodutivas (Latifi *et al.*, 2020). Dessa forma, é um método pertinente por permitir gerar a arquitetura genética de características de interesse, o número de marcadores usados para análise e fontes de variabilidade, como a deriva (Daetwyler *et al.*, 2013).

Essa metodologia se mostra relevante para testar diferentes estratégias para a implementação da seleção genômica em ovinos no Brasil, por ser de baixo custo e aplicável para diferentes características e cenários de estudo. Meuwissen *et al.* (2001) propuseram a seleção genômica através de um estudo de simulação onde mostraram que o valor genético poderia ser previsto com uma acurácia de 0,85 utilizando dados de marcadores. Assim, apesar de todos os desafios mencionados, a estratégia de simulação de dados oferece expectativas

animadoras nessa perspectiva, podendo resultar em importantes resultados para ovinocultura brasileira.

3 REFERÊNCIAS

AL-MAMUN, H. A. *et al.* Genome-wide linkage disequilibrium and genetic diversity in five populations of Australian domestic sheep. **Genet Sel Evol**, Armidale, v. 47, n. 90, p. 1-14, Nov. 2015. DOI <https://doi.org/10.1186/s12711-015-0169-6>. Disponível em: <https://doi.org/10.1186/s12711-015-0169-6>. Acesso em: 4 Jun. 2024.

ANDONOV, S. *et al.* Accuracy of breeding values in small genotyped populations using different sources of external information: A simulation study. **J Dairy Sci.**, United States, v. 100, n. 1, p. 395-401, Jan. 2017. DOI: [10.3168/jds.2016-11335](https://doi.org/10.3168/jds.2016-11335). Disponível em: <https://www.sciencedirect.com/science/article/pii/S0022030216307299>. Acesso em: 12 Fev. 2024.

ARAÚJO, A. C. *et al.* A Comprehensive Comparison of Haplotype-Based Single-Step Genomic Predictions in Livestock Populations With Different Genetic Diversity Levels: A Simulation Study. **Front. Genet.**, v. 12, n. 729867, Oct. 2021. DOI <https://doi.org/10.3389/fgene.2021.729867>. Disponível em: <https://www.frontiersin.org/journals/genetics/articles/10.3389/fgene.2021.729867/full#B34>. Acesso em: 22 Jun. 2024.

ARCO - ASSOCIAÇÃO BRASILEIRA DE CRIADORES DE OVINOS. Bagé: ARCO, 2019. Disponível em: <https://www.arcoovinos.com.br/PadraoRacial>. Acesso em: 8 mar. 2019.

BALOCHE, G. *et al.* Assessment of accuracy of genomic prediction for French Lacaune dairy sheep. **J Dairy Sci.**, Castanet-Tolosan, v. 97, n. 2, p. 1107-1116, Feb. 2014. DOI <https://doi.org/10.3168/jds.2013-7135>. Disponível em: <https://doi.org/10.3168/jds.2013-7135>. Acesso em: 4 Jun. 2024.

BOICHARD, D. *et al.* Genomic selection in domestic animals: principles, applications and perspectives. **C R Biol**, France, v. 339, n. 7-8, p. 274-277, Aug. 2016. DOI [10.1016/j.crv.2016.04.007](https://doi.org/10.1016/j.crv.2016.04.007). Disponível em: <https://pubmed.ncbi.nlm.nih.gov/27185591/>. Acesso em: 4 Jun. 2024.

BRITO, L. F. *et al.* Genetic parameters for various growth, carcass and meat quality traits in a New Zealand sheep population. **Small Rumin Res**, Amsterdam, v. 154, p. 81-91, July. 2017. DOI [10.1016/j.smallrumres.2017.07.011](https://doi.org/10.1016/j.smallrumres.2017.07.011). Disponível em: <https://doi.org/10.1016/j.smallrumres.2017.07.011>.

BUTTY, A. M. *et al.* Optimizing selection of the reference population for genotype imputation from array to sequence variants. **Front Genet**, [S.l.], v. 10, p. 510, May. 2019. DOI [10.3389/fgene.2019.00510](https://doi.org/10.3389/fgene.2019.00510). Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6554347/>. Acesso em: 4 Jun. 2024.

CARVALHEIRO, R. *et al.* Accuracy of genotype imputation in Nelore cattle. **Genet Sel Evol**, France, v. 46, n. 1, p. 69, Oct. 2014. DOI <https://doi.org/10.1186/s12711-014-0069->.

Disponível em: <https://gsejournal.biomedcentral.com/articles/10.1186/s12711-014-0069-1>. Acesso em: 4 Jun. 2024.

CHUD, T. C. S. *et al.* Strategies for genotype imputation in composite beef cattle. **BMC Genet.**, England, v. 16, n. 1, p. 1-10, Aug. 2015. DOI [10.1186/s12863-015-0251-7](https://doi.org/10.1186/s12863-015-0251-7). Disponível em: <https://pubmed.ncbi.nlm.nih.gov/26250698/>. Acesso em: 12 Mai. 2024.

CALUS, M. P. L. *et al.* Combining cow and bull reference populations to increase accuracy of genomic prediction and genome-wide association studies. **J Dairy Sci.**, United States, v. 96, n. 10, p. 6703–6715, Oct. 2013. DOI [10.3168/jds.2012-6013](https://doi.org/10.3168/jds.2012-6013). Disponível em: <https://www.sciencedirect.com/science/article/pii/S0022030213005237>. Acesso em: 3 Mar. 2022.

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. **J. Anim. Sci.**, [S.l.], v. 89, n. 1, p. 23-28, Jan. 2011. DOI <https://doi.org/10.2527/jas.2010-3071>. Disponível em: <https://academic.oup.com/jas/article-abstract/89/1/23/4764252?redirectedFrom=fulltext>. Acesso em: 4 Jun. 2024.

CORBIN, L. J. *et al.* Estimation of historical effective population size using linkage disequilibria with marker data. **J Anim Breed Genet**, [S.l.], v. 129, n. 4, p. 257-270, May. 2012. DOI [10.1111/j.1439-0388.2012.01003.x](https://doi.org/10.1111/j.1439-0388.2012.01003.x). Disponível em: <https://pubmed.ncbi.nlm.nih.gov/22775258/>. Acesso em: 4 Jun. 2024.

CURIK, I.; FERENČAKOVIĆ, M.; SÖLKNER, J. Inbreeding and runs of homozygosity: A possible solution to an old problem. **Livest Sci**, [S.l.], v. 166, p. 26-34, Aug. 2014. DOI <https://doi.org/10.1016/j.livsci.2014.05.034>. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S1871141314003060>. Acesso em: 4 Jun 2024.

DAETWYLER, H. D. *et al.* Accuracy of pedigree and genomic predictions of carcass and novel meat quality traits in multi-breed sheep data assessed by cross-validation. **Genet Sel Evol**, France, v. 44, n. 33, p. 1-11, Nov. 2012. DOI <https://doi.org/10.1186/1297-9686-44-33>. Disponível em: <https://gsejournal.biomedcentral.com/articles/10.1186/1297-9686-44-33>. Acesso em: 4 Jun 2024.

DAETWYLER, H. D. *et al.* Genomic prediction in animals and plants: simulation of data, validation, reporting, and benchmarking. **Genetics**, [S.l.], v. 193, n. 2, p. 347–365, Feb. 2013. DOI [10.1534/genetics.112.147983](https://doi.org/10.1534/genetics.112.147983). Disponível em: <https://pubmed.ncbi.nlm.nih.gov/23222650/>. Acesso em: 4 Jun. 2024.

DENISKOVA, T. *et al.* Population structure and genetic diversity of sheep breeds in the Kyrgyzstan. **Front Genet**, [S.l.], v. 10, p. 1311, Dec. 2019. DOI [10.3389/fgene.2019.01311](https://doi.org/10.3389/fgene.2019.01311). Disponível em: <https://www.frontiersin.org/journals/genetics/articles/10.3389/fgene.2019.01311/full>. Acesso em: 4 Jun. 2024.

DAS, S.; ABECASIS, G. R.; BROWNING, B. L. Genotype Imputation from Large Reference Panels. **Annu Rev Genomics Hum Genet.**, United States, v. 19, p. 73-96, Aug. 2018. DOI

10.1146/annurev-genom-083117-021602. Disponível em:
<https://pubmed.ncbi.nlm.nih.gov/29799802/>. Acesso em: 26 Jun. 2024.

FAO – Food and Agriculture Organization of the United Nations, 2021. **Meat market review**: Overview of global meat market developments in 2020. Rome: FAO, 2021. 15p. Disponível em: <https://www.fao.org/economic/est/est-commodities/meat/meat-and-meat-products-update/en/>. Acesso em: 23 Jun. 2024.

FAO – Food and Agriculture Organization of the United Nations, 2022. **Meat Market Review**: Emerging trends and outlook. 2022. Rome: FAO, 2022. 24p. Disponível em: <https://www.fao.org/markets-and-trade/publications/detail/es/c/1620239/>. Acesso em: 02 out. 2023.

FAOSTAT: **FAO's corporate database** [online]. Rome: FAO, 2021. Disponível em: <https://www.fao.org/faostat/en/#home>. Acesso em: 02 out. 2023.

GASPA, G. *et al.* Assessment of genomic selection for introgression of polledness into Holstein Friesian cattle by simulation. **Livest Sci**, [S.l.], v. 179, p. 86-95, Sept. 2015. DOI <https://doi.org/10.1016/j.livsci.2015.05.020>. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S1871141315002498>. Acesso em: 4 Jun. 2024.

GUINAN, F. L. *et al.* Changes in genetic trends in US dairy cattle since the implementation of genomic selection. **J Dairy Sci.**, United States, v. 106, n. 2, p. 1110-1129, Feb. 2023. DOI 10.3168/jds.2022-22205. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/36494224/>. Acesso em: 4 Jun. 2024.

HAYES, B. J. *et al.* Accuracy of genotype imputation in sheep breeds. **Anim. Genet.**, [S.l.], v. 43, n. 1, p. 72-80, Feb. 2012. DOI <https://doi.org/10.1111/j.1365-2052.2011.02208.x>. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1111/j.1365-2052.2011.02208.x>. Acesso em: 4 Jun 2024.

HENRYON, M.; BERG, P.; SORENSEN, A. C. Animal-breeding schemes using genomic information need breeding plans designed to maximise long-term genetic gains. **Livest Sci**, [S.l.], v. 166, p. 38-47, Aug. 2014. DOI <https://doi.org/10.1016/j.livsci.2014.06.016>. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1871141314003242?via%3Dihub>. Acesso em: 4 Jun. 2024.

HICKEY, J. M. *et al.* Factors affecting the accuracy of genotype imputation in populations from several maize breeding programs. **Crop Science**, [S.l.], v.52, n.2, p. 654-663, Mar. 2012. DOI <https://doi.org/10.2135/cropsci2011.07.0358>. Disponível em: <https://acess.onlinelibrary.wiley.com/doi/10.2135/cropsci2011.07.0358>. Acesso em: 4 Jun. 2024.

IBGE – Instituto Brasileiro de Geografia e Estatística – Sistema IBGE de recuperação automática – SIDRA, 2018. **Pesquisa Pecuária Municipal**: Total efetivo de ovinos de todos os estados. Distrito Federal: IBGE, 2018. Disponível em: <https://sidra.ibge.gov.br/tabela/3939>. Acesso em: janeiro de 2021.

IBGE – Instituto Brasileiro de Geografia e Estatística. **Pesquisa da Produção Pecuária Municipal 2019**, PMM 2019. Rio de Janeiro: IBGE, 2019, v. 47. 12p. Disponível em: <https://www.embrapa.br/cim-centro-de-inteligencia-e-mercado-em-aquicultura/pesquisa-pecuaria-municipal-ibge>. Acesso em: 09 Fev. 2023.

IBGE – Instituto Brasileiro de Geografia e Estatística. **Pesquisa da Produção Pecuária Municipal 2020**, PMM 2020. Rio de Janeiro: IBGE, 2020, v.48. 12p. Disponível em: <https://www.embrapa.br/cim-centro-de-inteligencia-e-mercado-em-aquicultura/pesquisa-pecuaria-municipal-ibge>. Acesso em: 09 Fev. 2023.

IBGE – Instituto Brasileiro de Geografia e Estatística - Estatística – Sistema IBGE de recuperação automática – SIDRA, 2017. **Pesquisa Pecuária Municipal: Efetivo dos rebanhos, 2017**, PMM 2017. Rio de Janeiro: IBGE, 2017. Disponível em: <http://www.sidra.ibge.gov.br/bda/tabela/listabl.asp?c=3939&z=t&o=24>. Acesso em: 10 Jan. 2021.

IBGE – Instituto Brasileiro de Geografia e Estatística. **Pesquisa da Produção Pecuária Municipal 2022**, PMM 2022. Rio de Janeiro: IBGE, 2022, v. 50. 12p. Disponível em: <https://biblioteca.ibge.gov.br/index.php/biblioteca-catalogo?view=detalhes&id=784>. Acesso em: 09 Fev. 2023.

IBTISHAM, F. *et al.* Genomic selection and its application in animal breeding. **The Thai J Vet Med**, [S.l.], v. 47, n. 3, p. 301-310, Sept. 2017. DOI [10.56808/2985-1130.2838](https://doi.org/10.56808/2985-1130.2838). Disponível em: <https://digital.car.chula.ac.th/tjvm/vol47/iss3/1/>. Acesso em 6 Jun. 2024.

KAI-LI, Z. *et al.* A comprehensive evaluation of factors affecting the accuracy of pig genotype imputation using a single or multi-breed reference population. **Journal of Integrative Agriculture**, [S.l.], v. 21, n. 2, p. 486-495, Feb. 2022. DOI [https://doi.org/10.1016/S2095-3119\(21\)63695-X](https://doi.org/10.1016/S2095-3119(21)63695-X). Disponível em: <https://www.sciencedirect.com/science/article/pii/S209531192163695X>. Acesso em: 10 Jun. 2024.

KIJAS, J. W. *et al.* Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. **PLoS Biol**, California, v. 10, n. 2, e1001258, Feb. 2012. DOI <https://doi.org/10.1371/journal.pbio.1001258>. Disponível em: <https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1001258>. Acesso em: 6 Jun. 2024.

LATIFI, M. *et al.* Comparison of different selection methods for improving litter size in sheep using computer simulation. **Span J Agric Res**, Spain, v. 18, n. 1, p. 403, e0403, Apr. 2020. DOI <https://doi.org/10.5424/sjar/2020181-15459>. Disponível em: <https://sjar.revistas.csic.es/index.php/sjar/article/view/15459>. Acesso em: 6 Jun. 2024.

LEGARRA, A.; REVERTER, A. Semi-parametric estimates of population accuracy and bias of predictions of breeding values and future phenotypes using the LR method. **Genet Sel Evol**, France, v. 50, p. 1-18, Nov. 2018. DOI <https://doi.org/10.1186/s12711-018-0426-6>. Disponível em: <https://gsejournal.biomedcentral.com/articles/10.1186/s12711-018-0426-6>. Acesso em: 6 Jun. 2024.

LI, Y. *et al.* Genotype imputation. **Annual review of genomics and human genetics.**, [S.l.] v. 10, p. 387-406, Sept. 2009. DOI <https://doi.org/10.1146/annurev.genom.9.081307.164242>. Disponível em: <https://www.annualreviews.org/content/journals/10.1146/annurev.genom.9.081307.164242>. Acesso em: 10 Jan. 2022.

LI, M-H. *et al.* A comparison of approaches to estimate the inbreeding coefficient and pairwise relatedness using genomic and pedigree data in a sheep population. **PloS one**, California, v. 6, n. 11, p. e26256, Nov. 2011. DOI [10.1371/journal.pone.0026256](https://doi.org/10.1371/journal.pone.0026256). Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3220595/>. Acesso em: 6 Jun. 2024.

LILLEHAMMER, M.; MEUWISSEN, T. H. E.; SONESSON, A. K. Genomic selection for maternal traits in pigs. **J Anim Sci**, Oxford, v. 89, n. 12, p. 3908-3916, Dec. 2011. DOI [10.2527/jas.2011-4044](https://doi.org/10.2527/jas.2011-4044). Disponível em: <https://academic.oup.com/jas/article-abstract/89/12/3908/4772003?redirectedFrom=fulltext&login=false>. Acesso em: 6 Jun. 2024.

LILLEHAMMER, M. *et al.* Genomic selection strategies to improve maternal traits in Norwegian White Sheep. **J Anim Breed Genet**, [S.l.], v. 137, n. 4, p. 384-394, Mar. 2020. DOI <https://doi.org/10.1111/jbg.12475>. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1111/jbg.12475>. Acesso em: 6 Jun. 2024.

LOBO, R. N. B.; LOBO, A. M. B. O. Evolução do melhoramento de caprinos e ovinos no Brasil. In: **Embrapa Caprinos e Ovinos-Artigo em anais de congresso (ALICE)**. In: ENCONTRO DE GENÉTICA DO NORDESTE, 19.; SIMPÓSIO DE GENÉTICA HUMANA E MÉDICA DO NORDESTE, 1.; GENÉTICA NA PRAÇA, 2012, Petrolina, Juazeiro. A genética, a natureza e o ser humano: mudando mentalidades e transformando vidas. Petrolina: Embrapa Semiárido: UNIVASF: SBG, 2012. 11 f. 1 CD-ROM., 2012. Disponível em: <https://www.embrapa.br/busca-de-publicacoes/-/publicacao/939270/evolucao-do-melhoramento-de-caprinos-e-ovinos-no-brasil>. Acesso em: 7 Jun. 2024.

LOBO, R. N. B. Opportunities for investment into small ruminant breeding programmes in Brazil. **J Anim Breed Genet**, [S.l.], v. 136, n. 5, p. 313-318, May. 2019. DOI <https://doi.org/10.1111/jbg.12396>. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1111/jbg.12396>. Acesso em: 6 Jun. 2024.

LOURENCO, D. *et al.* Single-step Genomic Evolutions from Theory to Practice: Using SNP Chips and Sequence Data in BLUPF90. **Genes**, Basel: MDPI, v. 11, n. 7, Jul. 2020. DOI <https://doi.org/10.3390/genes11070790>. Disponível em: <https://www.mdpi.com/2073-4425/11/7/790>. Acesso em: 20 Jun. 2024.

LUND, M. S. *et al.* A common reference population from four European Holstein populations increases reliability of genomic predictions. **Genet Sel Evol**, France, v. 43, n. 1, p. 1-8, Dec. 2011. DOI <https://doi.org/10.1186/1297-9686-43-43>. Disponível em: <https://gsejournal.biomedcentral.com/articles/10.1186/1297-9686-43-43>. Acesso em: 6 Jun. 2024.

MARINA, H. *et al.* Accuracy of Imputation of Microsatellite Markers from a 50K SNP Chip in Spanish Assaf Sheep. **Animals**, Basel, v. 11, n. 1, p. 86, Jan. 2021. DOI [10.3390/ani11010086](https://doi.org/10.3390/ani11010086). Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7824810/>. Acesso em: 6 Jun. 2024.

MARTIN, R. *et al.* Optimization Strategies to Adapt Sheep Breeding Programs to Pasture-Based Production Environments: A Simulation Study. **Animals**, Basel, v. 13, n. 22, p. 3476, Nov. 2023. DOI <https://doi.org/10.3390/ani13223476>. Disponível em: <https://www.mdpi.com/2076-2615/13/22/3476>. Acesso em: 6 Jun. 2024.

MEUWISSEN, T.; HAYES, B.J.; GODDARD, M. E. Prediction of total genetic value using genome-wide dense marker maps. **Genetics**, United States, v.157, p.1819-1829, Apr. 2001. DOI <https://doi.org/10.1093/genetics/157.4.1819>. Disponível em: <https://academic.oup.com/genetics/article/157/4/1819/6048353?login=false>. Acesso em: 7 Jun. 2024.

MEUWISSEN, T.; HAYES, B. J.; GODDARD, M. E. Accelerating improvement of livestock with genomic selection. **Annu Rev Anim Biosci**, United States, v. 1, p. 221-237, Jan. 2013. DOI 10.1146/annurev-animal-031412-103705. Disponível em: <https://www.annualreviews.org/content/journals/10.1146/annurev-animal-031412-103705>. Acesso em: 7 Jun. 2024.

MEUWISSEN, T.; HAYES, B. J.; GODDARD, M. E. Genomic selection: A paradigm shift in animal breeding. **Animal Front**, [S.l.], v.6, n.1, p. 6-14, Jan. 2016. DOI <https://doi.org/10.2527/af.2016-0002>. Disponível em: <https://academic.oup.com/af/article/6/1/6/4638797>. Acesso em: 6 Jun. 2024.

MISZTAL, I. *et al.* Methods to approximate reliabilities in single-step genomic evaluation. **J Dairy Sci.**, United States, v. 96, n. 1, p. 647-654, Jan. 2013. DOI 10.3168/jds.2012-5656. Disponível em: [https://www.journalofdairyscience.org/article/S0022-0302\(12\)00802-8/fulltext](https://www.journalofdairyscience.org/article/S0022-0302(12)00802-8/fulltext). Acesso em: 6 Jun. 2024.

MISZTAL, I. *et al.* **Manual for BLUPF90 family of programs**. Athens, June. 2022. Disponível em: <https://nce.ads.uga.edu/wiki/doku.php?id=documentation>. Acesso em: 26 Jan. 2021.

MORRIS, S. T. Overview of sheep production systems. In: Ferguson, D. M.; Lee, C.; Fisher, A. (Ed.). **Woodhead Publishing Series in Food Science, Technology and Nutrition, Advances in Sheep Welfare**. Woodhead Publishing, 2017, p. 19-35. DOI <https://doi.org/10.1016/B978-0-08-100718-1.00002-9>. Disponível em: <https://www.sciencedirect.com/science/article/pii/B9780081007181000029>. Acesso em: 20 Jun. 2024.

MUDADU, M. A. *et al.* **Imputação de genótipos em bovinos: um guia passo a passo**. Campinas: Embrapa Informática Agropecuária, v. 143, 2016. 31 p. ISSN 1677-9274. Disponível em: <https://www.embrapa.br/busca-de-publicacoes/-/publicacao/1064163/imputacao-de-genotipos-em-bovinos-um-guia-passo-a-passo>. Acesso em: 26 Jun. 2024.

NALÉRIO, E. S.; GIONGO, C. **Oportunidades de agregação de valor à carne ovina: da carcaça aos derivados cárneos**. Brasília, DF: Embrapa, 2022. 44p. Disponível em: <https://www.embrapa.br/busca-de-publicacoes/-/publicacao/1141830/oportunidades-de-agregacao-de-valor-a-carne-ovina-da-carcaca-aos-derivados-carneos>. Acesso em: 7 Jun. 2024.

NAVAJAS, E. A. Animal breeding and genetics: DNA Markers and Marker-Assisted Selection in the Genomic Era. In: Dikeman, M., Devine, C. (Ed.). **Encyclopedia of Meat Sciences (Second Edition)**. New York: Academic Press, 2014. v.1, n.2, p.12-18. Disponível em: <<https://doi.org/10.1016/B978-0-12-384731-7.00002-7>>. ISBN 9780123847348. Acesso em: 6 Jun. 2024.

NEVES, H. H. *et al.* Accuracy of genomic predictions in *Bos indicus* (Nellore) cattle. **Genet Sel Evol**, France, v. 46, n. 1, p. 1-13, Feb. 2014. DOI <https://doi.org/10.1186/1297-9686-46-17>. Disponível em: <https://gsejournal.biomedcentral.com/articles/10.1186/1297-9686-46-17>. Acesso em: 6 Jun. 2024.

OJEDA-MARÍN, C. *et al.* Differential patterns in runs of homozygosity in two mice lines under divergent selection for environmental variability for birth weight. **J Anim Breed Genet**, [S.l.], v. 141, n. 2, p. 193-206, Nov. 2024. DOI 10.1111/jbg.12835. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1111/jbg.12835>. Acesso em: 6 Jun. 2024.

PÉRTILE, S. F. N. *et al.* Seleção e associação genômica ampla para o melhoramento genético animal com uso do método ssGBLUP. **Pesq. Agropec. Bras.**, Brasília, v. 51, n. 10, p. 1729-1736, Oct. 2016. DOI <https://doi.org/10.1590/S0100-204X2016001000004>. Disponível em: <https://www.scielo.br/j/pab/a/tSrmQnJ44P79PRz9cBDHjtP/>. Acesso em: 6 Jun. 2024.

PICCOLI, M. L. *et al.* Accuracy of genome-wide imputation in Braford and Hereford beef cattle. **BMC Genet.**, London, v. 15, n. 157, p. 1-15, Dec. 2014. DOI <https://doi.org/10.1186/s12863-014-0157-9>. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/25543517/>. Acesso em: 5 Ago. 2022.

PICCOLI, M. L. *et al.* A comprehensive comparison between single-and two-step GBLUP methods in a simulated beef cattle population. **Can J Anim Sci**, Canada, v. 98, n. 3, p. 565-575, Sept. 2018. DOI <https://doi.org/10.1139/cjas-2017-0176>. Disponível em: <https://cdnsiencepub.com/doi/10.1139/cjas-2017-0176>. Acesso em: 6 Jun. 2024.

PIMENTEL, E. C. G.; KÖNIG, S. Genomic selection for the improvement of meat quality in beef. **J Anim Sci**, United States, v. 90, n. 10, p. 3418-3426, Oct. 2012. DOI <https://doi.org/10.2527/jas.2011-5005>. Disponível em: <https://academic.oup.com/jas/articleabstract/90/10/3418/4717860?redirectedFrom=fulltext&login=false>. Acesso em: 6 Jun. 2024.

PRYCE, J. E.; DAETWYLER, H. D. Designing dairy cattle breeding schemes under genomic selection: a review of international research. **Anim Prod Sci**, Australia, v. 52, n. 3, p. 107-114, 2012. DOI <https://doi.org/10.1071/AN11098>. Disponível em: <https://www.publish.csiro.au/AN/fulltext/AN11098>. Acesso em: 6 Jun. 2024.

RESHMA, R. S.; DAS, D. N. Molecular markers and its application in animal breeding. In: MONDAL, S.; SINGH, R. L. (Ed.). **Advances in Animal Genomics**. Academic Press, 2021. cap. 9, p. 123-140. DOI [10.1016/B978-0-12-820595-2.00009-6](https://doi.org/10.1016/B978-0-12-820595-2.00009-6). Disponível em: <<https://doi.org/10.1016/B978-0-12-820595-2.00009-6>>. Acesso em: 5 Jun. 2024.

RUPP, R. *et al.* Genomic application in sheep and goat breeding. **Animal front.**, [S.l.], v. 6, n. 1, p. 39-44, Jan. 2016. DOI <https://doi.org/10.2527/af.2016-0006>. Disponível em: <https://academic.oup.com/af/article/6/1/39/4638792>. Acesso em: 6 Jun. 2024.

SANTOS, N. P. S. *et al.* Estrutura de covariância para características de carcaça e tamanho corporal com medidas repetidas em ovinos de diferentes grupos genéticos. **Rev. Bras. Saúde e Prod. Anim.**, Salvador, v. 17, n. 4, p. 652-665, Dez. 2016. DOI <https://doi.org/10.1590/S1519-99402016000400009>. Disponível em: <https://www.scielo.br/j/rbspa/a/ZfrLVQsLMQyJXqw6SX5QCTh/abstract/?lang=pt>. Acesso em: 6 Jun. 2024.

SARMENTO, J. L. R.; SENA, L.S. Avanços no melhoramento genético na América Latina e perspectivas futuras. **Arch. latinoam. prod. anim.**, [S.l.], v.25, n. 1-2, p.81-89, Nov. 2017. Disponível em: https://ojs.alpa.uy/index.php/ojs_files. Acesso em: 6 Jun. 2024.

SCOTT, B. A. *et al.* How genomic selection has increased rates of genetic gain and inbreeding in the Australian national herd, genomic information nucleus, and bulls. **J Dairy Sci.**, Champaign, v.104, n. 11, p. 11832-11849, Nov. 2021. DOI [10.3168/jds.2021-20326](https://doi.org/10.3168/jds.2021-20326). Disponível em: <https://pubmed.ncbi.nlm.nih.gov/34454757/>. Acesso em: 4 Jun. 2024.

SENA, L. S. *et al.* Genetic evaluation of tropical climate-adapted sheep for carcass traits including genomic information. **Small Rumin. Res.**, [S.l.], v. 188, p. 106120, July. 2020. DOI <https://doi.org/10.1016/j.smallrumres.2020.106120>. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0921448820300754>. Acesso em: 7 Jun. 2024.

SHARMA, A. *et al.* Stories and Challenges of Genome Wide Association Studies in Livestock — A Review. **Asian-Australas J Anim Sci**, Korea, v. 28, n. 10, p. 1371-1379, Oct. 2015. DOI [10.5713/ajas.14.0715](https://doi.org/10.5713/ajas.14.0715). Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4554843/>. Acesso em: 7 Jun. 2024.

SHARMA, A. *et al.* Accuracy of Imputation of Microsatellite Markers from BovineSNP50 and BovineHD BeadChip in Hanwoo Population of Korea. **Genomics Inform.**, Korea, v. 16, n. 1, p. 10, Mar. 2018. DOI [10.5808/GI.2018.16.1.10](https://doi.org/10.5808/GI.2018.16.1.10). Disponível em: <https://pubmed.ncbi.nlm.nih.gov/29618182/>. Acesso em: 7 Jun. 2024.

SHUMBUSHO, F. *et al.* Potential benefits of genomic selection on genetic gain of small ruminant breeding programs. **J. Anim. Sci.**, United States, v. 91, n. 8, p. 3644-3657, Aug. 2013. DOI <https://doi.org/10.2527/jas.2012-6205>. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/23736059/>. Acesso em: 7 Jun. 2024.

SHUMBUSHO, F. *et al.* Economic evaluation of genomic selection in small ruminants: a sheep meat breeding program. **Animal**, Castanet-Tolosan, v. 10, n. 6, p. 1033-1041, Jun. 2016. DOI: [10.1017/S1751731115002049](https://doi.org/10.1017/S1751731115002049). Disponível em: <https://pubmed.ncbi.nlm.nih.gov/26446712/>. Acesso em: 4 Jun. 2024.

SILVEIRA, R. M. F. *et al.* Tipificação, caracterização e diferenciação de sistemas de produção de ovinos no semiárido brasileiro. **NJAS: Impacto nas Ciências Agrícolas e da Vida**, v. 1, pág. 48-73, 2021. Acesso em: 7 Jun. 2024.

SORIO, A.; RASI, L. Ovinocultura e abate clandestino: um problema fiscal ou uma solução de mercado? **Rev. Política Agrícola**, Brasília, v. 19, n. 1, p. 71–83, Fev. 2010. ISSN: 2317-

224X. Disponível em: <https://seer.sede.embrapa.br/index.php/RPA/article/view/336>. Acesso em: 7 Jun. 2024.

TEIXEIRA, W. C. *et al.* Perfil zoonosológico dos rebanhos caprinos e ovinos em três mesorregiões do estado do Maranhão, Brasil. **Acta Vet. Bras.**, Brasil, v. 9, n. 1, p. 34-42, 2015. Disponível em: <https://pesquisa.bvsalud.org/porta/resource/pt/vti-303487>. Acesso em: 22 Jun. 2024.

TORRES, T. S. *et al.* Genetic evaluation of sheep for resistance to gastrointestinal nematodes and body size including genomic information. **Anim Biosci**, Korea, v. 34, n. 4, p. 516-524, Apr. 2020. DOI 10.5713/ajas.19.0816. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/32299165/>. Acesso em: 7 Jun. 2024.

TSAIRIDOU, S. *et al.* Optimizing low-cost genotyping and imputation strategies for genomic selection in Atlantic salmon. **G3 (Bethesda)**, England, v. 10, n. 2, p. 581-590, Feb. 2020. DOI 10.1534/g3.119.400800. Disponível em: <https://academic.oup.com/g3journal/article/10/2/581/6026251?login=false>. Acesso em: 7 Jun. 2024.

VAN DEN BERG, I.; FRITZ, S.; BOICHARD, D. QTL fine mapping with Bayes C(π): a simulation study. **Genet Sel Evol**, France, v. 45, n.19, June. 2013. DOI <https://doi.org/10.1186/1297-9686-45-19>. Disponível em: <https://gsejournal.biomedcentral.com/articles/10.1186/1297-9686-45-19>. Acesso em: 7 Jun. 2024.

VANRADEN, P. M. Invited review: reliability of genomic predictions for North American Holstein bulls. **J Dairy Sci**, United States, v. 92, n. 1, p. 16-24, Jan. 2009. DOI 10.3168/jds.2008-1514. Disponível em: [https://www.journalofdairyscience.org/article/S0022-0302\(09\)70305-4/fulltext](https://www.journalofdairyscience.org/article/S0022-0302(09)70305-4/fulltext). Acesso em: 7 Jun. 2024.

VANRADEN, P. M.; SUN, C.; O'CONNELL, J. R. Fast imputation using medium or low-coverage sequence data. **BMC Genet**, England, v. 16, n. 1, p. 1-12, July. 2015. DOI 10.1186/s12863-015-0243-7. Disponível em: <https://bmcbgenomdata.biomedcentral.com/articles/10.1186/s12863-015-0243-7>. Acesso em: 7 Jun. 2024.

VANRADEN, P. M. Symposium review: How to implement genomic selection. **J Dairy Sci.**, United States, June. 2020. DOI <https://doi.org/10.3168/jds.2019-17684>. Disponível em: [https://www.journalofdairyscience.org/article/S0022-0302\(20\)30309-X/fulltext](https://www.journalofdairyscience.org/article/S0022-0302(20)30309-X/fulltext). Acesso em: 7 Jun. 2024.

VENTURA, R. V. *et al.* Impact of reference population on accuracy of imputation from 6K to 50K single nucleotide polymorphism chips in purebred and crossbreed beef cattle. **J Anim Sci**, United States v. 92, n. 4, p. 1433-1444, Apr. 2014. DOI 10.2527/jas.2013-6638. Disponível em: <https://academic.oup.com/jas/article-abstract/92/4/1433/4703266?redirectedFrom=fulltext&login=false>. Acesso em: 7 Jun. 2024.

VENTURA, R. V. *et al.* Assessing accuracy of imputation using different SNP panel densities in a multi-breed sheep population. **Genet Sel Evol**, France, v. 48, n. 1, p. 1-20, Sept. 2016. DOI 10.1186/s12711-016-0244-7. Disponível em:

<https://gsejournal.biomedcentral.com/articles/10.1186/s12711-016-0244-7>. Acesso em: 7 Jun. 2024.

VIANA, J. G. A.; MORAES, M. R. E.; DORNELES, J. P. Dinâmica das importações de carne ovina no Brasil: análise dos componentes temporais. **Semina: Ciências Agrárias**, Londrina, v. 36, n. 3, p. 2223-2233, Julho, 2015. DOI <https://doi.org/10.5433/1679-0359.2015v36n3Supl1p2223>. Disponível em: <https://ojs.uel.br/revistas/uel/index.php/semagrarias/article/view/16694>. Acesso em: 7 Jun. 2024.

VITEZICA, Z. *et al.* Bias in genomic predictions for populations under selection. **Genet Res (Camb)**, England, v. 93, n. 5, p. 357-366, Oct. 2011. DOI 10.1017/S001667231100022X. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/21767459/>. Acesso em: 6 Jun. 2024.

WANG, H. *et al.* Genome-wide association mapping including phenotypes from relatives without genotypes. **Genet Res (Camb)**, England, v. 94, n. 2, p. 73-83, Apr. 2012. DOI 10.1017/S0016672312000274. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/22624567/>. Acesso em: 6 Jun. 2024.

WANG, C. *et al.* Accuracy of genomic prediction using an evenly spaced, low-density single nucleotide polymorphism panel in broiler chickens. **Poult Sci.**, England, v. 92, n. 7, p. 1712-1723, July. 2013. DOI 10.3382/ps.2012-02941. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/23776257/>. Acesso em: 6 Jun. 2024.

WEILER, J. I.; FERNANDO, R. L. Strategies for the improvement of animal production using marker-assisted selection. In: SCHOOK, L. B.; LEWIN, H. A.; McLAREN, D. G. (Ed.). **Gene-Mapping Techniques and Applications**. Boca Raton: CRC Press, 2020. p. 305-328. DOI:10.1201/9781003066712-17. Disponível em: <https://www.researchgate.net/publication/342931214> Strategies for the Improvement of Animal Production Using Marker-Assisted Selection. Acesso em: 6 Jun. 2024.

WIGGANS, G. R.; VANRADEN, P. M.; COOPER, T. A. The genomic evaluation system in the United States: Past, present, future. **J Dairy Sci.**, United States, v. 94, n. 6, p. 3202-3211, June. 2011. DOI 10.3168/jds.2010-3866. Disponível em: [https://www.journalofdairyscience.org/article/S0022-0302\(11\)00307-9/fulltext](https://www.journalofdairyscience.org/article/S0022-0302(11)00307-9/fulltext). Acesso em: 6 Jun. 2024.

WOLC, A. *et al.* **Applications of Genomic Selection in Poultry**. In: Proc 10th World Congress of Genetics Applied to Livestock Production (WCGALP), 2014. Disponível em: <http://bit.ly/1QwDcUm>. Acesso em: 7 Jun. 2024.

WOLC, A. *et al.* Implementation of genomic selection in the poultry industry. **Animal front.**, [S.l.], v. 6, n. 1, p. 23-31, Jan. 2016. DOI <https://doi.org/10.2527/af.2016-0004>. Disponível em: <https://academic.oup.com/af/article/6/1/23/4638780>. Acesso em: 6 Jun. 2024.

YIN, S. *et al.* A comprehensive genomic analysis of Chinese indigenous ningxiang pigs: genomic breed compositions, runs of homozygosity, and beyond. **Int. J. Mol. Sci.**, Basel, v. 24, n. 19, 14550, Sept. 2023. DOI <https://doi.org/10.3390/ijms241914550>. Disponível em: <https://www.mdpi.com/1422-0067/24/19/14550>. Acesso em: 6 Jun. 2024.

YOSHIDA, G. M. *et al.* Accuracy of genotype imputation and genomic predictions in a two-generation farmed Atlantic salmon population using high-density and low-density SNP panels. **Aquaculture**, v. 491, p. 147-154, Apr. 2018. DOI <https://doi.org/10.1016/j.aquaculture.2018.03.004>. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0044848617304544?via%3Dihub>. Acesso em: 7 Jun. 2024.

CAPÍTULO I

Imputação de genótipos em população de ovinos de corte simuladas com diferentes estruturas de referência e validação

Imputação de genótipos em população de ovinos de corte simuladas com diferentes estruturas de referência e validação

MATIAS, Ana Paula Soares e Silva ^{a *}, André Campêlo Araujo^b José Lindenberg Rocha Sarmento^c

^aPrograma de Pós-graduação em Ciência Animal, Centro de Ciências Agrárias (CCA), Universidade Federal do Piauí (UFPI), Campus Ministro Petrônio Portella, CEP 64049-550, Teresina, Piauí (PI), Brasil. E-mail: ana-soares.bio@hotmail.com

^bAcuFast™, 4003 Millar Ave #8, Saskatoon, SK S7K 2K6, Canada. Email: andre.araujo@acufastswine.com

^cDepartamento de Zootecnia (DZO), Centro de Ciências Agrárias (CCA), Universidade Federal do Piauí (UFPI), Campus Ministro Petrônio Portella, CEP 64049-550, Teresina, Piauí (PI), Brasil. Pesquisador CNPq. E-mail: sarmento@ufpi.edu.br * Autor correspondente

RESUMO: A imputação de genótipos é um método robusto para minimizar custos com genotipagem em espécies de interesse econômico. No entanto, muitos fatores influenciam na acurácia de imputação. O objetivo desse trabalho foi testar diferentes cenários de imputação de genótipos para viabilizar a implementação da seleção genômica em sistema de produção de ovinos. Os dados para o estudo foram gerados a partir de simulação usando *software* QMSim. A imputação foi realizada de um chip de 6K para um de 50K com o *software* FImpute em um total de 32 cenários. Verificou-se que o tamanho e a estrutura da população de referência, a proporção de indivíduos genotipados e o tamanho efetivo populacional (N_e) influenciaram nos resultados de acurácia de imputação de SNPs e indivíduos. Populações com maior N_e tiveram acurácias de imputação de SNPs mínimas de 0,919 e máxima de 0,96. Cenários com maior número de animais genotipados as médias de acurácias de imputação de SNPs variaram entre 0,943 e 0,96. Resultados semelhantes foram observados para a média de acurácia de indivíduos. A distribuição da acurácia de SNPs foi melhor nos cenários com maior N_e e maior número de animais genotipados. A troca de reprodutores não influenciou nas medidas de acurácia em relação as populações sem troca. A imputação do chip de 6k para 50k foi eficiente e pode ser aplicado para estruturas populacionais semelhantes como meio de facilitar a implementação da seleção genômica.

Palavras-chave: Acurácia, Predição genômica, Ovino, Genotipagem.

Imputation of genotypes in a beef sheep population simulated with different reference and validation structures

^aPost-graduation program in Animal Science, Agrarian Sciences Center (CCA), Federal University of Piauí (UFPI), Ministro Petrônio Portella University Campus, CEP 64049-550, Teresina, Piauí (PI), Brazil. E-mail: ana-soares.bio@hotmail.com

^bAcuFast™, 4003 Millar Ave #8, Saskatoon, SK S7K 2K6, Canada. Email: andre.araujo@acufastswine.com

^cDepartment of Animal Science (DZO), Agrarian Sciences Center (CCA), Federal University of Piauí (UFPI), Ministro Petrônio Portella University Campus, CEP 64049-550, Teresina, Piauí (PI), Brazil. CNPq researcher. E-mail: sarmento@ufpi.edu.br * Corresponding author

ABSTRACT: Genotype imputation is a robust method to minimize genotyping costs in species of economic interest. However, many factors influence imputation accuracy. The objective of this work was to test different genotype imputation scenarios to enable the implementation of genomic selection in a sheep production system. The data for the study was generated from simulation using the software QMSim. The imputation was performed from a 6K chip to a 50K chip with the software FImpute in a total of 32 scenarios. The study found that the size and structure of the reference population, the proportion of genotyped individuals, and the effective population size (N_e) influenced the results of imputation accuracy for SNPs and individuals. Populations with higher N_e had SNP imputation accuracies ranging from a minimum of 0.919 to a maximum of 0.96. In scenarios with a higher number of genotyped animals, the average SNP imputation accuracies varied between 0.943 and 0.96. Similar results were observed for the average accuracy of individuals. The distribution of SNP accuracy improved in scenarios with higher N_e and a greater number of genotyped animals. The exchange of breeders did not affect accuracy measurements compared to populations without exchange. Chip imputation from 6K to 50K was efficient and can be applied to similar population structures as a means of facilitating the implementation of genomic selection.

Keywords: Accuracy, Genomic prediction, Sheep, genotyping.

1 Introdução

A imputação de genótipos é uma técnica que completa genótipos ausentes em um painel de menor densidade a partir de informações de genótipos presentes em um painel de referência, com maior densidade. Essa abordagem tem sido aplicada de maneira eficiente na análise de dados genômicos permitindo a obtenção de sequências genômicas amplas a partir de chips de *single nucleotide polymorphism* – SNPs de menor densidade. Isso é feito a partir do uso de informações de desequilíbrio de ligação (LD) e taxa de recombinação em uma população de referência para inferir genótipos ausentes em uma população de validação (Kojima *et al.*, 2019; Li *et al.*, 2009).

A acurácia da imputação de genótipos está associada ao tamanho da população de referência, a relação entre a população de referência e validação, a densidade dos painéis de marcadores, a frequência dos alelos menos frequentes (MAF), a composição da população de referência (raça única/múltiplas – única linhagem/múltiplas), relações genéticas entre os indivíduos genotipados, a herdabilidade da (s) característica sob seleção, dentre outros (Hickey *et al.*, 2012; Shin *et al.*, 2018; Ventura *et al.*, 2016). No entanto, existem poucos estudos que avaliam esses aspectos na imputação de genótipos na espécie ovina. Dessa maneira, analisar estratégias para formação de uma população de referência apropriada para a implementação de um programa de seleção genômica para países em desenvolvimento a custo relativamente baixo é recomendada.

Em países como a Austrália e a Nova Zelândia, a implementação de programas de melhoramento genético fez com que esses países se tornassem referência na exportação de carne ovina (FAOSTAT, 2021). No Brasil, programas de melhoramento genético nessa espécie são quase inexistentes, sendo que um dos principais entraves é o custo sobre a genotipagem (Lobo, 2019).

O uso da imputação de genótipos tem sido aplicado em alguns estudos com a espécie ovina. Bolormaa *et al.* (2015) utilizaram essa metodologia para imputar um chip de 50K a partir de um chip de baixa densidade (12K) e avaliar a acurácia de imputação sob diferentes cenários. Marina *et al.* (2022) provaram que a partir da imputação de um chip de 3K foi possível estimar valores genéticos de forma acurada para características de produção de leite e queijo. Marina *et al.* (2021) utilizaram um número mínimo de SNPs para imputar marcadores microssatélites e desenvolver um chip de baixa densidade para verificação de parentesco na raça ovina Assaf. O uso da imputação de genótipos foi crucial para implementação da seleção genômica em ovinos na Nova Zelândia 80, portanto, avaliar a imputação como estratégia para diminuição de custos

com a genotipagem torna-se uma tarefa importante (Yoshida *et al.*, 2018) quando se busca soluções para implementação de programas de melhoramento genético, especialmente em ovinos.

O objetivo com esta pesquisa foi comparar diferentes cenários de imputação de genótipos para viabilizar a implementação da seleção genômica em sistema de produção de ovinos. Para isso, foram utilizados cenários com herdabilidade baixa e moderada, duas estratégias de genotipagem a partir de um painel de moderada densidade (50K) e troca de reprodutores entre as populações estudadas.

2 Material e Métodos

Não foi necessária a aprovação do Comitê de Ética no Uso de Animais, pois todos os dados utilizados foram gerados computacionalmente através de simulações.

2.1 Simulação de dados

2.1.1 Estrutura populacional

Várias populações de ovinos sob um sistema de acasalamento fechado ou com compartilhamento de reprodutores foram simuladas no *Software* QMSim (Sargolzaei; Schenkel, 2009). As populações simuladas foram obtidas em três etapas, em que a primeira foi formação da população histórica com um total de 16.000 animais com números iguais de machos e fêmeas. Posteriormente, foi simulado um gargalo populacional em que o número de indivíduos diminuiu para 10.000 animais (5.000 machos e 5.000 fêmeas), na geração 1.000. Em seguida, a população aumentou para 12.000 animais (4.000 machos e 8.000 fêmeas), na geração 1.500. Os acasalamentos foram realizados de forma aleatória e o efeito de mutação-deriva foi utilizado para criar o desequilíbrio de ligação (LD) inicial na população histórica. Na Figura 1 pode-se observar o esquema de simulação completo.

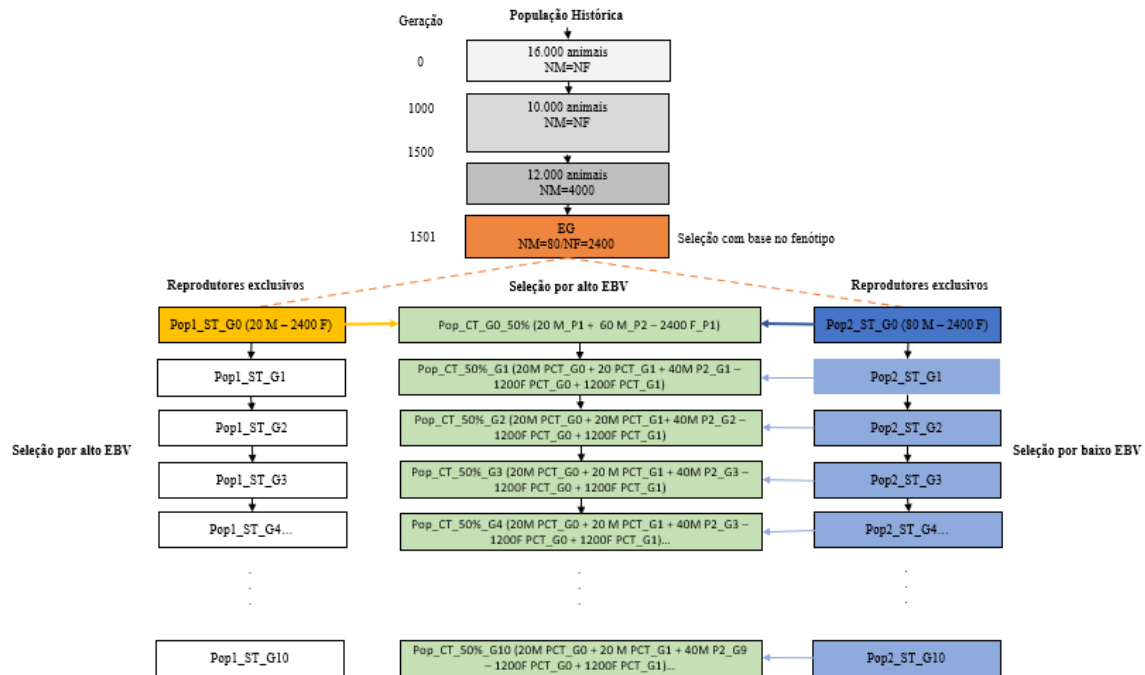


Figura 1. Esquema de simulação para obtenção de populações sem troca de reprodutores e com troca de reprodutores na espécie ovina. O mesmo *design* foi aplicado para obtenção de 10 gerações com troca de 5% de reprodutores. *POP_ST (população sem troca de reprodutor), POP_CT (população com troca de reprodutor).

Foram selecionados 80 machos e 2400 fêmeas da população histórica para formar a população de expansão (EG), em que o critério de seleção foi fenótipos altos e o descarte levou em consideração a idade dos animais na EG. A partir da última geração da EG foram criadas duas populações distintas com sistema de acasalamento fechado, ou seja, sem troca de reprodutores (POP1_ST e POP2_ST). Essas populações tiveram 10 gerações cada e número de fundadores diferentes (2.420 animais na POP1_ST e 2.480 para a POP2_ST). A seleção foi feita para EBV alto na população POP1_ST e baixo na população POP2_ST. Os reprodutores foram exclusivos em cada uma dessas populações. A proporção macho/fêmea nas populações selecionadas para alto EBV foi de 1/30, com uma taxa de substituição de 40% para reprodutores e 20% para matrizes, nas populações selecionadas para baixos EBV a proporção utilizada foi de 1/120. Essas proporções foram utilizadas para criar populações com tamanhos efetivos (N_e) diferentes ($N_e=80$ para POP1_ST e $N_e=300$ para POP2_ST). A taxa de nascimentos foi de 30% para parto simples, 50% para duplos e 20% para triplos (Araújo *et al.*, 2021).

Para obtenção das populações com troca de reprodutores, denominadas de POP_CT_50% (50% de troca de reprodutores) e POP_CT_5% (5% de troca de reprodutores), os fundadores foram trazidos da geração 0 das POP1_ST e POP2_ST e posteriormente das demais gerações da POP2_ST e das populações com troca recém-formadas (Figura 1). Também foram simuladas 10 gerações nas populações com troca. Houve sobreposição de gerações em

todas as gerações das populações sem e com troca de reprodutores. Nas populações sem troca, as taxas de reposição para machos e fêmeas foi de 20% e 40%, respectivamente. Para as populações com troca de reprodutores, eram mantidos 50% dos machos da população anterior e 50% eram trazidos da população sem troca de reprodutores (POP2_ST). A taxa de reposição de fêmeas nas populações com troca foi de 50%, com todas as fêmeas sendo selecionadas dentro da mesma população. A seleção e descarte de animais ocorreu por alto e baixo EBV na POP1_ST, respectivamente, assim como nas POP_CT e de maneira inversa na POP2_ST. As proporções de nascimentos nas POP_CT foram as mesmas utilizadas anteriormente para populações sem troca.

2.1.2 Características simuladas

Foram simuladas duas características, uma com herdabilidade de 0,30 (moderada) e 0,10 (baixa), com o intuito de representar a maioria das características de interesse zootécnico em ovinos. A variância fenotípica para ambas as características simuladas foi fixada em 100.

2.1.3 Genoma e edição de dados

O genoma simulado consistiu em 26 cromossomos autossômicos com tamanhos que variaram entre 43 e 301 centiMorgans (cM) totalizando em 2.656 cM, mimetizando o genoma ovino. As informações foram obtidas na plataforma NCBI (<https://www.ncbi.nlm.nih.gov/genome/?term=Ovis+aries>), *Assembly Oar_rambouillet* versão 4.0 (acesso em 2022). O *software* QMSim (Sargolzaei; Schenkel, 2009) também foi utilizado para a simulação do genoma.

Em todo o genoma autossômico simulado foram distribuídos 4.000 QTLs, que variou de 66 a 453 por cromossomo, de acordo com informações da plataforma *SheepQTLdb* (AnimalQTLdb, V.49, 2022). Os marcadores simulados foram bialélicos para serem semelhantes aos SNPs e o total de marcadores simulados foi de 46.827, conforme o painel de SNP de 50K para cromossomos autossômicos disponível na plataforma *SNPchiMp* v.3 (Nicolazzi *et al.*, 2015). Os SNPs simulados foram distribuídos de maneira uniforme pelo genoma a fim de gerar uma densidade de segregação com frequência de alelo menos frequente (MAF) > 0.05.

A distribuição gama com parâmetro 0,4 foi adotada para os efeitos genéticos aditivos dos QTLs. Não foram simulados efeitos para os SNPs. A herdabilidades dos QTL foram de 80 e 20% das herdabilidades das características de média e baixa herdabilidade, respectivamente.

Foram atribuídos efeitos poligênicos para a variância genética restante. Uma taxa de mutação recorrente foi aplicada na ordem de 1×10^{-4} para os QTL e marcadores. Com relação a ocorrências de genótipos faltantes e erros de genotipagem, taxas de 0,05 e 0,01 foram consideradas, respectivamente.

O controle de qualidade (CQ) foi feito no arquivo de genótipo das populações recentes com o uso do *software* PreGSf90, da família BLUPf90 (Misztal *et al.*, 2022). Os critérios adotados no CQ foram: diferença entre a heterozigosidade observada e esperada (menor que 0,15) e Equilíbrio de Hardy-Weinberg. Todos os SNPs das populações recentes estudadas passaram pelo CQ.

2.1.4 Populações de referência e validação

Dentro das populações estudadas com ou sem troca de reprodutores foram criados 2 grupos populacionais de referência e validação, a fim de testar a acurácia de imputação de um chip de baixa densidade (6K) para um de média densidade (50K), que é a principal proposta desse estudo. A população de referência inicial, com 100% dos animais genotipados, foi composta por, aproximadamente, 19.000 indivíduos amostrados aleatoriamente da geração seis até a nona e 5.000 na validação, provenientes da geração dez.

Para testar um outro cenário de imputação, mais realístico para a realidade de programas de melhoramento de ovinos em termos de número de animais genotipados, utilizou-se 5% da referência e da validação do total de animais genotipados no cenário inicial descrito anteriormente (aproximadamente 1.000 e 200 animais, respectivamente). O cenário inicial com 100% de animais genotipados com número grande de indivíduos foi definido para ter uma estimativa do que seria o valor esperado da acurácia de imputação em uma situação ótima.

2.1.5 Cenários avaliados

Foram avaliados um total de 16 cenários de estudo que consistiram em predições genômicas envolvendo 2 herdabilidades (0,30 e 0,10), 2 estratégias de genotipagem (100% e 5% dos animais genotipados), dois N_e ($N_e=80$ para POP1_ST e $N_e=300$ para POP2_ST) e duas estruturas de rebanho (com troca de 50% e 5% de reprodutores). Para cada um dos cenários descritos, foram testadas imputações levando em conta a população de maneira geral e o parentesco, nesse caso, totalizando 32 cenários de imputação. A Figura 2 traz um resumo dos cenários avaliados.

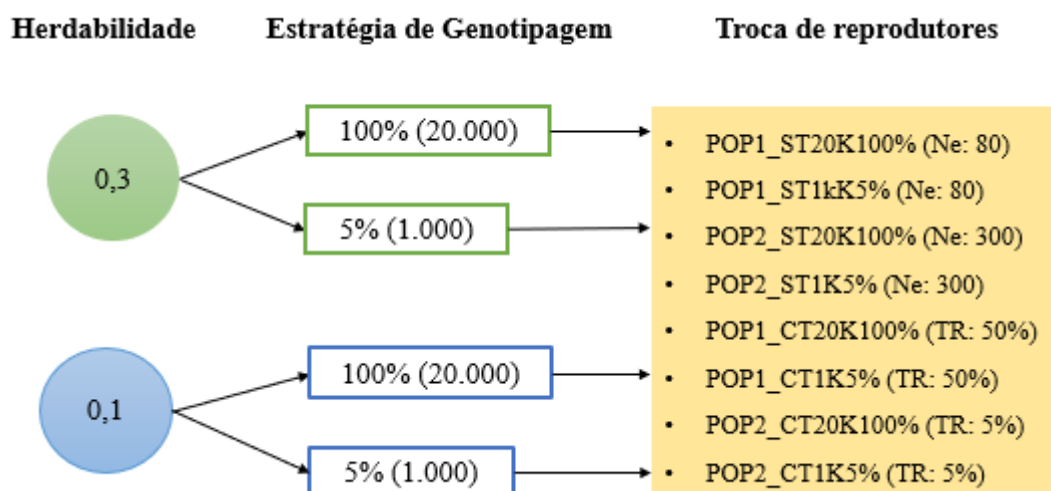


Figura 2. Cenários avaliados para imputação de genótipos em dados simulados de ovinos, considerando 2 herdabilidades (0,30 e 0,10); duas estratégias de genotipagem (100% e 5%), um painel de SNPS (50K) e 4 populações, sendo duas sem troca de reprodutores e com Ne distintos (300 e 80) e duas com troca de reprodutores (TR: 50% e TR: 5%) e mesmo Ne (300).

3 Imputação

Inicialmente um painel de 6K foi obtido a partir do painel simulado de 50K com a utilização de um programa desenvolvido pelos autores. Esse programa foi implementado em linguagem Linux e faz a seleção aleatória de um subconjunto de marcadores dentro de cada cromossomo. As informações necessárias são o número de cromossomos, número de marcadores presentes no cromossomo e número desejado de marcadores, sendo que o programa está disponível para o público por requerimento aos autores. A determinação da quantidade de marcadores em cada cromossomo foi feita com base na proporção do tamanho de cada cromossomo em relação ao tamanho total. Nesse sentido, foi dividido o tamanho do cromossomo pelo tamanho total do genoma (somatório do tamanho de cada cromossomo) e essa razão foi multiplicada pelo número de os marcadores no chip de menor densidade para cada cromossomo. O número de marcadores por cromossomo no chip de 50K e 6K pode ser observado na Tabela 1.

Tabela 1. Distribuição de marcadores SNPs por cromossomo em chip de moderada (50K) e baixa (6K) densidade.

Número de cromossomos	Número de SNP Chip 50K	Número de SNP Chip 6K
1	5244	675
2	4970	610
3	4465	549
4	2436	292
5	2109	264
6	2318	286
7	2005	245
8	1851	222
9	1901	232
10	1649	212
11	1077	152
12	1528	194
13	1527	203
14	1045	153
15	1506	198
16	1395	176
17	1291	177
18	1272	168
19	1124	148
20	998	125
21	768	122
22	984	124
23	1005	153
24	662	103
25	892	111
26	805	108
Total	46.827	6.002

A partir da obtenção do painel de menor densidade, foi realizado o controle de qualidade dos dados dos genótipos pelo *software Plink* v.1.9 (Purcell *et al.*, 2007). A imputação foi realizada em seguida utilizando o *software FImpute* v.3.0 (Sargolzaei *et al.*, 2014). A qualidade dos resultados foi medida pela acurácia de imputação dos SNPs e indivíduos utilizando a correlação da matriz padronizada dos genótipos imputados e reais que foram mascarados no processo de imputação, como sugerido por Calus *et al.* (2014). Todas as análises foram realizadas em um cluster de servidores de alta performance pertencente ao departamento de computação da Purdue University. Foram utilizados Rome CPUs com 2.0GHz e 1TB de memória para a simulação e Rome CPUs com 2.0GHz e 256GB de memória para as demais análises.

3 Resultados

3.1 Acurácia de imputação de SNPs

Os resultados apresentados levaram em consideração o método de imputação que considera população e família (pedigree), que foi ligeiramente melhor do que o método que considera apenas a população (dados não incluídos). A Tabela 2 traz os valores de acurácia de imputação para a herdabilidade 0,30 e 0,10. A partir das simulações realizadas é possível observar que as populações com maior N_e e maior número de animais na referência apresentaram maiores valores de acurácia (Cenários 3, 5, 7). Nos cenários com menor N_e , a acurácia foi menor quando o número de animais na referência foi menor (Cenário 2).

Tabela 2. Média de acurácia de imputação de SNPs para cada cenário avaliado contendo duas herdabilidades (0,30 e 0,10), valores mínimos (Mín), Máximos (Máx) e desvios padrão (SD).

Cenários	Acurácia de imputação h^2 0.30	Mín	Máx	SD	Acurácia de imputação h^2 0.10	Mín	Máx	SD
1- POP1_ST20K100%80	0,943	0,938	0,947	0,057	0,944	0,93	0,953	0,047
2- POP1_ST1kK5%80	0,927	0,921	0,933	0,093	0,928	0,915	0,938	0,087
3- POP2_ST20K100%300	0,956	0,956	0,958	0,014	0,957	0,957	0,959	0,013
4- POP2_ST1K5%300	0,927	0,924	0,934	0,048	0,922	0,919	0,93	0,047
5- POP1_CT20K100%300	0,954	0,952	0,955	0,020	0,955	0,954	0,956	0,016
6- POP1_CT1K5%300	0,928	0,926	0,933	0,051	0,926	0,924	0,927	0,048
7- POP2_CT20K100%300	0,957	0,956	0,958	0,023	0,959	0,958	0,96	0,017
8- POP2_CT1K5%300	0,934	0,933	0,935	0,054	0,933	0,931	0,936	0,049

Nas populações com troca de reprodutores (Cenários 5, 6, 7, 8), as acurácias foram semelhantes aos cenários sem troca (Cenários 1, 2, 3, 4). No entanto, o uso de menor proporção de troca de reprodutores e maior número de animais (Cenário 7), apresentou tendência de aumento na acurácia de imputação. Isso pode estar relacionado ao efeito da migração sobre a quebra e frequência dos haplótipos. Outro dado observado nessas populações é que em algumas populações com troca (Cenários 5, 8), a acurácia diminuiu conforme aumento da proporção de troca de reprodutores (Cenário 5). Esse efeito também pode estar relacionado ao efeito da migração sobre os haplótipos.

Com relação ao N_e nas populações sem troca (Cenário 1, 2, 3, 4), observou-se diferenças com relação a distribuição da acurácia dos SNPs, como pode ser observado na Figura 3.

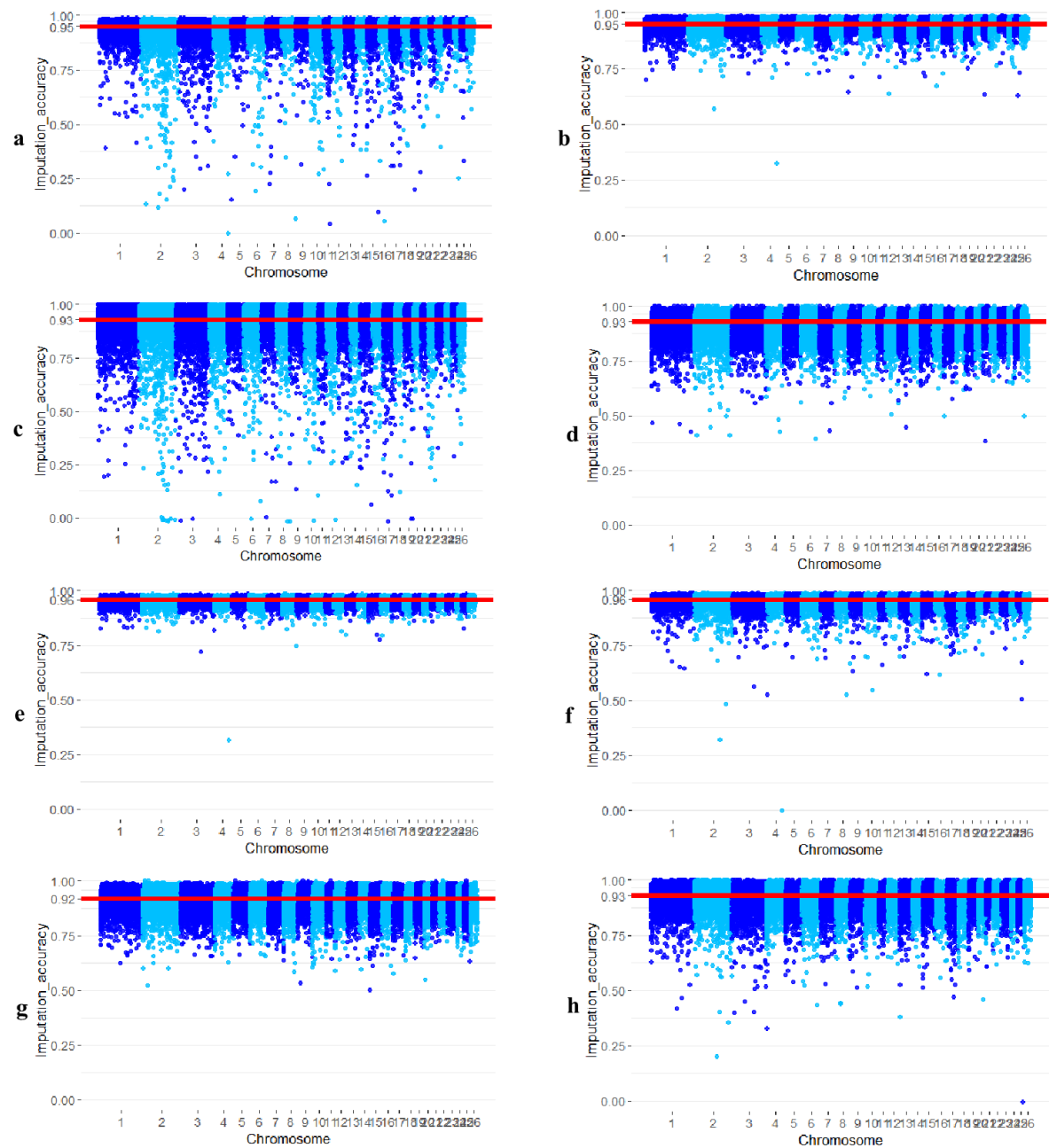


Figura 3. Acurácia de imputação de SNPs para a característica simulada com herdabilidade 0,30. **a:** POP1_ST20K100%80. **b:** POP1_CT20K100%300, TR: 50%. **c:** POP1_ST1K5%80. **d:** POP1_CT1K5%300, TR: 50%. **e:** POP2_ST20K100%300. **f:** POP2_CT20K100%300, TR: 5%. **g:** POP2_ST1K5%300. **h:** POP2_CT1K5%300, TR: 5%. * **Ne:** Número efetivo; **POP_ST20K100%80/300:** População sem troca de reprodutores com vinte mil animais, 100% genotipados, Ne=80 ou 300; **POP_ST1K5%80/300:** População sem troca de reprodutores, mil animais, 5% genotipados, Ne=80 ou 300; **TR:** Troca de reprodutores; **POP_CT20K100%300:** População com troca de reprodutores, vinte mil animais 100% genotipados, Ne=300; **POP_CT1K5%300:** População com troca de reprodutores, mil animais e 5% genotipados, Ne=300.

Os cenários com maior Ne apresentaram menores dispersão da acurácia dos SNPs, especialmente o que contava com maior número de animais (Figura 3e). O mesmo foi observado para as populações com troca com maior número de animais (Figuras 3b, f). O mesmo padrão foi observado com a herdabilidade mais baixa (0,10) (Figura 4).

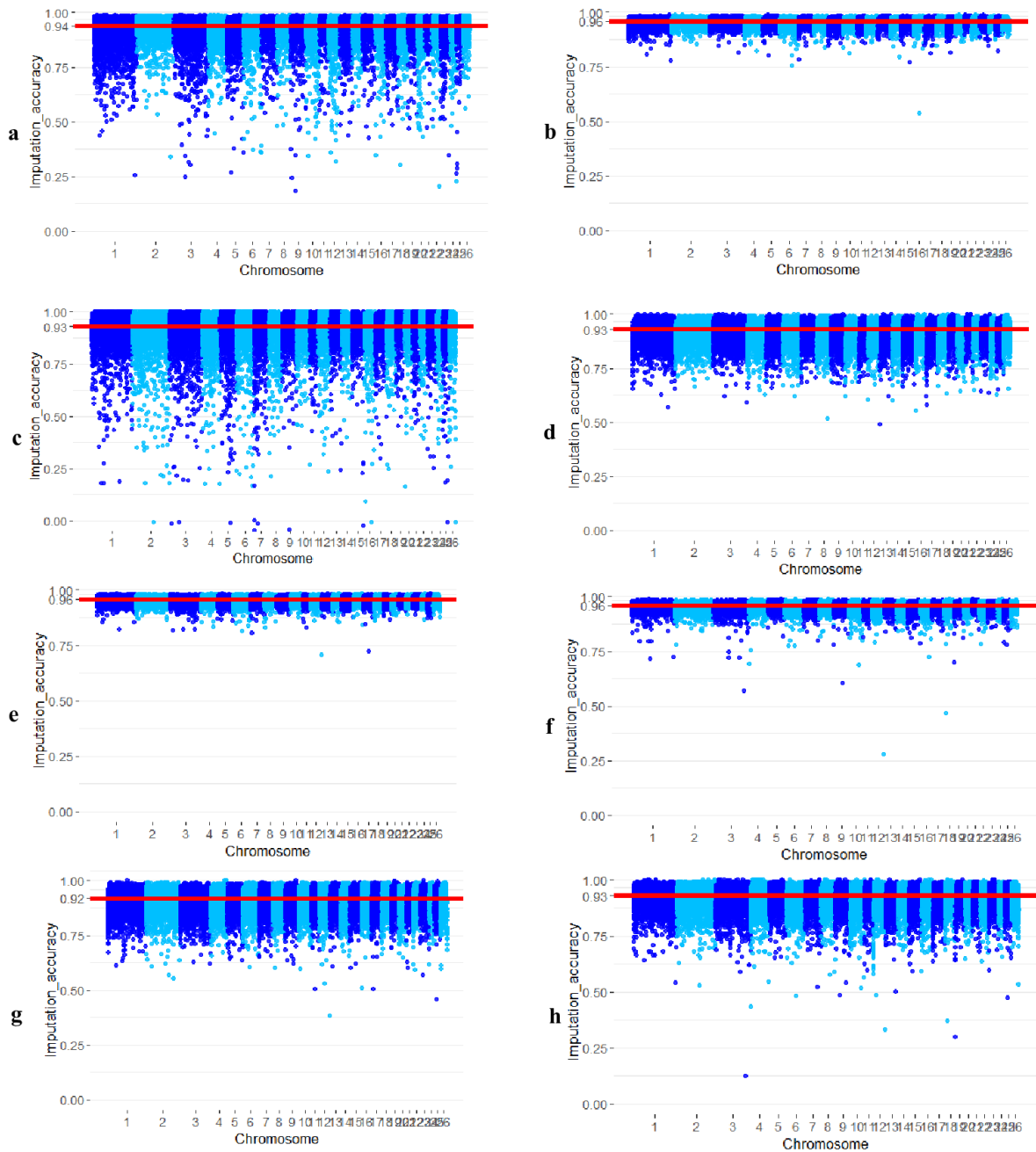


Figura 4. Acurácia de imputação de SNPs para a característica simulada com herdabilidade 0,10. **a:** POP1_ST20K100%80. **b:** POP1_CT20K100%300, TR: 50%. **c:** POP1_ST1K5%80. **d:** POP1_CT1K5%300, TR: 50%. **e:** POP2_ST20K100%300. **f:** POP2_CT20K100%300, TR: 5%. **g:** POP2_ST1K5%300. **h:** POP2_CT1K5%300, TR: 5%. * **Ne:** Número efetivo; **POP_ST20K100%80/300:** População sem troca de reprodutores com vinte mil animais, 100% genotipados, Ne=80 ou 300; **POP_ST1K5%80/300:** População sem troca de reprodutores, mil animais, 5% genotipados, Ne=80 ou 300; **TR:** Troca de reprodutores; **POP_CT20K100%300:** População com troca de reprodutores, vinte mil animais 100% genotipados, Ne=300; **POP_CT1K5%300:** População com troca de reprodutores, mil animais e 5% genotipados, Ne=300.

Outro fator observado é que a troca de reprodutores melhorou a distribuição da acurácia de SNPs (Figuras 3; 4b, d) quando comparado aos cenários sem troca de reprodutores com o mesmo número de animais e Ne menor (Figuras 3; 4a, c). Porém, os cenários com troca de reprodutores (Figuras 3; 4f, h) quando comparada com os cenários sem troca (Figuras 3; 4e, g),

todos com o mesmo N_e , a distribuição da acurácia de SNPs nas populações com troca piorou (maior dispersão). A Figura 5 mostra a distribuição das frequências de alelo menor (MAF) nas populações com menor e maior N_e .

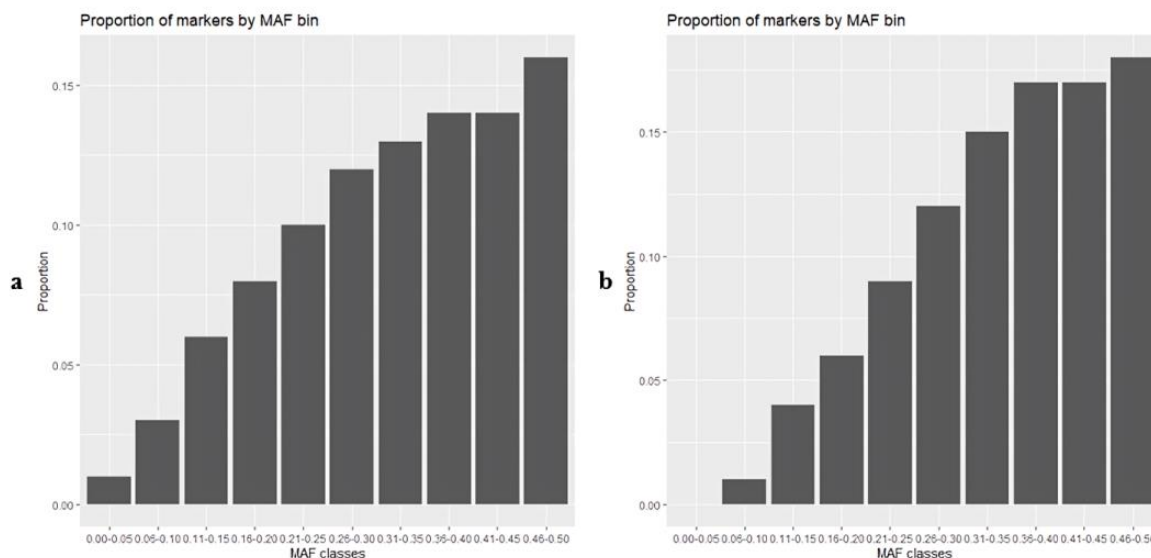


Figura 5. Distribuição da frequência de alelo menos frequente (MAF). **a:** população com menor número efetivo ($N_e=80$). **b:** população com maior número efetivo ($N_e=300$).

A comparação da distribuição da frequência alélica na população com menor e maior número efetivo evidenciou que na população com maior número efetivo (Figura 5b) a proporção de marcadores presentes na maior MAF foi maior. Maiores proporções de MAF menor foram observados na população com menor N_e (Figura 5a). É possível observar marcadores na faixa entre 0,00 e 0,05 na população com menor N_e (Figura 4a), que não foi visualizado nessa mesma escala na população com N_e maior (Figura 5b). Entre 0,06 e 0,10 a proporção de marcadores na população de maior N_e (Figura 5b) também foi menor.

O número de animais genotipados foi outro fator que interferiu nas médias de acurácias. Cenários em que se tinham maiores números de animais (~ 20.000) e 100% de animais genotipados no treinamento e na validação (Figuras 3; 4a, b, e, f) apresentaram as médias de acurácia acima de 0,94.

O cenário sem troca de reprodutores que obteve melhor média de acurácia foi o que contava com maior número efetivo e maior número de animais genotipados (POP2_ST20K100%300). O cenário com melhor média de acurácia das populações com troca foi o que teve menor proporção de troca e maior número de animais genotipados (POP2_CT20K100%300). A menor média de acurácia das populações sem troca foi observada no cenário com maior número efetivo e menor quantidade de animais na referência (POP1_ST1K5%80, POP2_ST1K5%300). Nos cenários com troca, médias de acurácias

menores foram observadas nos cenários com menor número de animais genotipados (POP1_CT1K5%300, POP2_CT1K5%300).

Neste estudo não houve grandes variações nas médias em relação as herdabilidade moderada e baixa. Em ambas herdabilidades a acurácia de imputação foi melhor na POP2_CT1K5%300 em relação a POP2_ST1K5%300.

3.2 Acurácia de imputação de indivíduos

Os valores de acurácia de indivíduos foram semelhantes aos valores de acurácia de SNPs (Tabela 3).

Tabela 3. Média de acurácia de indivíduos para cada cenário avaliado contendo duas herdabilidades (0.30 e 0.10), valores mínimos (Mín), Máximos (Máx) e desvios padrão (SD)

Cenários	Acurácia de imputação h^2 0.30	Mín	Máx	SD	Acurácia de imputação h^2 0.10	Mín	Máx	SD
1- POP1_ST20K100%80	0,942	0,937	0,946	0,009	0,943	0,929	0,953	0,010
2- POP1_ST1kK5%80	0,926	0,92	0,932	0,011	0,927	0,913	0,937	0,013
3- POP2_ST20K100%300	0,956	0,956	0,958	0,010	0,957	0,957	0,959	0,009
4- POP2_ST1K5%300	0,927	0,924	0,934	0,016	0,922	0,919	0,93	0,017
5- POP1_CT20K100%300	0,954	0,952	0,955	0,008	0,955	0,954	0,956	0,008
6- POP1_CT1K5%300	0,929	0,926	0,934	0,014	0,926	0,924	0,927	0,017
7- POP2_ST20K100%300	0,957	0,956	0,958	0,009	0,959	0,958	0,96	0,007
8- POP2_CT1K5%300	0,934	0,933	0,935	0,015	0,933	0,931	0,936	0,015

A acurácia de imputação de indivíduos nos cenários com herdabilidade 0,30, manteve-se estável, independente do parâmetro de análise utilizado (Figura 6). O mesmo comportamento foi observado nos cenários com herdabilidade 0,10, conforme se observa na Figura 7.

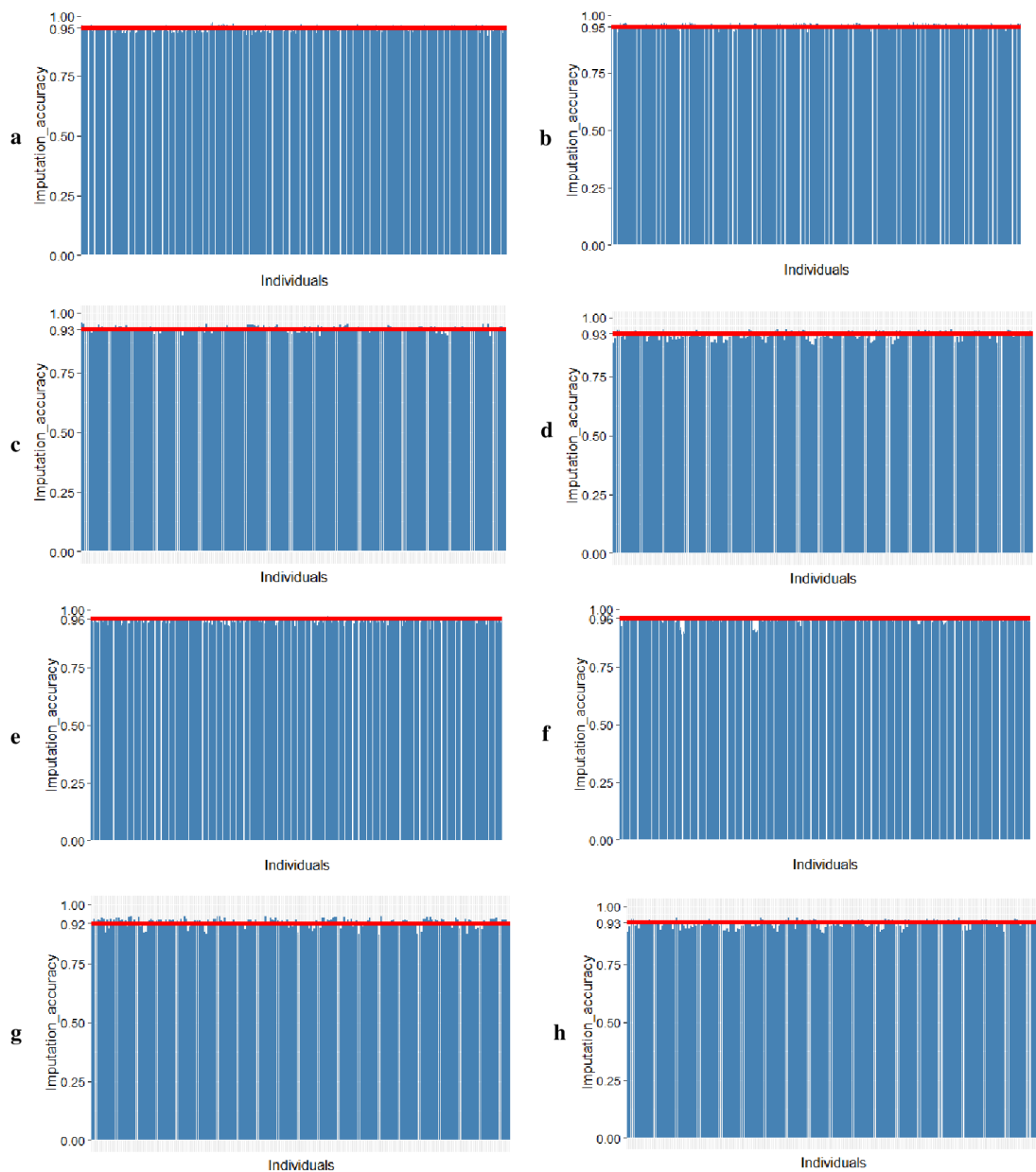


Figura 6. Acurácia de imputação de indivíduos para a característica simulada com herdabilidade 0,30. **a:** POP1_ST20K100%80. **b:** POP1_CT20K100%300, TR: 50%. **c:** POP1_ST1K5%80. **d:** POP1_CT1K5%300, TR: 50%. **e:** POP2_ST20K100%300. **f:** POP2_CT20K100%300, TR: 5%. **g:** POP2_ST1K5%300. **h:** POP2_CT1K5%300, TR: 5%. * **Ne:** Número efetivo; **POP_ST20K100%80/300:** População sem troca de reprodutores com vinte mil animais, 100% genotipados, Ne=80 ou 300; **POP_ST1K5%80/300:** População sem troca de reprodutores, mil animais, 5% genotipados, Ne=80 ou 300; **TR:** Troca de reprodutores; **POP_CT20K100%300:** População com troca de reprodutores, vinte mil animais 100% genotipados, Ne=300; **POP_CT1K5%300:** População com troca de reprodutores, mil animais e 5% genotipados, Ne=300.

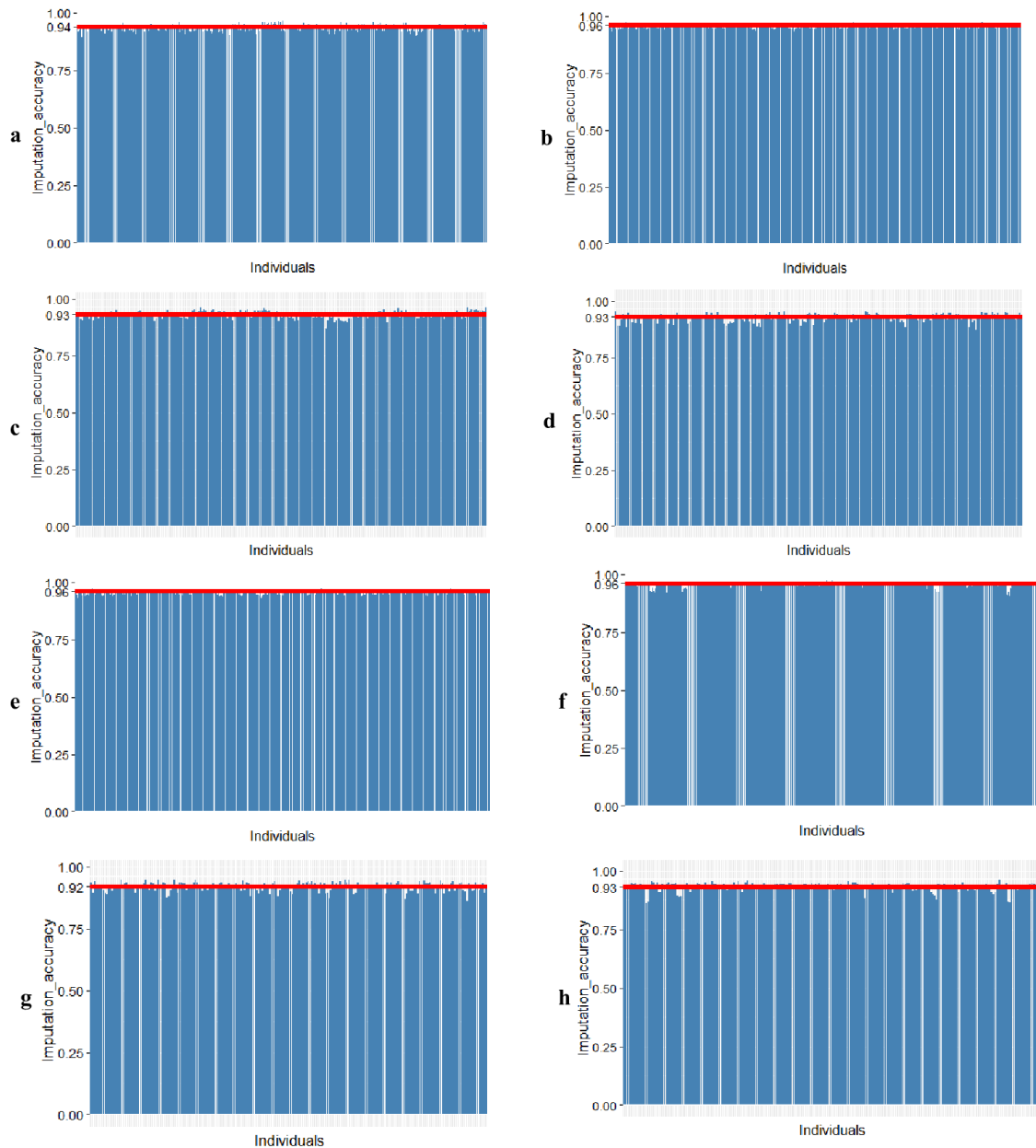


Figura 7. Acurácia de imputação de indivíduos para a característica simulada com herdabilidade 0,10. **a:** POP1_ST20K100%80. **b:** POP1_CT20K100%300, TR: 50%. **c:** POP1_ST1K5%80. **d:** POP1_CT1K5%300, TR: 50%. **e:** POP2_ST20K100%300. **f:** POP2_CT20K100%300, TR: 5%. **g:** POP2_ST1K5%300. **h:** POP2_CT1K5%300, TR: 5%. * **Ne:** Número efetivo; **POP_ST20K100%80/300:** População sem troca de reprodutores com vinte mil animais, 100% genotipados, Ne=80 ou 300; **POP_ST1K5%80/300:** População sem troca de reprodutores, mil animais, 5% genotipados, Ne=80 ou 300; **TR:** Troca de reprodutores; **POP_CT20K100%300:** População com troca de reprodutores, vinte mil animais 100% genotipados, Ne=300; **POP_CT1K5%300:** População com troca de reprodutores, mil animais e 5% genotipados, Ne=30

4 Discussão

Neste estudo, foram analisados fatores que afetam a acurácia de imputação, sendo eles a estrutura e tamanho da população de referência, a característica em estudo, o método de imputação e o relacionamento entre os indivíduos genotipados. O método de imputação utilizado com base em informações de marcadores e pedigree geraram resultados de acurácias

melhores, isso provavelmente porque a inclusão de informações de família permite rastrear haplótipos longos, que ocorrem de maneira mais frequente dentro da família e tem baixa frequência na população (Alillo *et al.*, 2018). De acordo com Sargolzaei *et al.* (2014) a inclusão de informações de família na imputação de genótipos é especialmente importante quando se considera alelos raros, que apresentam baixa acurácia quando a imputação é baseada apenas na população.

Foi possível observar que a acurácia de imputação aumentou nos cenários com maior número efetivo de animais. Provavelmente esse resultado está relacionado a MAF dos marcadores, que foi maior em populações com maior N_e . Menor N_e , frequentemente, resulta em maior acurácia de imputação devido maior extensão de blocos de haplótipos, que podem ser melhor preditos devido a maior frequência relativamente. No entanto, a acurácia de imputação também depende da MAF dos marcadores, que se espera que seja maior na população com maior N_e (Kai-Li *et al.*, 2022). Resultado semelhante foi demonstrado no trabalho de Kai-Li *et al.* (2022), que observaram aumento na acurácia de imputação conforme aumento da MAF em suínos. De forma semelhante, Bolormaa *et al.* (2018) relataram que a imputação do genoma completo em ovinos utilizando três conjuntos de referência evidenciou que a MAF baixa ($<0,05$) diminuiu a acurácia de imputação em todos os cenários avaliados.

O efeito do número de indivíduos genotipados nas populações de referência e validação nos valores de acurácia de imputação foi outro dado importante, em que maiores números de animais genotipados contribuíram positivamente para o aumento dessa média, assim como foi demonstrado em outros trabalhos (Calus *et al.*, 2014; Granado-Tajada; Varona; Ugarte, 2021; Ventura., 2016). Kai-Li *et al.* (2022), ao avaliarem fatores que afetam a acurácia de imputação de genótipos em suínos e Butty *et al.* (2019), ao trabalharem com bovinos, também relataram aumento de acurácia ao utilizar população de referência maiores no processo de imputação de genótipos.

A troca de reprodutores não proporcionou diferenças na média das acurácias de imputação em relação as populações sem troca na maior parte dos cenários avaliados. Isso pode ter acontecido devido ao banco de dados utilizado, que por se tratar de simulação, o pedigree é completo e não apresenta erros. No entanto, o cenário com menor troca de reprodutores e menor número de animais genotipados (POP2_CT5%300) teve média de acurácia melhor tanto na herdabilidade 0.10 como na herdabilidade 0.30. Provavelmente, a troca de reprodutores dentro de populações de referência menores e com alto N_e possa ter contribuído para esse resultado. Em ambos os cenários, na população com maior N_e , maior troca de reprodutores e maior número de animais genotipados (POP1_CT100%300), a acurácia de imputação também foi

maior do que a população sem troca. É provável que o número efetivo possa ter contribuído para o aumento da acurácia.

Tamanhos efetivos maiores nas populações sem troca estudadas melhorou a distribuição da acurácia dos SNPs por promover maior diversidade genética dentro da população e reduzir chances de fixação de alelos. Em análise mais específica, esses resultados podem ser úteis como parâmetro para estudos em que se pretende marcar regiões gênicas ou QTLs específicos, de modo que é possível melhorar valores de acurácia para regiões de interesse a partir da definição de estratégias mais elaboradas.

A redução na acurácia de imputação observada na população com maior proporção de troca de reprodutores pode estar relacionada ao aumento da diversidade genética devido a mistura de populações. Resultado semelhante foi relatado por Kai-Li *et al.* (2022), em que houve diminuição gradativa na acurácia de imputação à medida em que se aumentava a mistura de animais na população de referência.

O aumento na proporção de animais genotipados melhorou os valores de acurácias em todos os cenários avaliados. Em ovinos, a genotipagem de animais ainda é um entrave para os produtores, que geralmente investem pouco nessa tecnologia, devido ao alto custo da genotipagem relativo ao custo do animal. Portanto esses resultados são importantes de serem analisados quando se pensa na implantação de programas de melhoramento para essa espécie. Um resultado interessante neste estudo é que nos demais cenários, mesmo com menor número de animais na referência (~ 1000) e utilizando apenas 5% de animais genotipados no treinamento e na validação, a média da acurácia de imputação permaneceu acima de 90%. Provavelmente o pedigree completo e sem erros contribuiu para esse resultado. De acordo com Bolormaa *et al.* (2015) para se obter bons resultados na acurácia de predição a partir de dados imputados, é necessário o limite mínimo de 90% na acurácia de imputação.

O uso do chip de 6K para imputar o painel de 50K utilizados neste estudo alcançou altas acurácias. Esse resultado é promissor devido a possibilidade de usar essa metodologia para facilitar a implementação da seleção genômica na espécie ovina no Brasil. O principal interesse dos programas de melhoramento animal é maximizar a produção ao tempo em que reduz custos para obtenção desse ganho. Isso é especialmente importante para espécies que possuem baixo custo comercial e países em desenvolvimento. Estratégias de genotipagem que envolvam imputação de genótipos é uma alternativa interessante para implementação da seleção genômica nessas condições. Em um estudo feito por Marina *et al.* (2021), o uso da imputação a partir de baixo número de marcadores SNPs para imputar marcadores multialélicos também alcançou

altas acurácias e serviu como uma possibilidade de baixo custo para os criadores na identificação individual e de parentesco em ovinos.

A estabilidade observada na acurácia de indivíduos provavelmente se deve ao número de marcadores envolvidos, que considera todos os marcadores imputados presentes nos indivíduos. O aumento da acurácia de indivíduos observado apenas na POP1_CT1K5%300 com herdabilidade 0,30, se comparado aos valores de acurácia de SNPs, pode ter sido influenciado pelo aumento da diversidade genética devido a maior proporção de troca de reprodutores e menor quantidade de animais na referência.

5 Conclusão

Este trabalho demonstrou que é possível a obtenção de altas acurácias de imputação de acordo com os cenários simulados de populações de ovinos propostos. O método de imputação utilizando o chip de baixa densidade (6K) para média densidade (50K) mostrou-se eficiente. A troca de reprodutores, de maneira geral, não interferiu na acurácia de imputação. Os resultados descritos neste trabalho reforçam a ideia do uso do método de imputação como alternativa viável para facilitar a implementação da seleção genômica em populações com baixa quantidade de ovinos genotipados. No entanto, esse estudo serve de base para novos trabalhos nessa vertente, pois investigações adicionais serão necessárias para recomendações mais assertivas.

Referências

ALILOO, H. *et al.* The feasibility of using low-density marker panels for genotype imputation and genomic prediction of crossbred dairy cattle of East Africa. **J Dairy Sci.**, United States, v. 101, n. 10, p. 9108-9127, Oct. 2018. DOI [10.3168/jds.2018-14621](https://doi.org/10.3168/jds.2018-14621). Disponível em: <https://pubmed.ncbi.nlm.nih.gov/30077450/>. Acesso em: 20 Jun. 2024.

ARAÚJO, A. C. *et al.* A Comprehensive Comparison of Haplotype-Based Single-Step Genomic Predictions in Livestock Populations With Different Genetic Diversity Levels: A Simulation Study. **Front. Genet.**, v. 12, n. 729867, Oct. 2021. DOI <https://doi.org/10.3389/fgene.2021.729867>. Disponível em: <https://www.frontiersin.org/journals/genetics/articles/10.3389/fgene.2021.729867/full#B34>. Acesso em: 22 Jun. 2024.

AnimalQTLdb, release 49, 2022. Disponível em: <https://www.animalgenome.org/cgi-bin/QTLdb/OA/index>. Acesso em: 26 Jun. 2024.

BOLORMAA, S. *et al.* Non-additive genetic variation in growth, carcass and fertility traits of beef cattle. **Genet Sel Evol**, France, v. 47, n. 26, Apr. 2015. DOI [10.1186/s12711-015-0114-8](https://doi.org/10.1186/s12711-015-0114-8). Disponível em: <https://pubmed.ncbi.nlm.nih.gov/25880217/>. Acesso em 20 June. 2024.

BOLORMAA, S. *et al.* Evaluating the accuracy of imputed whole genome sequence in sheep. In: 11th Proceedings of the World Congress on Genetics Applied to Livestock Production. V. 11. Auckland: New Zealand; 2018:263.

BUTTY, A. M. *et al.* Optimizing Selection of the Reference Population for Genotype Imputation From Array to Sequence Variants. **Front. Genet.**, [S.l.], v. 10, n. 510, May. 2019. DOI [10.3389/fgene.2019.00510](https://doi.org/10.3389/fgene.2019.00510). Disponível em: <https://www.frontiersin.org/journals/genetics/articles/10.3389/fgene.2019.00510/full>. Acesso em: 20 Jun. 2024.

CALUS, M. P. L. *et al.* Evaluation of measures of correctness of genotype imputation in the context of genomic prediction: a review of livestock applications. **Animal**, Cambridge, v. 8, n. 11, p. 1743-1753, Nov. 2014. DOI [10.1017/S1751731114001803](https://doi.org/10.1017/S1751731114001803). Disponível em: <https://pubmed.ncbi.nlm.nih.gov/25045914/>. Acesso em: 20 Jun. 2024.

FAOSTAT: **FAO's corporate database** [online]. Rome: FAO, 2021. Disponível em: <https://www.fao.org/faostat/en/#home>. Acesso em: 02 out. 2023.

GRANADO-TAJADA, I.; VARONA, L.; UGARTE, E. Genotyping strategies for maximizing genomic information in evaluations of the Latxa dairy sheep breed. **J Dairy Sci.**, United States, v. 104, n. 6, pág. 6861-6872, June. 2021. DOI <https://doi.org/10.3168/jds.2020-19978>. Disponível em: [https://www.journalofdairyscience.org/article/S0022-0302\(21\)00463-X/fulltext](https://www.journalofdairyscience.org/article/S0022-0302(21)00463-X/fulltext). Acesso em: 20 Jun. 2024.

HICKEY, J. M. *et al.* Factors affecting the accuracy of genotype imputation in populations from several maize breeding programs. **Crop Sci.** [S.l.], v. 52, n. 2, p. 654-663, Mar. 2021. DOI <https://doi.org/10.2135/cropsci2011.07.0358>. Disponível em: <https://acess.onlinelibrary.wiley.com/doi/10.2135/cropsci2011.07.0358>. Acesso em: 23 Jun. 2024.

KAI-LI, Z. *et al.* A comprehensive evaluation of factors affecting the accuracy of pig genotype imputation using a single or multi-breed reference population. **Journal of Integrative Agriculture**, [S.l.] v. 21, n. 2, p. 486-495, Feb. 2022. DOI [https://doi.org/10.1016/S2095-3119\(21\)63695-X](https://doi.org/10.1016/S2095-3119(21)63695-X). Disponível em: <https://www.sciencedirect.com/science/article/pii/S209531192163695X>. Acesso em: 10 Jun. 2024.

KOJIMA, K. *et al.* A recurrent neural network based method for genotype imputation on phased genotype data. **bioRxiv**, [S.l.] p. 821504, Oct. 2019. DOI <https://doi.org/10.1101/821504>. Disponível em: <https://www.biorxiv.org/content/10.1101/821504v1.abstract>. Acesso em: 10 Mai. 2024.

LI, Y. *et al.* Genotype imputation. **Annu Rev Genomics Hum Genet.**, United States, v. 10, p. 387-406, Sep. 2009. DOI <https://doi.org/10.1146/annurev.genom.9.081307.164242>. Disponível em: <https://www.annualreviews.org/content/journals/10.1146/annurev.genom.9.081307.164242>. Acesso em: 10 Jan. 2022.

LOBO, R. N. B. Opportunities for investment into small ruminant breeding programmes in Brazil. **J Anim Breed Genet**, [S.l.], v. 136, n. 5, p. 313-318, May. 2019. DOI <https://doi.org/10.1111/jbg.12396>. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1111/jbg.12396>. Acesso em: 6 Jun. 2024.

MARINA, H. *et al.* Accuracy of Imputation of Microsatellite Markers from a 50K SNP Chip in Spanish Assaf Sheep. **Animals**, Basel, v. 11, n. 1, p. 86, Jan. 2021. DOI [10.3390/ani11010086](https://doi.org/10.3390/ani11010086). Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7824810/>. Acesso em: 6 Jun. 2024.

MARINA, H. *et al.* Low-density SNP panel for efficient imputation and genomic selection of milk production and technological traits in dairy sheep. **J Dairy Sci.**, United States, v. 105, n. 10, p. 8199-8217, Oct. 2022. DOI <https://doi.org/10.3168/jds.2021-21601>. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0022030222004726>. Acesso em: 20 Jun. 2024.

MISZTAL, I. *et al.* **Manual for BLUPF90 family of programs**. Athens, June. 2022. Disponível em: <https://nce.ads.uga.edu/wiki/doku.php?id=documentation>. Acesso em: 05 Jan. 2021.

MORRIS, S. T.; KENYON, P. R. Intensive sheep and beef production from pasture – A New Zealand perspective of concerns, opportunities and challenges. **Meat Sci.**, England, v. 98, p. 330- 335, Nov. 2014. DOI <https://doi.org/10.1016/j.meatsci.2014.06.011>. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/24998778/>. Acesso em: 8 Mar. 2023.

MORRIS, S. T. Overview of sheep production systems. In: Ferguson, D. M.; Lee, C.; Fisher, A. (Ed.). **Woodhead Publishing Series in Food Science, Technology and Nutrition, Advances in Sheep Welfare**. Woodhead Publishing, 2017, p. 19-35. DOI <https://doi.org/10.1016/B978-0-08-100718-1.00002-9>. Disponível em: <https://www.sciencedirect.com/science/article/pii/B9780081007181000029>. Acesso em: 20 Jun. 2024.

NICOLAZZI, E. L. *et al.* SNPchiMp v.3: integrating and standardizing single nucleotide polymorphism data for livestock species. **BMC Genomics**, London, v. 16, n. 283, Apr. 2015. DOI: [10.1186/s12864-015-1497-1](https://doi.org/10.1186/s12864-015-1497-1). Disponível em: <https://pubmed.ncbi.nlm.nih.gov/25881165/>. Acesso em: 20 Jun. 2024.

PURCELL, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. **Am J Hum Genet**, United States, v. 81, n. 3, p. 559-575, Jul. 2007. DOI [10.1086/519795](https://doi.org/10.1086/519795). Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1950838/>. Acesso em: 26 Jun. 2024.

SARGOLZAEI, M., SCHENKEL, F. S. QMSim: a large-scale genome simulator for livestock. **Bioinformatics**, England, v. 25, n. 5, p. 680-681, Mar. 2009. DOI <https://doi.org/10.1093/bioinformatics/btp045>. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/19176551/>. Acesso em: 20 Jun. 2024.

SARGOLZAEI, M.; CHESNAIS, J. P.; SCHENKEL, F. S. A new approach for efficient genotype imputation using information from relatives. **BMC Genomics**, London, v. 15, n. 478, June. 2014. DOI <https://doi.org/10.1186/1471-2164-15-478>. Disponível em:

<https://bmcbgenomics.biomedcentral.com/articles/10.1186/1471-2164-15-478>. Acesso em: 20 Jun. 2024.

SHIN, D. *et al.* Extent of linkage disequilibrium and effective population size of Korean Yorkshire swine. **Asian-Australas J Anim Sci**, Korea (South), v. 31, n. 12, p. 1843, Dec. 2018. DOI [10.5713/ajas.17.0258](https://doi.org/10.5713/ajas.17.0258). Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6212734/>. Acesso em: 12 Mai. 2023.

SOUZA, K. C. *et al.* Escore de condição corporal em ovinos visando a sua eficiência reprodutiva e produtiva. **Pubvet**, v. 5, n. 1, Jan. 2011. Disponível em: <https://ojs.pubvet.com.br/index.php/revista/article/view/2419>. Acesso em: 12 Mai. 2023.

VENTURA, R. V. *et al.* Assessing accuracy of imputation using different SNP panel densities in a multi-breed sheep population. **Genet Sel Evol**, France, v. 48, n. 1, p. 1-20, Sept. 2016. DOI [10.1186/s12711-016-0244-7](https://doi.org/10.1186/s12711-016-0244-7). Disponível em: <https://gsejournal.biomedcentral.com/articles/10.1186/s12711-016-0244-7>. Acesso em: 7 Jun. 2024.

YOSHIDA, G. M. *et al.* Accuracy of genotype imputation and genomic predictions in a two-generation farmed Atlantic salmon population using high-density and low-density SNP panels. **Aquaculture**, v. 491, p. 147-154, Apr. 2018. DOI <https://doi.org/10.1016/j.aquaculture.2018.03.004>. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0044848617304544?via%3Dihub>. Acesso em: 7 Jun. 2024.

CAPÍTULO II

**Predição genômica com dados simulados de ovinos em programa de melhoramento
usando diferentes estratégias de imputação**

Predição genômica com dados simulados de ovinos em programa de melhoramento usando diferentes estratégias de imputação

MATIAS, Ana Paula Soares e Silva ^{a *}, André Campêlo Araújo^b José Lindenberg Rocha Sarmento^c

^aPrograma de Pós-graduação em Ciência Animal, Centro de Ciências Agrárias (CCA), Universidade Federal do Piauí (UFPI), Campus Ministro Petrônio Portella, CEP 64049-550, Teresina, Piauí (PI), Brasil. E-mail: ana-soares.bio@hotmail.com

^bAcuFast™, 4003 Millar Ave #8, Saskatoon, SK S7K 2K6, Canada. Email: andre.araujo@acufastswine.com

^cDepartamento de Zootecnia (DZO), Centro de Ciências Agrárias (CCA), Universidade Federal do Piauí (UFPI), Campus Ministro Petrônio Portella, CEP 64049-550, Teresina, Piauí (PI), Brasil. Pesquisador CNPq. E-mail: sarmento@ufpi.edu.br * Autor correspondente

RESUMO: A predição genômica tornou os programas de melhoramento mais rápidos e eficientes. No entanto, esse método ainda não completamente estabelecido em algumas espécies de animais de interesse econômico, principalmente em países em desenvolvimento. O objetivo desse trabalho foi avaliar a qualidade da predição genômica em dados simulados de ovinos provenientes de diferentes estratégias de imputação de dados genotípicos. Os dados para este estudo foram gerados via simulação utilizando o *software* QMSim, que serviram de base para um estudo prévio de imputação de genótipos com o *software* FImpute. A predição foi realizada considerando os dados reais e imputados para uma característica de baixa herdabilidade (0,10). O método single-step GBLUP foi utilizado para a predição dos GEBVs. Para as medidas de validação, a acurácia de predição foi obtida pela correlação de Pearson entre o TBV e o GEBV na população de validação. O viés de predição foi avaliado pela diferença entre a média do GEBV e TBV e a dispersão dos GEBVs foi medida como 1 menos o coeficiente de regressão do TBV sobre o GEBV. As médias dos cenários foram comparadas pelo teste t e tukey a 5%. Os resultados mostraram que os cenários com maior proporção de troca de reprodutores resultaram em melhores valores de acurácia, acima de 0,50 chegando a 0,85. Os vieses tiveram melhores valores nas populações com maior proporção de troca de reprodutores e piores nas populações sem troca. As dispersões, no geral, foram baixas em todos os cenários avaliados. O método ssGBLUP mostrou-se eficiente para analisar os cenários de estudos, no entanto, serão necessários estudos posteriores para uma melhor compreensão sobre os fatores que afetaram as métricas de predição desse estudo. A imputação foi eficiente para todos os cenários de predição avaliados, especialmente os cenários que incluiu a troca de reprodutores.

Palavras-chave: Genotipagem, Acurácia, Seleção genômica

Genomic prediction with simulated data from sheep in a breeding program using different imputation strategies

^aPost-graduation program in Animal Science, Agrarian Sciences Center (CCA), Federal University of Piauí (UFPI), Ministro Petrônio Portella University Campus, CEP 64049-550, Teresina, Piauí (PI), Brazil. E-mail: ana-soares.bio@hotmail.com

^bAcuFast™, 4003 Millar Ave #8, Saskatoon, SK S7K 2K6, Canada. Email: andre.araujo@acufastswine.com

^cDepartment of Animal Science (DZO), Agrarian Sciences Center (CCA), Federal University of Piauí (UFPI), Ministro Petrônio Portella University Campus, CEP 64049-550, Teresina, Piauí (PI), Brazil. CNPq researcher. E-mail: sarmento@ufpi.edu.br * Corresponding author

ABSTRACT: Genomic prediction has made breeding programs faster and more efficient. However, this method has not yet been fully established in some animal species of economic interest, particularly in developing countries. The objective of this work was to evaluate the quality of genomic prediction in simulated sheep data from different genotypic data imputation strategies. The data for this study were generated via simulation using the QMSim software, which served as the basis for a prior genotype imputation study with the FImpute software. The prediction was carried out using real and imputed data for a trait with low heritability (0.10). The single-step GBLUP (ssGBLUP) method was used to predict GEBVs. For validation, prediction accuracy was obtained by Pearson's correlation between TBV and GEBV in the validation population. Prediction bias was assessed by the difference between the means of GEBV and TBV, and the dispersion of GEBVs was measured as 1 minus the regression coefficient of TBV on GEBV. The means of the scenarios were compared using the t-test and Tukey's test at a 5% significance level. The results showed that scenarios with a higher proportion of reproductive exchange resulted in better accuracy values, ranging from above 0.50 to 0.85. Biases were lower in populations with a higher proportion of reproductive exchange and higher in populations without exchange. Dispersion values were generally low across all evaluated scenarios. The ssGBLUP method proved efficient for analyzing the study scenarios, although further studies are needed to better understand the factors affecting the prediction metrics in this study. Imputation was efficient across all prediction scenarios evaluated, particularly those involving reproductive exchange.

Keywords: Genotyping, Accuracy, Genomic Selection.

1 Introdução

O melhoramento genético clássico, baseado no fenótipo e pedigree, proporcionou importantes ganhos na produção animal. Devido a algumas limitações no melhoramento genético clássico, a seleção genômica (*genomic selection*, GS) foi desenvolvida e implementada, proporcionando maior eficiência na avaliação de características de difícil mensuração, baixa herdabilidade ou que se manifestam tardiamente na vida do animal (Meuwissen; Hayes; Goddard, 2001). A GS ocupa lugar de destaque entre as metodologias utilizadas para aumentar a eficiência produtiva através do melhoramento genético dos rebanhos (Meuwissen; Hayes; Goddard, 2016).

A implementação da GS baseia-se na formação de uma população de treinamento, que tem genótipo e fenótipo conhecidos (Meuwissen; Hayes; Goddard, 2001). Na população de treinamento, os efeitos dos marcadores genéticos são estimados, e posteriormente serão usados para a predição de valores genéticos em indivíduos jovens. Normalmente, esse processo é feito com base em marcadores moleculares do tipo *Single Nucleotide Polymorphisms* (SNPs) (MEUWISSEN; HAYES; GODDARD, 2013). Segundo os mesmos autores, isso é possível devido os SNPs estarem em equilíbrio de ligação com os loci de características quantitativa (QTLs). No entanto, é importante validar os resultados da predição de valores genéticos genômicos estimados (GEBVs) pelo uso de populações de treinamento e métricas adequadas (acurácia, viés, dispersão, entre outras) (Legarra; Reverter, 2018).

Embora os benefícios da seleção genômica sejam evidentes, sua implementação depende de um estudo detalhado antes de sua realização (escolha de animais, método estatístico, painel de genotipagem, entre outros), bem como o controle de qualidade adequado dos dados a serem utilizados (Sharma *et al.*, 2015). Em espécies em que os efetivos populacionais são menores, a carência de coleta consistente de fenótipos e controle zootécnico também é um fator importante (Blasco; Toro, 2014; Meuwissen; Hayes; Goddard, 2016; Sharma *et al.*, 2015). No entanto, a parte estrutural e financeira é tida como maior obstáculo, devido a necessidade de grandes populações de treinamento e limitações quanto a investimentos em genotipagem (Blasco; Toro, 2014), especialmente em países em desenvolvimento (Mrode *et al.*, 2018, 2019).

Em espécies cujos animais apresentam relativamente baixo valor econômico, como é o caso dos ovinos, a decisão sobre a viabilidade da implementação da genômica é ainda mais relevante, especialmente quando se considera os custos com a genotipagem (Blasco; Toro, 2014; Lobo, 2019; Rupp *et al.*, 2016). Para isso, a imputação de genótipos é uma aliada nesse processo, pela diminuição do custo da genotipagem de animais. No entanto, a avaliação da

acurácia de imputação é crucial para estudos de predição genômica, pois afeta diretamente a confiabilidade das análises. Alta acurácia de imputação é essencial para otimizar o uso eficiente dos dados genômicos (Bolormaa *et al.*, 2015), assim como o uso de medidas de qualidade de predição genômica, como acurácia, viés e dispersão, considerando diferentes estratégias de imputação de genótipos.

O uso da simulação de dados é uma estratégia interessante para testar cenários de predição genômica quando comparado ao uso de dados reais, pois requer apenas recursos computacionais e definição de parâmetros estatísticos adequados para análise (Daetwyler *et al.*, 2013). Diversos estudos genômicos com base em simulação de dados têm sido utilizados na espécie ovina. Para avaliar o potencial reprodutivo e a resistência às mudanças climáticas (Karadavut *et al.*, 2022), comparar o efeito de diferentes estratégias de genotipagem na acurácia de predição de características de interesse econômico (Araujo *et al.*, 2021; Granado-Tajada; Varona; Ugarte, 2021) e também para testar e comparar diferentes estratégias de implementação da seleção genômica (Lillehammer *et al.*, 2020).

Na maioria dos países em desenvolvimento, o uso da seleção genômica e sua implementação na espécie ovina ainda é incipiente. Nesse sentido, estudos que buscam avaliar alternativas para implementação dessas estratégias são necessários. Dessa forma, o objetivo principal desse trabalho foi avaliar a qualidade da predição genômica em dados simulados de ovinos provenientes de diferentes estratégias de imputação de dados genotípicos.

2 Material e Métodos

2.1 Simulação de dados

A simulação de dados foi a mesma realizada no capítulo anterior (Capítulo I) para as características com herdabilidade 0,10. Brevemente, o *software* QMSim (Sargolzaei; Schenkel, 2009) foi utilizado para simular bancos de dados de ovinos em programa de melhoramento com diferentes estruturas populacionais. Para formação dos cenários, foram utilizados diferentes números de animais genotipados (20K e 1K), números efetivos (300, 80) e estruturas de rebanho (com e sem troca de reprodutores). O total de indivíduos com fenótipos foi de aproximadamente 41.000 animais no treinamento e 4.500 na validação, totalizando ~ 45.500. A variância fenotípica foi fixada em 100. Foram simuladas 4 populações, sendo duas com troca de reprodutores e duas sem troca de reprodutores, cada uma com 10 gerações. Foram criados 8 cenários de estudo, sendo simulados quatro cenários de populações sem troca de reprodutores: População sem troca de reprodutores, vinte mil animais, tamanho efetivo (N_e) 80

(POP1_ST20K_100%80); população sem troca de reprodutores, mil animais, tamanho efetivo (Ne) 80 (POP1_ST1K_5%80). As duas outras populações sem troca (POP2_ST), seguiram as mesmas condições, diferindo apenas o número efetivo, que foi de 300 animais (POP2_ST20K_100%300, POP2_ST1K_5%300). Outros quatro cenários de populações com troca de reprodutores foram criados: População com troca de 50% reprodutores, vinte mil animais, (POP1_CT20K_100%300); população com troca de 5% de reprodutores, mil animais (POP1_CT1K_5%300). As duas outras populações sem troca (POP2_CT), seguiram as mesmas condições, diferindo apenas a porcentagem de reprodutores, que foi de 5% (POP2_CT20K_100%300, POP2_CT1K_5%300). Para cada um dos cenários foram feitas cinco repetições.

2.2 Predição dos valores genéticos genômicos

2.2.1 Estimação dos componentes de variância

Os dados gerados dos cenários simulados foram base para análises nos programas da família BLUPF90 (Misztal *et al.*, 2022). Primeiramente, os dados foram renumerados no *software* RENUMF90 para servirem como entradas nos *softwares* subsequentes da família. Sem a inclusão das informações dos SNPs, foi executado o BLUPF90+ visando, primariamente, estimar os componentes de variância, utilizando uma combinação dos algoritmos AI-REML e EM-REML. No modelo unicaracterístico usado para estimação dos componentes de variância, foram considerados: como efeitos fixos, o sexo dos animais, com dois níveis, e geração, com dez níveis; e como efeito aleatório, os animais e o resíduo, com níveis variando de acordo com o número de indivíduos em cenário. Matricialmente, o modelo pode ser representado da seguinte forma:

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

em que: $\mathbf{y}$ é o vetor de observações para a característica simulada; $\boldsymbol{\beta}$ é o vetor de efeitos fixos; $\boldsymbol{\alpha}$ é o vetor de efeitos genéticos aditivos com $\boldsymbol{\alpha} \sim N(0, \mathbf{A}\sigma_a^2)$, sendo $\mathbf{A}$ a matriz de parentesco, $\mathbf{X}$ é a matriz de incidência dos efeitos fixos; $\mathbf{Z}$ é a matriz de incidência dos efeitos genéticos aditivos; $\mathbf{e}$ é o vetor de efeitos residuais, de modo que $\mathbf{e} \sim N(0, \mathbf{I}\sigma_e^2)$, sendo $\mathbf{I}$ a matriz identidade.

Após estimar os componentes de variância, A seguinte fórmula foi utilizada para calcular a herdabilidade:

$$h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2},$$

onde a herdabilidade no sentido restrito (h^2) é a razão entre a variância genética aditiva (σ_a^2) e a variância fenotípica ($\sigma_a^2 + \sigma_e^2$).

2.2.2 BLUP genômico

O método BLUP genômico de passo único (*single-step* GBLUP ou ssGBLUP), proposto por Legarra *et al.* (2009) e Christensen e Lund (2010) foi utilizado para a predição dos GEBVs. Antes da predição genômica, um controle de qualidade dos dados genotípicos foi realizado com o PreGSf90 (Misztal *et al.*, 2022). Os critérios utilizados no controle de qualidade foram: *Call rate* para marcadores <95%, *call rate* por animal <0,90, Frequência do alelo menos frequente (MAF) <0,05 e diferença entre heterosigozidade observada e esperada > 0.15 para remoção de SNPs. Apenas marcadores presentes em autossômicos foram simulados e utilizados nas análises (veja o Capítulo I para maiores detalhes do processo de simulação).

Os componentes de variância, estimados como descrito na seção anterior, pedigree, fenótipos mais as informações dos SNPs foram utilizadas nas análises de predição, que também foram realizadas com o *software* BLUPF90+. A inversa da matriz de parentesco híbrida ($\mathbf{H}^{-1}$), resultante da combinação das matrizes $\mathbf{A}$ (parentesco de pedigree) e matriz $\mathbf{G}$ (parentesco genômico), foi calculada de acordo com Aguilar *et al.* (2010) e utilizada diretamente as equações do modelo misto, apresentando a seguinte forma:

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

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

A matriz de parentesco genômico $\mathbf{G}$ foi calculada de acordo como o método I proposto por VanRaden (2008), sendo:

$$\mathbf{G} = \frac{\mathbf{MM}'}{2\sum p_i(1-p_i)},$$

onde $\mathbf{M}$ é a matriz de conteúdo genético com a dimensão de n animais genotipados por m SNPs (codificados como 0, 1 ou 2 para a ausência, presença de uma cópia, ou presença de duas cópias do alelo de referência) centrada com base na frequência alélicas e p_i e a frequência do alelo p .

2.3 Medidas de validação

A qualidade das predições foi avaliada utilizando os GEBVs gerados na população de validação, com base nas seguintes métricas: acurácia e viés de predição e dispersão dos GEBVs. A acurácia de predição foi obtida pela correlação de Pearson entre o valor genético verdadeiro (*True Breeding Value* - TBV) e o GEBV. O viés foi avaliado como a diferença entre a média do GEBV e do TBV. A dispersão dos GEBVs foi medida como 1 menos o coeficiente de regressão do TBV sobre o GEBV.

2.4 Cenários avaliados

Para cada um dos 8 cenários gerados com a simulação através do *software* QMSim (Sargolzaei; Schenkel, 2009), foram realizadas imputações pelo método que considerou família e população, gerando 8 cenários de imputação de genótipos. Em cada um desses cenários de imputação foram realizadas predições genômicas mascarando os fenótipos da geração 10, assim como nos cenários com genótipos reais (que não houve imputação), totalizando 16 cenários de avaliação. Os resultados da predição genômica dos cenários imputados e reais foram utilizados para avaliar e comparar a capacidade preditiva do modelo.

2.5 Comparação dos cenários

A comparação dos cenários foi realizada através de um teste t ao nível de 5% para comparação das médias ajustadas dos cenários em que não houve imputação com os cenários em que houve imputação (pares de média). O ajuste das médias no modelo foi feito incluindo os efeitos utilizados para gerar os cenários de imputação.

$$y_{ijklm} = \mu + ER_j + NE_k + PG_l + \varepsilon_{ijklm}$$

onde y_{ijklm} é a medida de avaliação da qualidade da predição (acurácia, viés e dispersão), em que ER_j é a estrutura de rebanho, NE_k é o número efetivo, PG_l é a porcentagem de animais genotipados e ε_{ijklm} é o efeito residual aleatório do modelo, onde $\varepsilon \sim N(0, \mathbf{I}\sigma_e^2)$.

As médias foram comparadas aos pares pelo test t entre os cenários reais e imputados. Para a comparação entre cenários imputados com imputados, foi utilizado o teste de tukey a 5%.

3 Resultados

3.1 Comparações entre cenários reais e imputados

Os resultados das comparações entre os cenários reais e imputados mostrou que não houve diferenças entre as médias dos cenários para as três métricas utilizadas, acurácia, viés e dispersão (Tabela 1).

Tabela 1. Medias e desvios-padrão (SD) da acurácia e viés de predição e dispersão dos GEBVs utilizando dados reais e imputados de ovinos obtidos a partir de simulação.

Cenários	Acurácia		Viés		Dispersão	
	Real (SD)	Imputado (SD)	Real (SD)	Imputado (SD)	Real (SD)	Imputado (SD)
POP1_CT20k100%300	0,8362 ^a (0,0219)	0,8360 ^a (0,0215)	-13,2304 ^a (1,0210)	-13,2610 ^a (1,0599)	-0,0404 ^a (0,0506)	-0,0404 ^a (0,0507)
POP1_CT1K5%300	0,8502 ^a (0,0230)	0,8500 ^a (0,0231)	-14,7642 ^a (0,8497)	-14,6858 ^a (0,9649)	-0,0938 ^a (0,0650)	-0,0970 ^a (0,0644)
POP2_CT20k100%300	0,5904 ^a (0,0249)	0,5902 ^a (0,0249)	-16,2276 ^a (0,9654)	-16,0294 ^a (0,9550)	-0,0414 ^a (0,1449)	-0,0436 ^a (0,1466)
POP2_CT1K5%300	0,6310 ^a (0,0351)	0,6304 ^a (0,0352)	-17,4894 ^a (1,1071)	-17,3802 ^a (1,1712)	-0,1794 ^a (0,1390)	-0,1818 ^a (0,1392)
POP1_ST20K100%80	0,4354 ^a (0,0618)	0,4360 ^a (0,0613)	-17,1986 ^a (1,5312)	-17,3188 ^a (1,4942)	0,2344 ^a (0,1250)	0,2358 ^a (0,1243)
POP1_ST1K5%80	0,4142 ^a (0,0907)	0,4140 ^a (0,0907)	-17,6626 ^a (1,3134)	-17,6120 ^a (1,3443)	0,0634 ^a (0,1576)	0,0634 ^a (0,1576)
POP2_ST20K100%300	0,4464 ^a (0,0330)	0,4466 ^a (0,0333)	-18,3120 ^a (0,7557)	-18,5096 ^a (0,7786)	0,1504 ^a (0,0367)	0,1482 ^a (0,0352)
POP2_ST1K5%300	0,4806 ^a (0,0318)	0,4804 ^a (0,0319)	-17,9350 ^a (0,5041)	-17,8986 ^a (0,5266)	-0,0172 ^a (0,0775)	-0,0164 ^a (0,0781)

*As comparações foram feitas nas linhas (horizontal)

* Valores com a mesma letra não diferem estatisticamente pelo teste t.

***POP_CT20K100%80/300:** População com troca de reprodutores, vinte mil animais, 100% genotipados, Ne=80 ou 300; **POP_CT1K5%80/300:** População com troca de reprodutores, mil animais, 5% genotipados, Ne=80 ou 300; **POP_ST20K100%300:** População sem troca de reprodutores, vinte mil animais 100% genotipados, Ne=300; **POP_ST1K5%300:** População sem troca de reprodutores, mil animais e 5% genotipados, Ne=300.

3.2 Comparações entre cenários imputados com imputados

Nas comparações entre cenários imputados houve diferenças entre os cenários tanto para medidas de acurácia, quanto para o viés e a dispersão (Tabela 2).

Tabela 2. Comparações dos cenários de predição em estudo com valores imputados utilizando acurácia, viés, dispersão e os desvios-padrão (SD) para cada métrica.

Cenários	Acurácia	Viés	Dispersão
POP1_CT20k100%300_imputado	0.8360 ^a (0,0215)	-13,2610 ^a (1,0599)	-0,0404 ^{bcd} (0,0507)
POP1_CT1K5%300_imputado	0.8500 ^a (0,0231)	-14.6858 ^{ab} (0,9649)	-0,0970 ^{cd} (0,06448)
POP2_CT20K100%300_imputado	0.5902 ^b (0,0249)	-16,0294 ^{bc} (0,9550)	-0,0436 ^{bcd} (0,1466)
POP2_CT1K5%300_imputado	0.6304 ^b (0,0352)	-17.3802 ^{cd} (1,1712)	-0,1818 ^d (0,1392)
POP1_ST20K100%80_imputado	0.4360 ^c (0,0613)	-17,3188 ^{cd} (1,4942)	0,2358 ^a (0,1243)
POP1_ST1K5%80_imputado	0.4140 ^c (0,0907)	-17,6120 ^{cd} (1,3443)	0.0634 ^{abcd} (0,1576)
POP2_ST20K100%300_imputado	0.4466 ^c (0,0333)	-18,5096 ^d (0,7786)	0,1482 ^{abc} (0,0352)
POP2_ST1K5%300_imputado	0.4804 ^c (0,0319)	-17,8986 ^{cd} (0,5266)	-0,0164 ^{bcd} (0,0781)

*As comparações foram feitas nas colunas (vertical)

* Valores com a mesma letra não diferem estatisticamente pelo teste de tukey a 5%

***POP_CT20K100%80/300:** População com troca de reprodutores, vinte mil animais, 100% genotipados, Ne=80 ou 300; **POP_CT1K5%80/300:** População com troca de reprodutores, mil animais, 5% genotipados, Ne=80 ou 300; **POP_ST20K100%300:** População sem troca de reprodutores, vinte mil animais 100% genotipados, Ne=300; **POP_ST1K5%300:** População sem troca de reprodutores, mil animais e 5%

As acurácias de predição foram melhores nas populações com troca de reprodutores (POP_CT) e piores nas populações sem troca (POP_ST). As populações com 50% de troca de reprodutores (POP1_CT) apresentaram maiores acurácias, acima de 0,80. Quando a troca foi de 5% (POP2_CT), observou-se resultados variando entre 0,5902 e 0,6304. O número de animais genotipados não influenciou nessa métrica, pois as populações com menor número de animais genotipados mostraram acurácias mais altas, embora, em média, tenham sido iguais às populações com maior número de animais genotipados. Todas as populações em que não houve troca de reprodutores apresentaram acurácia abaixo de 0,50, sendo os mais baixos (0.4360; 0,4041) nas populações com menor número efetivo (POP1_ST).

Houve tendência de aumento de viés nas populações sem troca de reprodutores. As populações com maior troca de reprodutores (POP1_CT) tiveram valores menores. Maiores médias (-17,8986; -18,5096) foram observadas nas POP2_ST e menores nas POP_CT (-

13,2610; -14.6858). O número efetivo não influenciou o viés, no entanto, a média foi menor (-13,2610) no cenário com maior número de animais genotipados (POP1_CT20k100%300).

A dispersão foi baixa em todos os cenários avaliados em geral. Houve diferença entre a POP1_ST20K100%80 e a POP2_ST1K5%300. sendo o valor mais alto (0,2358) observado na primeira, e mais baixo (-0,0164) na segunda.

4 Discussão

Neste estudo, foram investigados os impactos da imputação sobre a predição genômica, seguindo uma abordagem semelhante à adotada por Boloorma *et al.* (2015). As acurácias mínimas observadas nos cenários testados foram de 0,90, em conformidade com os valores estabelecidos por esses autores ao testar diferentes cenários preditivos.

Foram observadas maiores acurácias em populações com uma maior proporção de troca de reprodutores, em comparação com populações sem troca. A diversidade genética dessas populações pode ter contribuído para esse efeito. De modo similar, Al Kalaldehy *et al.* (2019) demonstraram que as acurácias das predições genômicas eram geralmente elevadas em populações mistas.

Em um estudo realizado por Brito *et al.* (2017) em populações de ovinos com estrutura multirracial foram verificadas melhores acurácias de predição em populações de treinamento mistas. Moghaddar *et al.* (2023) estudaram características de qualidade da carne em ovinos da raça Merino, comparando o uso de reprodutores multirraciais com genótipos comuns analisados por um chip de 50K com marcadores selecionados de uma população multirracial. Os autores concluíram que a acurácia da predição genômica para a característica de gordura de carcaça aumentou com o uso de reprodutores multirraciais. Em bovinos, Londono-Gil *et al.* (2024) compararam predições genômicas em raças mistas e únicas. Os autores concluíram que as predições realizadas com raças mistas apresentaram maior acurácia e reduziram os valores de dispersão utilizando o método ssGBLUP.

Níveis mais baixos de viés foram verificados nas populações com troca de reprodutores, em relação as populações sem troca, provavelmente a variabilidade genética pode ter contribuído para a melhoria dos valores. Alguns fatores de interferência na medida de viés já foram relatados. A interação genótipo x ambiente é uma das causas, quando não é considerada nas predições (Hayes *et al.*, 2009). Outra causa é o uso de uma amostra não representativa da população geral em estudos genômicos (Van Raden *et al.*, 2009), ou quando os padrões de LD diferem entre o estudo e a população-alvo (Goddard; Hayes, 2009). A utilização de um painel de referência não correspondente à ancestralidade da população estudada pode levar a

imputações e subseqüentes predições tendenciosas (Daetwyler *et al.*, 2014). Além disso, a amostra não aleatória de indivíduos para o estudo da população-alvo também pode influenciar o viés (Saatchi *et al.*, 2011), assim como imprecisões nas medições fenotípicas ou genotípicas (Wiggans *et al.*, 2011). Neste estudo, outra possível causa de interferência pode ter influenciado os resultados de viés, de modo que as frequências alélicas podem não estar de acordo com o esperado. Isso pode resultar em incompatibilidade entre o parentesco de pedigree e o genômico (Kijas *et al.*, 2012).

Características de baixa herdabilidade, como é o alvo deste trabalho, também podem apresentar vieses mais altos devido a variância genética explicar apenas uma pequena parte da variação fenotípica. Outros fatores que afetam essa medida são erros de amostragem de genótipos e pedigree (Jibrila *et al.*, 2021), que não se aplicam ao caso, por se tratar de um estudo simulado.

O viés de predição, utilizado na população de validação é definido como a diferença entre o GEBV e o TBV, valores próximos de zero correspondem a medidas ideais (Legarra; Reverter, 2018). Diferentes fatores podem interagir e resultar em aumento do viés nas predições genômicas. Em um trabalho de Nordbo *et al.* (2019), com dados reais da espécie bovina, foram relatadas altas estimativas de viés (0,768; 0,812) nos resultados de predição genômica. O estudo abordou os fatores que poderiam ter levado a esses resultados e discutiu como resolvê-los, sugerindo a inclusão de parâmetros com maior ou menor significado biológico. No entanto, os mesmos autores refletem que, devido as predições genômicas de etapa única dependerem de muitas suposições e parâmetros, identificar o motivo para a causa de valores elevados de viés não é algo simples, assim também como ajustá-los. Em estudos sobre predições genômicas em ovinos, os vieses calculados com a diferença entre as medias dos GEBVs preditos e observados (Legarra; Reverter *et al.*, 2018) ainda não são frequentemente observados (Araujo *et al.*, 2022). No entanto, duas pesquisas identificaram a presença de vieses em predições genômicas para ovelhas leiteiras (Macedo *et al.*, 2020; 2022).

Outros trabalhos, como o de Astruc *et al.* (2014) e Baloché *et al.* (2014) em ovinos relataram superestimação do valor genético com maior efeito nas características sob importante pressão de seleção por um método diferente (coeficiente de regressão linear), no entanto, nem sempre se conhece a origem desses vieses. Para tentar descobrir as causas para os altos valores de viés obtidos nesse estudo serão necessárias investigações futuras mais minuciosas.

As dispersões para todos os cenários avaliados foram baixas relativamente. Esses resultados são importantes, pois está associado à precisão dos valores preditos pelo modelo utilizado nas análises. O uso de informações genômicas através do uso de marcadores SNPS

pode ter contribuído para esse resultado, como também foi relatado por Araújo *et al.* (2022), que observaram valores de dispersão mais próximos de zero em modelos baseados na genômica em ovinos Rambouillet dos Estados Unidos.

A comparação entre cenários reais e imputados neste estudo não apresentou diferenças para as medidas de acurácia, viés e dispersão. Esse resultado é relevante, pois indica que a imputação do chip de 6k para o de 50K pode ser recomendada sem prejuízo nas predições genômicas. Dessa maneira, o produtor pode economizar com custos de genotipagem usando a estratégia de genotipar para uma densidade média e baixa e obter bons resultados.

Dentre os cenários imputados, os que proporcionaram melhores resultados foram os que continham maior número efetivo e maior proporção de troca de reprodutores. Provavelmente o maior número efetivo nas populações estudadas foram mais representativos, de maneira que as predições genômicas feitas para novos indivíduos possivelmente serão mais acuradas, por ter sido treinado em um conjunto de dados que refletem melhor a população-alvo. Não diferente, maior troca de reprodutores está intimamente relacionado a maior diversidade genética, que melhora as predições genômicas ao fornecer uma base de dados mais rica e variada, permitindo que os modelos capturem de maneira mais precisa a complexidade das influências genéticas sobre as características.

O sucesso da seleção genômica (SG) é multifatorial, depende da acurácia das predições genômicas, que é influenciada pela quantidade de marcadores, dos efetivos populacionais, do desequilíbrio de ligação (LD), da relação de parentesco entre a referência e a validação, da arquitetura genética da característica e da herdabilidade (Meuwissen; Hayes; Goddard, 2001). Neste estudo, a estrutura populacional gerada e a arquitetura genética da característica simulada podem ter influenciado nos valores discrepantes de viés. Desse modo, como esse parâmetro é uma preocupação atual em análises genômicas, vale a pena rever alguns pontos da simulação para ponderar os resultados obtidos. No entanto, o viés não deve ser considerado para tomada de decisão de forma isolada. Assim, os altos valores de acurácia obtidos nos cenários com troca e os baixos valores de dispersão alcançados são suficientes para guiar novos estudos a fim de testar tamanhos populacionais intermediários aos que foram estudados nesse trabalho.

5 Conclusão

Neste estudo, foram alcançadas altas acurácias de predição genômica utilizando cenários imputados a partir de um chip de 6k. Embora os elevados vieses observados precisem ser investigados mais detalhadamente, as informações geradas oferecem uma base valiosa para futuros trabalhos, com o objetivo de identificar e recomendar os melhores cenários para a

implementação da seleção genômica em populações ovinas com um número limitado de animais genotipados.

Referências

AGUILAR, *et al.* Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. **J Dairy Sci.**, United States, v. 93, n. 2, p. 743-752, Feb. 2010. DOI [10.3168/jds.2009-2730](https://doi.org/10.3168/jds.2009-2730). Disponível em: <https://pubmed.ncbi.nlm.nih.gov/20105546/>. Acesso em: 24 Jun. 2024.

AL KALALDEH *et al.* Using imputed whole-genome sequence data to improve the accuracy of genomic prediction for parasite resistance in Australian sheep. **Genetics Selection Evolution**, v. 51, p. 1-13, 2019. DOI: [10.1186/s12711-019-0476-4](https://doi.org/10.1186/s12711-019-0476-4) Disponível em: <https://link.springer.com/article/10.1186/s12711-019-0476-4>. Acesso em: 29 Jul. 2024.

ARAUJO, *et al.* A comprehensive comparison of haplotype-based single-step genomic predictions in livestock populations with different genetic diversity levels: a simulation study. **Frontiers in Genetics**, v. 12, p. 729867, 2021. DOI. [10.3389/fgene.2021.729867](https://doi.org/10.3389/fgene.2021.729867). Disponível em: <https://www.frontiersin.org/journals/genetics/articles/10.3389/fgene.2021.729867/full>. Acesso em: 20 Nov. 2023.

ARAUJO, *et al.* SNP-and haplotype-based single-step genomic predictions for body weight, wool, and reproductive traits in North American Rambouillet sheep. **Journal of Animal Breeding and Genetics**, v. 140, n. 2, p. 216-234, 2023. DOI: [10.1111/jbg.12748](https://doi.org/10.1111/jbg.12748). Disponível em: <https://onlinelibrary.wiley.com/doi/full/10.1111/jbg.12748>. Acesso em: 04 Ago. 2024.

ASTRUC, *et al.* Genomic evaluation validation test proposed by Interbull is necessary but not sufficient because it does not check the correct genetic trend. In: Proceedings of the 39th annual conference of the international committee for animal recording (ICAR). 2014. p. 19-23. Disponível em: https://www.icar.org/Documents/Berlin_2014/PPTs/Presented/Astruc.pdf. Acesso em: 01 Ago. 2024

BALOCHE, G. *et al.* Assessment of accuracy of genomic prediction for French Lacaune dairy sheep. **J Dairy Sci.**, Castanet-Tolosan, v. 97, n. 2, p. 1107-1116, Feb. 2014. DOI <https://doi.org/10.3168/jds.2013-7135>. Disponível em: <https://doi.org/10.3168/jds.2013-7135>. Acesso em: 4 Jun. 2024.

BLASCO, A.; TORO, M. A. A short critical history of the application of genomics to animal breeding. **Livest Sci**, Netherlands, v. 166, p. 4-9, Aug. 2014. DOI <https://doi.org/10.1016/j.livsci.2014.03.015>. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S187114131400167X>. Acesso em: 7 Jun. 2024.

BRITO, *et al.* Prediction of genomic breeding values for growth, carcass and meat quality traits in a multi-breed sheep population using a HD SNP chip. **BMC genetics**, v. 18, p. 1-17, 2017. DOI: [10.1186/s12863-017-0476-8](https://doi.org/10.1186/s12863-017-0476-8). Disponível em: <https://link.springer.com/article/10.1186/s12863-017-0476-8>. Acesso em: 27 Jul. 2024.

DAETWYLER, *et al.* Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. **Nature genetics**, v. 46, n. 8, p. 858-865, 2014. DOI:

10.1038/ng.3034. Disponível em: <https://www.nature.com/articles/ng.3034>. Acesso em 07 Ago. 2024

DAETWYLER, H. D. *et al.* Genomic prediction in animals and plants: simulation of data, validation, reporting, and benchmarking. **Genetics**, [S.l.], v. 193, n. 2, p. 347–365, Feb. 2013. DOI 10.1534/genetics.112.147983. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/23222650/>. Acesso em: 4 Jun. 2024.

GODDARD, Michael E.; HAYES, Ben J. Mapping genes for complex traits in domestic animals and their use in breeding programmes. **Nature Reviews Genetics**, v. 10, n. 6, p. 381–391, 2009. DOI: 10.1038/nrg2575. Disponível em: <https://www.nature.com/articles/nrg2575>. Acesso em 07 Ago. 2024.

GRANADO-TAJADA, I.; VARONA, L.; UGARTE, E. Genotyping strategies for maximizing genomic information in evaluations of the Latxa dairy sheep breed. **J Dairy Sci.**, United States, v. 104, n. 6, pág. 6861–6872, June. 2021. DOI <https://doi.org/10.3168/jds.2020-19978>. Disponível em: [https://www.journalofdairyscience.org/article/S0022-0302\(21\)00463-X/fulltext](https://www.journalofdairyscience.org/article/S0022-0302(21)00463-X/fulltext). Acesso em: 20 Jun. 2024.

HAYES, *et al.* Invited review: Genomic selection in dairy cattle: Progress and challenges. **Journal of dairy science**, v. 92, n. 2, p. 433–443, 2009. DOI: 10.3168/jds.2008-1646. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0022030209703479>. Acesso em: 04 Ago. 2024.

JIBRILA, Ibrahim *et al.* Avoiding preselection bias in subsequent single-step genomic BLUP evaluations of genomically preselected animals. **Journal of Animal Breeding and Genetics**, v. 138, n. 4, p. 432–441, 2021. DOI: [10.1111/jbg.12533](https://doi.org/10.1111/jbg.12533). Disponível em: <https://onlinelibrary.wiley.com/doi/full/10.1111/jbg.12533>. Acesso em: 02 Ago. 2024.

KARADAVUT, U. *et al.* A simulation study to determine the response of Morkaraman sheep

KIJAS, J. W. *et al.* Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. **PLoS Biol**, California, v. 10, n. 2, e1001258, Feb. 2012. DOI <https://doi.org/10.1371/journal.pbio.1001258>. Disponível em: <https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1001258>. Acesso em: 6 Jun. 2024.

LEGARRA, A.; AGUILAR, I.; MISZTAL, I. A relationship matrix including full pedigree and genomic information. **J Dairy Sci.**, United States, v. 92, p. 4656–4663, Sept. 2009. DOI [10.3168/jds.2009-2061](https://doi.org/10.3168/jds.2009-2061). Disponível em: <https://pubmed.ncbi.nlm.nih.gov/19700729/>. Acesso em: 24 Jun. 2024.

LEGARRA, A.; REVERTER, A. Semi-parametric estimates of population accuracy and bias of predictions of breeding values and future phenotypes using the LR method. **Genet Sel Evol**, France, v. 50, p. 1–18, Nov. 2018. DOI <https://doi.org/10.1186/s12711-018-0426-6>. Disponível em: <https://gsejournal.biomedcentral.com/articles/10.1186/s12711-018-0426-6>. Acesso em: 6 Jun. 2024.

LILLEHAMMER, M. *et al.* Genomic selection strategies to improve maternal traits in Norwegian White Sheep. **J Anim Breed Genet**, [S.l.], v. 137, n. 4, p. 384–394, Mar. 2020. DOI <https://doi.org/10.1111/jbg.12475>. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1111/jbg.12475>. Acesso em: 6 Jun. 2024.

LOBO, R. N. B. Opportunities for investment into small ruminant breeding programmes in Brazil. **J Anim Breed Genet**, [S.l.], v. 136, n. 5, p. 313-318, May. 2019. DOI <https://doi.org/10.1111/jbg.12396>. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1111/jbg.12396>. Acesso em: 6 Jun. 2024.

LONDOÑO-GIL, *et al.* Strategies for genomic predictions of an indicine multi-breed population using single-step GBLUP. *Journal of Animal Breeding and Genetics*, 2024. DOI: [10.1111/jbg.12882](https://doi.org/10.1111/jbg.12882). Disponível em: <https://onlinelibrary.wiley.com/doi/full/10.1111/jbg.12882>. Acesso em: 04 Ago. 2024.

MEUWISSEN, T.; HAYES, B. J.; GODDARD, M. E. Accelerating improvement of livestock with genomic selection. **Annu Rev Anim Biosci**, United States, v. 1, p. 221-237, Jan. 2013. DOI [10.1146/annurev-animal-031412-103705](https://doi.org/10.1146/annurev-animal-031412-103705). Disponível em: <https://www.annualreviews.org/content/journals/10.1146/annurev-animal-031412-103705>. Acesso em: 7 Jun. 2024.

MEUWISSEN, T.; HAYES, B. J.; GODDARD, M. E. Genomic selection: A paradigm shift in animal breeding. **Animal Front**, [S.l.], v.6, n.1, p. 6-14, Jan. 2016. DOI <https://doi.org/10.2527/af.2016-0002>. Disponível em: <https://academic.oup.com/af/article/6/1/6/4638797>. Acesso em: 6 Jun. 2024.

MEUWISSEN, T.; HAYES, B.J.; GODDARD, M. E. Prediction of total genetic value using genome-wide dense marker maps. **Genetics**, United States, v.157, p.1819-1829, Apr. 2001. DOI <https://doi.org/10.1093/genetics/157.4.1819>. Disponível em: <https://academic.oup.com/genetics/article/157/4/1819/6048353?login=false>. Acesso em: 7 Jun. 2024.

MISZTAL, I. *et al.* **Manual for BLUPF90 family of programs**. Athens, June. 2022. Disponível em: <https://nce.ads.uga.edu/wiki/doku.php?id=documentation>. Acesso em: Jan. 2021.

MISZTAL, I.; LEGARRA, A.; AGUILAR, I. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. **J Dairy Sci.**, United States, v. 92, n. 9, p. 4648–4655, Sept. 2009. DOI [10.3168/jds.2009-2064](https://doi.org/10.3168/jds.2009-2064). Disponível em: <https://pubmed.ncbi.nlm.nih.gov/19700728/>. Acesso em: 24 Jun. 2024.

MOGHADDAR, *et al.* Genomic prediction of purebred Merinos from an admixed terminal population using selected sequence variants. In: **Proceedings of 12th World Congress on Genetics Applied to Livestock Production (WCGALP) Technical and species orientated innovations in animal breeding, and contribution of genetics to solving societal challenges**. Wageningen Academic Publishers, 2022. p. 1110-1113. DOI: [10.3920/978-90-8686-940-4](https://doi.org/10.3920/978-90-8686-940-4) Disponível em: https://www.wageningenacademic.com/doi/abs/10.3920/978-90-8686-940-4_263. Acesso em: 04 Ago. 2024.

MRODE, R. *et al.* Genomic selection and use of molecular tools in breeding programs for indigenous and crossbred cattle in developing countries: current status and future prospects. **Front Genet**, Lausanne, v. 9, n. 694, Jan. 2019. DOI [10.3389/fgene.2018.00694](https://doi.org/10.3389/fgene.2018.00694). Disponível em: <https://pubmed.ncbi.nlm.nih.gov/30687382/>. Acesso em: 20 Jun. 2024.

MRODE, R. *et al.* Invited review: Genomic selection for small ruminants in developed countries: how applicable for the rest of the world? **Animal**, Cambridge, v. 12, n. 7, p. 1333-1340, Jul. 2018. DOI [10.1017/S1751731117003688](https://doi.org/10.1017/S1751731117003688). Disponível em: <https://pubmed.ncbi.nlm.nih.gov/29343308/>. Acesso em: 20 Jun. 2024.

NORDBO, *et al.* Level-biases in estimated breeding values due to the use of different SNP panels over time in ssGBLUP. *Genetics Selection Evolution*, v. 51, p. 1-8, 2019. DOI: [10.1186/s12711-019-0467-6](https://doi.org/10.1186/s12711-019-0467-6). Disponível em: <https://link.springer.com/article/10.1186/s12711-019-0517-z>. Acesso em: 02 Ago. 2024.

raised in central and eastern Anatolian to changing climate conditions. **J. Glob. Innov. Agric. Sci.**, [S.l.], v. 9, n. 4, p. 173-181, Dec. 2021. DOI [10.22194/JGIAS/9.959](https://doi.org/10.22194/JGIAS/9.959). Disponível em: https://www.researchgate.net/publication/357534736_A_SIMULATION_STUDY_TO_DETERMINE_THE_RESPONSE_OF_MORKARAMAN_SHEEP_RAISED_IN_CENTRAL_AND_EASTERN_ANATOLIAN_TO_CHANGING_CLIMATE_CONDITIONS. Acesso em: 20 Jun. 2024.

RUPP, R. *et al.* Genomic application in sheep and goat breeding. **Animal front.**, [S.l.], v. 6, n. 1, p. 39-44, Jan. 2016. DOI <https://doi.org/10.2527/af.2016-0006>. Disponível em: <https://academic.oup.com/af/article/6/1/39/4638792>. Acesso em: 6 Jun. 2024.

SAATCHI, *et al.* Accuracies of genomic breeding values in American Angus beef cattle using K-means clustering for cross-validation. **Genetics Selection Evolution**, v. 43, p. 1-16, 2011. DOI: [10.1186/1297-9686-43-40](https://doi.org/10.1186/1297-9686-43-40). Disponível em: <https://link.springer.com/article/10.1186/1297-9686-43-40>. Acesso em: 07 Ago. 2024

SHARMA, A. *et al.* Stories and Challenges of Genome Wide Association Studies in Livestock — A Review. **Asian-Australas J Anim Sci**, Korea, v. 28, n. 10, p. 1371-1379, Oct. 2015. DOI [10.5713/ajas.14.0715](https://doi.org/10.5713/ajas.14.0715). Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4554843/>. Acesso em: 7 Jun. 2024.

VANRADEN, P. M. Invited review: reliability of genomic predictions for North American Holstein bulls. **J Dairy Sci**, United States, v. 92, n. 1, p. 16-24, Jan. 2009. DOI [10.3168/jds.2008-1514](https://doi.org/10.3168/jds.2008-1514). Disponível em: [https://www.journalofdairyscience.org/article/S0022-0302\(09\)70305-4/fulltext](https://www.journalofdairyscience.org/article/S0022-0302(09)70305-4/fulltext). Acesso em: 7 Jun. 2024.

WIGGANS, George R. *et al.* Genomic selection in dairy cattle: the USDA experience. **Annual review of animal biosciences**, v. 5, n. 1, p. 309-327, 2017. DOI: [10.1146/annurev-animal-021815-111422](https://doi.org/10.1146/annurev-animal-021815-111422). Disponível em: <https://www.annualreviews.org/content/journals/10.1146/annurev-animal-021815-111422>. Acesso em: 07 Ago. 2024.

ANEXOS

**Cartão de parâmetros para simulação no *software* QMSim baseado no genoma ovino
com painel de moderada densidade e 4.000 QTLs**

```

/*****
**   Global parameters   **
*****/
title = "Ovinos herdabilidade 0.30";
nrep = 1;           //Number of replicates
h2 = 0.30;          //Heritability
qtlh2 = 0.24; //QTL heritability
phvar = 100; //Phenotypic variance
nthread = 1;
/*****
** Historical population **
*****/
begin_hp;
hg_size = 16000 [0] 10000 [1000] 12000 [1500]; //Size of the historical generations
nmfhg = 8000; //Number of males in the first generation
nmlhg = 4000; //Number of males in the last generation
end_hp;
/*****
** Expantion population**
*****/
begin_pop = "EG";
begin_founder;
  male [n = 80, pop = "hp"];
  female [n = 2400, pop = "hp"];
end_founder;
ng = 20;
ls = 1 2[0.50] 3[0.20];
md = rnd;
sr = 0.40;
dr = 0.20 0.1 /e;
sd = phen /h;
cd = age;
begin_popoutput;
stat;
ld /gen 0 20 /maft 0.05 /max_distance 25; // ##### pedir distancia maxima pra calcular o
LD ate 25 cM
end_popoutput;
end_pop;
/*****
**   Population 1 no exchange **
*****/
begin_pop = "POP1_ST";
begin_founder;

```

```

    male [n = 20, pop = "EG", gen = 20];
    female [n = 2400, pop = "EG", gen = 20];
end_founder;
ng = 10;
ls = 1 2[0.50] 3[0.20];
md = rnd;
sr = 0.20;
dr = 0.40; // ##### nao precisa de aumento exponencial nas populacoes que serao
estudadas, remover
sd = ebv /h;
ebv_est = blup;
cd = ebv /l;
begin_popoutput;
stat;
ld /gen 0 5 6 7 8 9 10 /maft 0.05 /max_distance 25; // ##### pedir distancia maxima pra
calcular o LD ate 25 cM ate nas ultimas 5 geracoes anteriores
data /gen 1 2 3 4 5 6 7 8 9 10;
genotype /snpcode /gen 1 2 3 4 5 6 7 8 9 10;
allele_freq /gen 1 2 3 4 5 6 7 8 9 10;
end_popoutput;
end_pop;
/*****
**    Population 2 no exchange **
*****/
begin_pop = "POP2_ST";
begin_founder;
    male [n = 80, pop = "EG", gen = 20] /not_founder_yet;
    female [n = 2400, pop = "EG", gen = 20];
end_founder;
ng = 10;
ls = 1 2[0.50] 3[0.20];
md = rnd;
sr = 0.20;
dr = 0.40; // ##### nao precisa de aumento exponencial nas populacoes que serao
estudadas, remover
sd = ebv /l;
ebv_est = blup;
cd = ebv /h;
begin_popoutput;
stat;
ld /gen 0 5 6 7 8 9 10 /maft 0.05 /max_distance 25; // ##### pedir distancia maxima pra
calcular o LD ate 25 cM ate nas ultimas 5 geracoes anteriores
data /gen 1 2 3 4 5 6 7 8 9 10;
genotype /snpcode /gen 1 2 3 4 5 6 7 8 9 10;
allele_freq /gen 1 2 3 4 5 6 7 8 9 10;
end_popoutput;
end_pop;
/*****
**    Population 3 no exchange **
*****/

```

```

begin_pop = "POP3_ST";
begin_founder;
  male [n = 20, pop = "EG", gen = 20] /not_founder_yet;
  female [n = 2400, pop = "EG", gen = 20];
end_founder;
ng = 10;
ls = 1 2[0.50] 3[0.20];
md = rnd;
sr = 0.20;
dr = 0.40; // ##### nao precisa de aumento exponencial nas populacoes que serao
estudadas, remover
sd = ebv /h;
ebv_est = blup;
cd = ebv /l;
begin_popoutput;
stat;
ld /gen 0 5 6 7 8 9 10 /maft 0.05 /max_distance 25; // ##### pedir distancia maxima pra
calcular o LD ate 25 cM ate nas ultimas 5 geracoes anteriores
data /gen 1 2 3 4 5 6 7 8 9 10;
genotype /snpcode /gen 1 2 3 4 5 6 7 8 9 10;
allele_freq /gen 1 2 3 4 5 6 7 8 9 10;
end_popoutput;
end_pop;
/*****
**   Population 4 no exchange **
*****/
begin_pop = "POP4_ST";
begin_founder;
  male [n = 80, pop = "EG", gen = 20] /not_founder_yet;
  female [n = 2400, pop = "EG", gen = 20];
end_founder;
ng = 10;
ls = 1 2[0.50] 3[0.20];
md = rnd;
sr = 0.20;
dr = 0.40; // ##### nao precisa de aumento exponencial nas populacoes que serao
estudadas, remover
sd = ebv /l;
ebv_est = blup;
cd = ebv /h;
begin_popoutput;
stat;
ld /gen 0 5 6 7 8 9 10 /maft 0.05 /max_distance 25; // ##### pedir distancia maxima pra
calcular o LD ate 25 cM ate nas ultimas 5 geracoes anteriores
data /gen 1 2 3 4 5 6 7 8 9 10;
genotype /snpcode /gen 1 2 3 4 5 6 7 8 9 10;
allele_freq /gen 1 2 3 4 5 6 7 8 9 10;
end_popoutput;
end_pop;
/*****

```

```

**    Population 5 no exchange **
*****/
begin_pop = "POP5_ST";
begin_founder;
  male [n = 20, pop = "EG", gen = 20] /not_founder_yet;
  female [n = 2400, pop = "EG", gen = 20];
end_founder;
ng = 10;
ls = 1 2[0.50] 3[0.20];
md = rnd;
sr = 0.20;
dr = 0.40; // ##### nao precisa de aumento exponencial nas populacoes que serao
estudadas, remover
sd = ebv /h;
ebv_est = blup;
cd = ebv /l;
begin_popoutput;
stat;
ld /gen 0 5 6 7 8 9 10 /maft 0.05 /max_distance 25; // ##### pedir distancia maxima pra
calcular o LD ate 25 cM ate nas ultimas 5 geracoes anteriores
data /gen 1 2 3 4 5 6 7 8 9 10;
genotype /snpcode /gen 1 2 3 4 5 6 7 8 9 10;
allele_freq /gen 1 2 3 4 5 6 7 8 9 10;
end_popoutput;
end_pop;

/*****/
**    Population 1 with 50% exchange**
*****/

/** Generation 0 **/ // ##### essa e a geracao 0, sempre comeca a partir da 0
begin_pop = "POP1_CT_G0"; // change the name of the population in the generation
begin_founder;
  male [n = 20, pop = "POP1_ST", gen = 0];
  male [n = 60, pop = "POP2_ST", gen = 0];
  female [n = 2400, pop = "POP1_ST", gen = 0];
end_founder;
ng = 1;
ls = 1 2[0.50] 3[0.20];
pmp = 0.45;
md = rnd;
sr = 0.20;
dr = 0.40;
sd = ebv /h;
ebv_est = blup;
cd = ebv /l;
begin_popoutput;
stat;
ld /gen 0 1 /maft 0.05 /max_distance 25;
data /gen 0 1;

```

```

    genotype /snpcode /gen 0 1;
    allele_freq /gen 0 1;
    end_popoutput;
    end_pop;

/** Generation 1 **/
begin_pop = "POP1_CT_G1"; // change the name of the population in the generation
begin_founder;
    male [n = 20, pop = "POP1_CT_G0", gen = 0, select = ebv /h];
    male [n = 20, pop = "POP1_CT_G0", gen = 1, select = ebv /h];
    male [n = 40, pop = "POP2_ST", gen = 1, select = ebv /h];
    female [n = 1200, pop = "POP1_CT_G0", gen = 0, select = ebv /h];
    female [n = 1200, pop = "POP1_CT_G0", gen = 1, select = ebv /h];
end_founder;
ng = 1;
ls = 1 2[0.50] 3[0.20];
pmp = 0.45;
md = rnd;
sr = 0.20;
dr = 0.40;
sd = ebv /h;
ebv_est = blup;
cd = ebv /l;
begin_popoutput;
stat;
ld /gen 1 /maft 0.05 /max_distance 25;
data /gen 1;
    genotype /snpcode /gen 1;
    allele_freq /gen 1;
    end_popoutput;
    end_pop;

/** Generation 2**/
begin_pop = "POP1_CT_G2";
begin_founder;
    male [n = 20, pop = "POP1_CT_G1", gen = 0, select = ebv /h];
    male [n = 20, pop = "POP1_CT_G1", gen = 1, select = ebv /h];
    male [n = 40, pop = "POP2_ST", gen = 2, select = ebv /h];
    female [n = 1200, pop = "POP1_CT_G1", gen = 0, select = ebv /h];
    female [n = 1200, pop = "POP1_CT_G1", gen = 1, select = ebv /h];
end_founder;
ng = 1;
ls = 1 2[0.50] 3[0.20];
pmp = 0.45;
md = rnd;
sr = 0.20;
dr = 0.40;
sd = ebv /h;
ebv_est = blup;
cd = ebv /l;

```

```

begin_popoutput;
stat;
ld /gen 1 /maft 0.05 /max_distance 25;
data /gen 1;
genotype /snpcode /gen 1;
allele_freq /gen 1;
end_popoutput;
end_pop;

/** Generation 3**/
begin_pop = "POP1_CT_G3";
begin_founder;
  male [n = 20, pop = "POP1_CT_G2", gen = 0, select = ebv /h];
  male [n = 20, pop = "POP1_CT_G2", gen = 1, select = ebv /h];
  male [n = 40, pop = "POP2_ST", gen = 3, select = ebv /h];
  female [n = 1200, pop = "POP1_CT_G2", gen = 0, select = ebv /h];
  female [n = 1200, pop = "POP1_CT_G2", gen = 1, select = ebv /h];
end_founder;
ng = 1;
ls = 1 2[0.50] 3[0.20];
pmp = 0.45;
md = rnd;
sr = 0.20;
dr = 0.40;
sd = ebv /h;
ebv_est = blup;
cd = ebv /l;
begin_popoutput;
stat;
ld /gen 1 /maft 0.05 /max_distance 25;
data /gen 1;
genotype /snpcode /gen 1;
allele_freq /gen 1;
end_popoutput;
end_pop;

/** Generation 4**/
begin_pop = "POP1_CT_G4";
begin_founder;
  male [n = 20, pop = "POP1_CT_G3", gen = 0, select = ebv /h];
  male [n = 20, pop = "POP1_CT_G3", gen = 1, select = ebv /h];
  male [n = 40, pop = "POP2_ST", gen = 4, select = ebv /h];
  female [n = 1200, pop = "POP1_CT_G3", gen = 0, select = ebv /h];
  female [n = 1200, pop = "POP1_CT_G3", gen = 1, select = ebv /h];
end_founder;
ng = 1;
ls = 1 2[0.50] 3[0.20];
pmp = 0.45;
md = rnd;
sr = 0.20;

```

```

dr = 0.40;
sd = ebv /h;
ebv_est = blup;
cd = ebv /l;
begin_popoutput;
stat;
ld /gen 1 /maft 0.05 /max_distance 25;
data /gen 1;
genotype /snp_code /gen 1;
allele_freq /gen 1;
end_popoutput;
end_pop;

/** Generation 5**/
begin_pop = "POP1_CT_G5";
begin_founder;
male [n = 20, pop = "POP1_CT_G4", gen = 0, select = ebv /h];
male [n = 20, pop = "POP1_CT_G4", gen = 1, select = ebv /h];
male [n = 40, pop = "POP2_ST", gen = 5, select = ebv /h];
female [n = 1200, pop = "POP1_CT_G4", gen = 0, select = ebv /h];
female [n = 1200, pop = "POP1_CT_G4", gen = 1, select = ebv /h];
end_founder;
ng = 1;
ls = 1 2[0.50] 3[0.20];
pmp = 0.45;
md = rnd;
sr = 0.20;
dr = 0.40;
sd = ebv /h;
ebv_est = blup;
cd = ebv /l;
begin_popoutput;
stat;
ld /gen 1 /maft 0.05 /max_distance 25;
data /gen 1;
genotype /snp_code /gen 1;
allele_freq /gen 1;
end_popoutput;
end_pop;

/** Generation 6**/
begin_pop = "POP1_CT_G6";
begin_founder;
male [n = 20, pop = "POP1_CT_G5", gen = 0, select = ebv /h];
male [n = 20, pop = "POP1_CT_G5", gen = 1, select = ebv /h];
male [n = 40, pop = "POP2_ST", gen = 6, select = ebv /h];
female [n = 1200, pop = "POP1_CT_G5", gen = 0, select = ebv /h];
female [n = 1200, pop = "POP1_CT_G5", gen = 1, select = ebv /h];
end_founder;
ng = 1;

```

```

ls = 1 2[0.50] 3[0.20];
pmp = 0.45;
md = rnd;
sr = 0.20;
dr = 0.40;
sd = ebv /h;
ebv_est = blup;
cd = ebv /l;
begin_popoutput;
stat;
ld /gen 1 /maft 0.05 /max_distance 25;
data /gen 1;
genotype /snp_code /gen 1;
allele_freq /gen 1;
end_popoutput;
end_pop;

/** Generation 7**/
begin_pop = "POP1_CT_G7";
begin_founder;
male [n = 20, pop = "POP1_CT_G6", gen = 0, select = ebv /h];
male [n = 20, pop = "POP1_CT_G6", gen = 1, select = ebv /h];
male [n = 40, pop = "POP2_ST", gen = 7, select = ebv /h];
female [n = 1200, pop = "POP1_CT_G6", gen = 0, select = ebv /h];
female [n = 1200, pop = "POP1_CT_G6", gen = 1, select = ebv /h];
end_founder;
ng = 1;
ls = 1 2[0.50] 3[0.20];
pmp = 0.45;
md = rnd;
sr = 0.20;
dr = 0.40;
sd = ebv /h;
ebv_est = blup;
cd = ebv /l;
begin_popoutput;
stat;
ld /gen 1 /maft 0.05 /max_distance 25;
data /gen 1;
genotype /snp_code /gen 1;
allele_freq /gen 1;
end_popoutput;
end_pop;

/** Generation 8**/
begin_pop = "POP1_CT_G8";
begin_founder;
male [n = 20, pop = "POP1_CT_G7", gen = 0, select = ebv /h];
male [n = 20, pop = "POP1_CT_G7", gen = 1, select = ebv /h];
male [n = 40, pop = "POP2_ST", gen = 8, select = ebv /h];

```

```

    female [n = 1200, pop = "POP1_CT_G7", gen = 0, select = ebv /h];
    female [n = 1200, pop = "POP1_CT_G7", gen = 1, select = ebv /h];
end_founder;
ng = 1;
ls = 1 2[0.50] 3[0.20];
pmp = 0.45;
md = rnd;
sr = 0.20;
dr = 0.40;
sd = ebv /h;
ebv_est = blup;
cd = ebv /l;
begin_popoutput;
stat;
ld /gen 1 /maft 0.05 /max_distance 25;
data /gen 1;
genotype /snp_code /gen 1;
allele_freq /gen 1;
end_popoutput;
end_pop;

/** Generation 9**/
begin_pop = "POP1_CT_G9";
begin_founder;
    male [n = 20, pop = "POP1_CT_G8", gen = 0, select = ebv /h];
    male [n = 20, pop = "POP1_CT_G8", gen = 1, select = ebv /h];
    male [n = 40, pop = "POP2_ST", gen = 9, select = ebv /h];
    female [n = 1200, pop = "POP1_CT_G8", gen = 0, select = ebv /h];
    female [n = 1200, pop = "POP1_CT_G8", gen = 1, select = ebv /h];
end_founder;
ng = 1;
ls = 1 2[0.50] 3[0.20];
pmp = 0.45;
md = rnd;
sr = 0.20;
dr = 0.40;
sd = ebv /h;
ebv_est = blup;
cd = ebv /l;
begin_popoutput;
stat;
ld /gen 1 /maft 0.05 /max_distance 25;
data /gen 1;
genotype /snp_code /gen 1;
allele_freq /gen 1;
end_popoutput;
end_pop;

/** Generation 10**/

```

```

begin_pop = "POP1_CT_G10";
begin_founder;
  male [n = 20, pop = "POP1_CT_G9", gen = 0, select = ebv /h];
  male [n = 20, pop = "POP1_CT_G9", gen = 1, select = ebv /h];
  male [n = 40, pop = "POP2_ST", gen = 10, select = ebv /h];
  female [n = 1200, pop = "POP1_CT_G9", gen = 0, select = ebv /h];
  female [n = 1200, pop = "POP1_CT_G9", gen = 1, select = ebv /h];
end_founder;
ng = 1;
ls = 1 2[0.50] 3[0.20];
pmp = 0.45;
md = rnd;
sr = 0.20;
dr = 0.40;
sd = ebv /h;
ebv_est = blup;
cd = ebv /l;
begin_popoutput;
stat;
ld /gen 1 /maft 0.05 /max_distance 25;
data /gen 1;
genotype /snp_code /gen 1;
allele_freq /gen 1;
end_popoutput;
end_pop;

/*****
**   Population 1 with 5% exchange**
*****/

/** Generation_0-_5 **/ // ##### essa e a geracao 0, sempre comeca a partir da 0
begin_pop = "POP1_CT_G0_5"; // change the name of the population in the generation
begin_founder;
  male [n = 4, pop = "POP1_ST", gen = 0];
  male [n = 76, pop = "POP2_ST", gen = 0];
  female [n = 2400, pop = "POP1_ST", gen = 0];
end_founder;
ng = 1;
ls = 1 2[0.50] 3[0.20];
pmp = 0.45;
md = rnd;
sr = 0.20;
dr = 0.40;
sd = ebv /h;
ebv_est = blup;
cd = ebv /l;
begin_popoutput;
stat;
ld /gen 0 1 /maft 0.05 /max_distance 25;
data /gen 0 1;

```

```

    genotype /snp_code /gen 0 1;
    allele_freq /gen 0 1;
    end_popoutput;
    end_pop;

```

```

/** Generation_1_5**/
begin_pop = "POP1_CT_G1_5";
begin_founder;
    male [n = 38, pop = "POP1_CT_G0_5", gen = 0, select = ebv /h];
    male [n = 38, pop = "POP1_CT_G0_5", gen = 1, select = ebv /h];
    male [n = 4, pop = "POP2_ST", gen = 1, select = ebv /h];
    female [n = 1200, pop = "POP1_CT_G0_5", gen = 0, select = ebv /h];
    female [n = 1200, pop = "POP1_CT_G0_5", gen = 1, select = ebv /h];
end_founder;
ng = 1;
ls = 1 2[0.50] 3[0.20];
pmp = 0.45;
md = rnd;
sr = 0.20;
dr = 0.40;
sd = ebv /h;
ebv_est = blup;
cd = ebv /l;
begin_popoutput;
stat;
ld /gen 1 /maft 0.05 /max_distance 25;
data /gen 1;
genotype /snp_code /gen 1;
allele_freq /gen 1;
end_popoutput;
end_pop;

```

```

/** Generation 2_5**/
begin_pop = "POP1_CT_G2_5";
begin_founder;
    male [n = 38, pop = "POP1_CT_G1_5", gen = 0, select = ebv /h];
    male [n = 38, pop = "POP1_CT_G1_5", gen = 1, select = ebv /h];
    male [n = 4, pop = "POP2_ST", gen = 2, select = ebv /h];
    female [n = 1200, pop = "POP1_CT_G1_5", gen = 0, select = ebv /h];
    female [n = 1200, pop = "POP1_CT_G1_5", gen = 1, select = ebv /h];
end_founder;
ng = 1;
ls = 1 2[0.50] 3[0.20];
pmp = 0.45;
md = rnd;
sr = 0.20;
dr = 0.40;
sd = ebv /h;
ebv_est = blup;

```

```

cd = ebv /l;
begin_popoutput;
stat;
ld /gen 1 /maft 0.05 /max_distance 25;
data /gen 1;
genotype /snpcode /gen 1;
allele_freq /gen 1;
end_popoutput;
end_pop;

/** Generation 3_5**/
begin_pop = "POP1_CT_G3_5";
begin_founder;
  male [n = 38, pop = "POP1_CT_G2_5", gen = 0, select = ebv /h];
  male [n = 38, pop = "POP1_CT_G2_5", gen = 1, select = ebv /h];
  male [n = 4, pop = "POP2_ST", gen = 3, select = ebv /h];
  female [n = 1200, pop = "POP1_CT_G2_5", gen = 0, select = ebv /h];
  female [n = 1200, pop = "POP1_CT_G2_5", gen = 1, select = ebv /h];
end_founder;
ng = 1;
ls = 1 2[0.50] 3[0.20];
pmp = 0.45;
md = rnd;
sr = 0.20;
dr = 0.40;
sd = ebv /h;
ebv_est = blup;
cd = ebv /l;
begin_popoutput;
stat;
ld /gen 1 /maft 0.05 /max_distance 25;
data /gen 1;
genotype /snpcode /gen 1;
allele_freq /gen 1;
end_popoutput;
end_pop;

/** Generation 4_5**/
begin_pop = "POP1_CT_G4_5";
begin_founder;
  male [n = 38, pop = "POP1_CT_G3_5", gen = 0, select = ebv /h];
  male [n = 38, pop = "POP1_CT_G3_5", gen = 1, select = ebv /h];
  male [n = 4, pop = "POP2_ST", gen = 4, select = ebv /h];
  female [n = 1200, pop = "POP1_CT_G3_5", gen = 0, select = ebv /h];
  female [n = 1200, pop = "POP1_CT_G3_5", gen = 1, select = ebv /h];
end_founder;
ng = 1;
ls = 1 2[0.50] 3[0.20];
pmp = 0.45;
md = rnd;

```

```

sr = 0.20;
dr = 0.40;
sd = ebv /h;
ebv_est = blup;
cd = ebv /l;
begin_popoutput;
stat;
ld /gen 1 /maft 0.05 /max_distance 25;
data /gen 1;
genotype /snpcode /gen 1;
allele_freq /gen 1;
end_popoutput;
end_pop;

/** Generation 5_5**/
begin_pop = "POP1_CT_G5_5";
begin_founder;
male [n = 38, pop = "POP1_CT_G4_5", gen = 0, select = ebv /h];
male [n = 38, pop = "POP1_CT_G4_5", gen = 1, select = ebv /h];
male [n = 4, pop = "POP2_ST", gen = 5, select = ebv /h];
female [n = 1200, pop = "POP1_CT_G4_5", gen = 0, select = ebv /h];
female [n = 1200, pop = "POP1_CT_G4_5", gen = 1, select = ebv /h];
end_founder;
ng = 1;
ls = 1 2[0.50] 3[0.20];
pmp = 0.45;
md = rnd;
sr = 0.20;
dr = 0.40;
sd = ebv /h;
ebv_est = blup;
cd = ebv /l;
begin_popoutput;
stat;
ld /gen 1 /maft 0.05 /max_distance 25;
data /gen 1;
genotype /snpcode /gen 1;
allele_freq /gen 1;
end_popoutput;
end_pop;

/** Generation 6_5**/
begin_pop = "POP1_CT_G6_5";
begin_founder;
male [n = 38, pop = "POP1_CT_G5_5", gen = 0, select = ebv /h];
male [n = 38, pop = "POP1_CT_G5_5", gen = 1, select = ebv /h];
male [n = 4, pop = "POP2_ST", gen = 6, select = ebv /h];
female [n = 1200, pop = "POP1_CT_G5_5", gen = 0, select = ebv /h];
female [n = 1200, pop = "POP1_CT_G5_5", gen = 1, select = ebv /h];
end_founder;

```

```

ng = 1;
ls = 1 2[0.50] 3[0.20];
pmp = 0.45;
md = rnd;
sr = 0.20;
dr = 0.40;
sd = ebv /h;
ebv_est = blup;
cd = ebv /l;
begin_popoutput;
stat;
ld /gen 1 /maft 0.05 /max_distance 25;
data /gen 1;
genotype /snpcode /gen 1;
allele_freq /gen 1;
end_popoutput;
end_pop;

/** Generation 7_5**/
begin_pop = "POP1_CT_G7_5";
begin_founder;
  male [n = 38, pop = "POP1_CT_G6_5", gen = 0, select = ebv /h];
  male [n = 38, pop = "POP1_CT_G6_5", gen = 1, select = ebv /h];
  male [n = 4, pop = "POP2_ST", gen = 7, select = ebv /h];
  female [n = 1200, pop = "POP1_CT_G6_5", gen = 0, select = ebv /h];
  female [n = 1200, pop = "POP1_CT_G6_5", gen = 1, select = ebv /h];
end_founder;
ng = 1;
ls = 1 2[0.50] 3[0.20];
pmp = 0.45;
md = rnd;
sr = 0.20;
dr = 0.40;
sd = ebv /h;
ebv_est = blup;
cd = ebv /l;
begin_popoutput;
stat;
ld /gen 1 /maft 0.05 /max_distance 25;
data /gen 1;
genotype /snpcode /gen 1;
allele_freq /gen 1;
end_popoutput;
end_pop;

/** Generation 8_5**/
begin_pop = "POP1_CT_G8_5";
begin_founder;
  male [n = 38, pop = "POP1_CT_G7_5", gen = 0, select = ebv /h];
  male [n = 38, pop = "POP1_CT_G7_5", gen = 1, select = ebv /h];

```

```

    male [n = 4, pop = "POP2_ST", gen = 8, select = ebv /h];
    female [n = 1200, pop = "POP1_CT_G7_5", gen = 0, select = ebv /h];
    female [n = 1200, pop = "POP1_CT_G7_5", gen = 1, select = ebv /h];
end_founder;
ng = 1;
ls = 1 2[0.50] 3[0.20];
pmp = 0.45;
md = rnd;
sr = 0.20;
dr = 0.40;
sd = ebv /h;
ebv_est = blup;
cd = ebv /l;
begin_popoutput;
stat;
ld /gen 1 /maft 0.05 /max_distance 25;
data /gen 1;
genotype /snp_code /gen 1;
allele_freq /gen 1;
end_popoutput;
end_pop;

/** Generation 9_5**/
begin_pop = "POP1_CT_G9_5";
begin_founder;
    male [n = 38, pop = "POP1_CT_G8_5", gen = 0, select = ebv /h];
    male [n = 38, pop = "POP1_CT_G8_5", gen = 1, select = ebv /h];
    male [n = 4, pop = "POP2_ST", gen = 9, select = ebv /h];
    female [n = 1200, pop = "POP1_CT_G8_5", gen = 0, select = ebv /h];
    female [n = 1200, pop = "POP1_CT_G8_5", gen = 1, select = ebv /h];
end_founder;
ng = 1;
ls = 1 2[0.50] 3[0.20];
pmp = 0.45;
md = rnd;
sr = 0.20;
dr = 0.40;
sd = ebv /h;
ebv_est = blup;
cd = ebv /l;
begin_popoutput;
stat;
ld /gen 1 /maft 0.05 /max_distance 25;
data /gen 1;
genotype /snp_code /gen 1;
allele_freq /gen 1;
end_popoutput;
end_pop;

/** Generation 10_5**/

```

```

begin_pop = "POP1_CT_G10_5";
begin_founder;
  male [n = 38, pop = "POP1_CT_G9_5", gen = 0, select = ebv /h];
  male [n = 38, pop = "POP1_CT_G9_5", gen = 1, select = ebv /h];
  male [n = 4, pop = "POP2_ST", gen = 10, select = ebv /h];
  female [n = 1200, pop = "POP1_CT_G9_5", gen = 0, select = ebv /h];
  female [n = 1200, pop = "POP1_CT_G9_5", gen = 1, select = ebv /h];
end_founder;
ng = 1;
ls = 1 2[0.50] 3[0.20];
pmp = 0.45;
md = rnd;
sr = 0.20;
dr = 0.40;
sd = ebv /h;
ebv_est = blup;
cd = ebv /l;
begin_popoutput;
stat;
ld /gen 1 /maft 0.05 /max_distance 25;
data /gen 1;
genotype /snpcode /gen 1;
allele_freq /gen 1;
end_popoutput;
end_pop;

/*****
** Genome **
*****/
begin_genome;
  begin_chr = 1; // CHR 1 //number of chromosomes with the same parameter
    chrln = 301; //Chromosome length cM
    nmloci = 5244; //Number of markers
    mpos = even; //Marker positions
    nma = all 2; //Number of marker alleles
    maf = eql; //Marker allele frequencies
    nqloci = 453; //Number of QTL
    qpos = rnd; //QTL positions
    nqa = rnd 2 3 4 5 6 7; //Number of QTL alleles
    qaf = eql; //QTL allele frequencies
    qae = rndg 0.4; //QTL allele effects ~ Gamma
  end_chr;
//
  begin_chr = 1; // CHR 2 //number of chromosomes with the same parameter
    chrln = 266; //Chromosome length cM
    nmloci = 4970; //Number of markers
    mpos = even; //Marker positions
    nma = all 2; //Number of marker alleles
    maf = eql; //Marker allele frequencies
    nqloci = 400; //Number of QTL

```

```

qpos = rnd;           //QTL positions
nqa = rnd 2 3 4 5 6 7;           //Number of QTL alleles
qaf = eql;           //QTL allele frequencies
qae = rndg 0.4;       //QTL allele effects ~ Gamma
end_chr;
//
begin_chr = 1; // CHR 3 //number of chromosomes with the same parameter
chrln = 241;           //Chromosome length cM
nmloci = 4465;         //Number of markers
mpos = even;          //Marker positions
nma = all 2;           //Number of marker alleles
maf = eql;            //Marker allele frequencies
nqloci = 363;          //Number of QTL
qpos = rnd;
nqa = rnd 2 3 4 5 6 7;           //Number of QTL alleles
qaf = eql;            //QTL allele frequencies
qae = rndg 0.4;       //QTL allele effects ~ Gamma
end_chr;
//
begin_chr = 1; // CHR 4 //number of chromosomes with the same parameter
chrln = 130;           //Chromosome length cM
nmloci = 2436;         //Number of markers
mpos = even;          //Marker positions
nma = all 2;           //Number of marker alleles
maf = eql;            //Marker allele frequencies
nqloci = 197;          //Number of QTL
qpos = rnd;
nqa = rnd 2 3 4 5 6 7;           //Number of QTL alleles
qaf = eql;            //QTL allele frequencies
qae = rndg 0.4;       //QTL allele effects ~ Gamma
end_chr;
//
begin_chr = 1; // CHR 5 //number of chromosomes with the same parameter
chrln = 118;           //Chromosome length cM
nmloci = 2109;         //Number of markers
mpos = even;          //Marker positions
nma = all 2;           //Number of marker alleles
maf = eql;            //Marker allele frequencies
nqloci = 178;          //Number of QTL
qpos = rnd;
nqa = rnd 2 3 4 5 6 7;           //Number of QTL alleles
qaf = eql;            //QTL allele frequencies
qae = rndg 0.4;       //QTL allele effects ~ Gamma
end_chr;
//
begin_chr = 1; // CHR 6 //number of chromosomes with the same parameter
chrln = 130;           //Chromosome length cM
nmloci = 2318;         //Number of markers
mpos = even;          //Marker positions
nma = all 2;           //Number of marker alleles

```

```

maf = eql; //Marker allele frequencies
nqloci = 196; //Number of QTL
qpos = rnd;
nqa = rnd 2 3 4 5 6 7; //Number of QTL alleles
qaf = eql; //QTL allele frequencies
qae = rndg 0.4; //QTL allele effects ~ Gamma
end_chr;
//
begin_chr = 1; // CHR 7 //number of chromosomes with the same parameter
chrln = 108; //Chromosome length cM
nmloci = 2005; //Number of markers
mpos = even; //Marker positions
nma = all 2; //Number of marker alleles
maf = eql; //Marker allele frequencies
nqloci = 162; //Number of QTL
qpos = rnd;
nqa = rnd 2 3 4 5 6 7; //Number of QTL alleles
qaf = eql; //QTL allele frequencies
qae = rndg 0.4; //QTL allele effects ~ Gamma
end_chr;
//
begin_chr = 1; // CHR 8 //number of chromosomes with the same parameter
chrln = 99; //Chromosome length cM
nmloci = 1851; //Number of markers
mpos = even; //Marker positions
nma = all 2; //Number of marker alleles
maf = eql; //Marker allele frequencies
nqloci = 149; //Number of QTL
qpos = rnd;
nqa = rnd 2 3 4 5 6 7; //Number of QTL alleles
qaf = eql; //QTL allele frequencies
qae = rndg 0.4; //QTL allele effects ~ Gamma
end_chr;
//
begin_chr = 1; // CHR 9 //number of chromosomes with the same parameter
chrln = 105; //Chromosome length cM
nmloci = 1901; //Number of markers
mpos = even; //Marker positions
nma = all 2; //Number of marker alleles
maf = eql; //Marker allele frequencies
nqloci = 158; //Number of QTL
qpos = rnd;
nqa = rnd 2 3 4 5 6 7; //Number of QTL alleles
qaf = eql; //QTL allele frequencies
qae = rndg 0.4; //QTL allele effects ~ Gamma
end_chr;
//
begin_chr = 1; // CHR 10 //number of chromosomes with the same parameter
chrln = 97; //Chromosome length cM
nmloci = 1649; //Number of markers

```

```

mpos = even;          //Marker positions
nma  = all 2;          //Number of marker alleles
maf  = eql;           //Marker allele frequencies
nqloci = 146;          //Number of QTL
qpos = rnd;
nqa  = rnd 2 3 4 5 6 7; //Number of QTL alleles
qaf  = eql;           //QTL allele frequencies
qae  = rndg 0.4;       //QTL allele effects ~ Gamma
end_chr;
//
begin_chr = 1; // CHR 11 //number of chromosomes with the same parameter
chrln = 61;      //Chromosome length cM
nmloci = 1077;    //Number of markers
mpos = even;      //Marker positions
nma  = all 2;      //Number of marker alleles
maf  = eql;        //Marker allele frequencies
nqloci = 92;       //Number of QTL
qpos = rnd;
nqa  = rnd 2 3 4 5 6 7; //Number of QTL alleles
qaf  = eql;        //QTL allele frequencies
qae  = rndg 0.4;    //QTL allele effects ~ Gamma
end_chr;
//
begin_chr = 1; // CHR 12 //number of chromosomes with the same parameter
chrln = 85;      //Chromosome length cM
nmloci = 1528;    //Number of markers
mpos = even;      //Marker positions
nma  = all 2;      //Number of marker alleles
maf  = eql;        //Marker allele frequencies
nqloci = 128;     //Number of QTL
qpos = rnd;
nqa  = rnd 2 3 4 5 6 7; //Number of QTL alleles
qaf  = eql;        //QTL allele frequencies
qae  = rndg 0.4;    //QTL allele effects ~ Gamma
end_chr;
//
begin_chr = 1; // CHR 13 //number of chromosomes with the same parameter
chrln = 87;      //Chromosome length cM
nmloci = 1527;    //Number of markers
mpos = even;      //Marker positions
nma  = all 2;      //Number of marker alleles
maf  = eql;        //Marker allele frequencies
nqloci = 131;     //Number of QTL
qpos = rnd;
nqa  = rnd 2 3 4 5 6 7; //Number of QTL alleles
qaf  = eql;        //QTL allele frequencies
qae  = rndg 0.4;    //QTL allele effects ~ Gamma
end_chr;
//
begin_chr = 1; // CHR 14 //number of chromosomes with the same parameter

```

```

chrln = 71;                //Chromosome length cM
nmloci = 1045;             //Number of markers
mpos = even;              //Marker positions
nma = all 2;              //Number of marker alleles
maf = eql;                //Marker allele frequencies
nqloci = 107;             //Number of QTL
qpos = rnd;
nqa = rnd 2 3 4 5 6 7;    //Number of QTL alleles
qaf = eql;                //QTL allele frequencies
qae = rndg 0.4;           //QTL allele effects ~ Gamma
end_chr;

//
begin_chr = 1; // CHR 15 //number of chromosomes with the same parameter
chrln = 90;                //Chromosome length cM
nmloci = 1506;             //Number of markers
mpos = even;              //Marker positions
nma = all 2;              //Number of marker alleles
maf = eql;                //Marker allele frequencies
nqloci = 135;             //Number of QTL
qpos = rnd;
nqa = rnd 2 3 4 5 6 7;    //Number of QTL alleles
qaf = eql;                //QTL allele frequencies
qae = rndg 0.4;           //QTL allele effects ~ Gamma
end_chr;

//
begin_chr = 1; // CHR 16 //number of chromosomes with the same parameter
chrln = 78;                //Chromosome length cM
nmloci = 1395;             //Number of markers
mpos = even;              //Marker positions
nma = all 2;              //Number of marker alleles
maf = eql;                //Marker allele frequencies
nqloci = 117;             //Number of QTL
qpos = rnd;
nqa = rnd 2 3 4 5 6 7;    //Number of QTL alleles
qaf = eql;                //QTL allele frequencies
qae = rndg 0.4;           //QTL allele effects ~ Gamma
end_chr;

//
begin_chr = 1; // CHR 17 //number of chromosomes with the same parameter
chrln = 83;                //Chromosome length cM
nmloci = 1291;             //Number of markers
mpos = even;              //Marker positions
nma = all 2;              //Number of marker alleles
maf = eql;                //Marker allele frequencies
nqloci = 125;             //Number of QTL
qpos = rnd;
nqa = rnd 2 3 4 5 6 7;    //Number of QTL alleles
qaf = eql;                //QTL allele frequencies
qae = rndg 0.4;           //QTL allele effects ~ Gamma
end_chr;

```

```

//
begin_chr = 1; // CHR 18 //number of chromosomes with the same parameter
chrln = 71; //Chromosome length cM
nmloci = 1272; //Number of markers
mpos = even; //Marker positions
nma = all 2; //Number of marker alleles
maf = eql; //Marker allele frequencies
nqloci = 107; //Number of QTL
qpos = rnd;
nqa = rnd 2 3 4 5 6 7; //Number of QTL alleles
qaf = eql; //QTL allele frequencies
qae = rndg 0.4; //QTL allele effects ~ Gamma
end_chr;

//
begin_chr = 1; // CHR 19 //number of chromosomes with the same parameter
chrln = 63; //Chromosome length cM
nmloci = 1124; //Number of markers
mpos = even; //Marker positions
nma = all 2; //Number of marker alleles
maf = eql; //Marker allele frequencies
nqloci = 95; //Number of QTL
qpos = rnd;
nqa = rnd 2 3 4 5 6 7; //Number of QTL alleles
qaf = eql; //QTL allele frequencies
qae = rndg 0.4; //QTL allele effects ~ Gamma
end_chr;

//
begin_chr = 1; // CHR 20 //number of chromosomes with the same parameter
chrln = 56; //Chromosome length cM
nmloci = 998; //Number of markers
mpos = even; //Marker positions
nma = all 2; //Number of marker alleles
maf = eql; //Marker allele frequencies
nqloci = 84; //Number of QTL
qpos = rnd;
nqa = rnd 2 3 4 5 6 7; //Number of QTL alleles
qaf = eql; //QTL allele frequencies
qae = rndg 0.4; //QTL allele effects ~ Gamma
end_chr;

//
begin_chr = 1; // CHR 21 //number of chromosomes with the same parameter
chrln = 53; //Chromosome length cM
nmloci = 768; //Number of markers
mpos = even; //Marker positions
nma = all 2; //Number of marker alleles
maf = eql; //Marker allele frequencies
nqloci = 80; //Number of QTL
qpos = rnd;
nqa = rnd 2 3 4 5 6 7; //Number of QTL alleles
qaf = eql; //QTL allele frequencies

```

```

    qae = rndg 0.4;           //QTL allele effects ~ Gamma
    end_chr;
//

    begin_chr = 1; // CHR 22 //number of chromosomes with the same parameter
    chrln = 56;           //Chromosome length cM
    nmloci = 984;         //Number of markers
    mpos = even;          //Marker positions
    nma = all 2;           //Number of marker alleles
    maf = eql;            //Marker allele frequencies
    nqloci = 84;           //Number of QTL
    qpos = rnd;
    nqa = rnd 2 3 4 5 6 7; //Number of QTL alleles
    qaf = eql;            //QTL allele frequencies
    qae = rndg 0.4;       //QTL allele effects ~ Gamma
    end_chr;
//

    begin_chr = 1; // CHR 23 //number of chromosomes with the same parameter
    chrln = 68;           //Chromosome length cM
    nmloci = 1005;        //Number of markers
    mpos = even;          //Marker positions
    nma = all 2;           //Number of marker alleles
    maf = eql;            //Marker allele frequencies
    nqloci = 102;         //Number of QTL
    qpos = rnd;
    nqa = rnd 2 3 4 5 6 7; //Number of QTL alleles
    qaf = eql;            //QTL allele frequencies
    qae = rndg 0.4;       //QTL allele effects ~ Gamma
    end_chr;
//

    begin_chr = 1; // CHR 24 //number of chromosomes with the same parameter
    chrln = 43;           //Chromosome length cM
    nmloci = 662;         //Number of markers
    mpos = even;          //Marker positions
    nma = all 2;           //Number of marker alleles
    maf = eql;            //Marker allele frequencies
    nqloci = 66;          //Number of QTL
    qpos = rnd;
    nqa = rnd 2 3 4 5 6 7; //Number of QTL alleles
    qaf = eql;            //QTL allele frequencies
    qae = rndg 0.4;       //QTL allele effects ~ Gamma
    end_chr;
//

    begin_chr = 1; // CHR 25 //number of chromosomes with the same parameter
    chrln = 48;           //Chromosome length cM
    nmloci = 892;         //Number of markers
    mpos = even;          //Marker positions
    nma = all 2;           //Number of marker alleles
    maf = eql;            //Marker allele frequencies
    nqloci = 72;          //Number of QTL

```

```

qpos = rnd;
nqa = rnd 2 3 4 5 6 7;           //Number of QTL alleles
qaf = eql;                       //QTL allele frequencies
qae = rndg 0.4;                  //QTL allele effects ~ Gamma
end_chr;
//
begin_chr = 1; // CHR 26 //number of chromosomes with the same parameter
chrln = 49;           //Chromosome length cM
nmloci = 805;         //Number of markers
mpos = even;          //Marker positions
nma = all 2;          //Number of marker alleles
maf = eql;            //
nqloci = 74;          //Number of QTL
qpos = rnd;
nqa = rnd 2 3 4 5 6 7;           //Number of QTL alleles
qaf = eql;                       //QTL allele frequencies
qae = rndg 0.4;                  //QTL allele effects ~ Gamma
end_chr;

rmmg = 0.001; //rate of misssing marker genotypes
rmge = 0.005; //rate of marker genotyping error
mmutr = 1e-4 /recurrent; //Marker mutation rate
qmutr = 1e-4 /recurrent; //QTL mutation rate
end_genome;

/*****
** Output options **
*****/
begin_output;
linkage_map; //save marker and QTL linkage map
hp_stat;     //save brief statistics on historical population
allele_effect; //save allele effects
monitor_hp_homo; //Marker mean and QTL homozugosity on the hp
end_output;

```