

## Utilização de dados de genealogia e moleculares na seleção em programas de melhoramento

Brenda Vieira de Oliveira<sup>1</sup>, Renata Dourado Roque<sup>1</sup>, Guilherme Flaviano Pereira<sup>1</sup>, Camila Ferreira Azevedo<sup>2†</sup>, Ana Carolina Campana Nascimento<sup>2</sup>, Moysés Nascimento<sup>2</sup>

<sup>1</sup> Universidade Federal de Viçosa; Centro de Ciências Exatas e Tecnológicas; Programa de Pós-Graduação em Estatística Aplicada e Biometria; Viçosa – Minas Gerais, Brasil.

<sup>2</sup> Universidade Federal de Viçosa; Centro de Ciências Exatas e Tecnológicas; Departamento de Estatística; Viçosa – Minas Gerais, Brasil.

**Resumo:** Com os avanços tecnológicos em programas de melhoramento e biologia molecular, o uso de dados de pedigree (*A*) e moleculares (*G*) tornou-se parte essencial desses programas. As relações baseadas em pedigree refletem o parentesco por descendência, enquanto as relações baseadas em marcadores moleculares indicam o parentesco por estado. A integração dessas informações é um desafio, pois as diferentes fontes podem fornecer estimativas distintas dos valores genéticos dos indivíduos. Este estudo teve como objetivo avaliar o impacto do uso isolado e combinado das matrizes *A* e *G* na seleção de indivíduos superiores em programas de seleção recorrente, considerando o impacto no ganho genético, diversidade genética e endogamia ao longo de 15 ciclos de melhoramento. Foram simuladas cinco características fenotípicas com diferentes níveis de herdabilidade e graus de correlação positiva entre elas. As abordagens analisadas incluíram o uso isolado de *A* e *G*, a combinação das duas matrizes em uma matriz *H* e o uso de um índice baseado em *A* e *G* com pesos oriundos das confiabilidades dos modelos genômicos. A seleção baseada na matriz *H* demonstrou menor endogamia e maior diversidade genética, especialmente nos cenários com maior porcentagem de indivíduos genotipados. O índice baseado em *A* e *G* resultou em um aumento gradual da endogamia, especialmente nos cenários com correlação alta entre as características. Embora não tenham sido observadas diferenças no ganho genético, a matriz *H* apresentou uma tendência de destacar-se na manutenção da diversidade e controle da endogamia, mas sua significância não pode ser afirmada sem testes estatísticos adicionais.

**Palavras-chave:** Genotipagem; Parentesco; Pedigree; Marcadores moleculares.

## Use of pedigree and molecular data in selection in breeding programs

**Abstract:** With technological advances in breeding programs and molecular biology, the use of pedigree (*A*) and molecular (*G*) data has become an essential part of these programs. Pedigree-based relationships reflect the relationship by descent, while molecular marker-based relationships indicate the relationship by state. Integrating this information is a challenge, as different sources may provide distinct estimates of the genetic values of individuals. This study aimed to evaluate the impact of using *A* and *G* matrices both separately and in combination on the selection of superior individuals in recurrent selection programs, considering the impact on genetic gain, genetic diversity, and inbreeding over 15 breeding cycles. Five phenotypic traits were simulated, with varying levels of heritability and degrees of positive correlation between them. The approaches analyzed included the use of *A* and *G* separately, combining the two matrices into a matrix *H*, and using an index based on *A* and *G* with weights derived from the reliability of genomic models. Selection based on matrix *H* demonstrated lower inbreeding and higher genetic diversity, particularly in scenarios with a higher percentage of genotyped individuals. The index based on *A* and *G* resulted in a gradual increase in inbreeding, especially in scenarios with high correlation between traits. Although no differences in genetic gain were observed, matrix *H* showed a tendency to stand out in maintaining diversity and controlling inbreeding, but its significance cannot be asserted without additional statistical tests.

**Keywords:** Genotyping; Parentage; Pedigree; Molecular markers.

---

† Autor correspondente: [camila.azevedo@ufv.br](mailto:camila.azevedo@ufv.br)

Manuscript received on: 31/01/2025; revised on: 28/03/2025; accepted on: 01/04/2025.

## Introdução

Os programas de melhoramento genético têm como objetivo principal identificar indivíduos com elevado potencial genético, promovendo avanços significativos em populações de interesse (Resende *et al.*, 2012). Nesse contexto, a seleção genômica (SG) emerge como uma ferramenta eficiente, permitindo a seleção baseada em valores genéticos preditos, mesmo sem a avaliação fenotípica dos indivíduos, o que aumenta a acurácia na identificação de indivíduos geneticamente superiores em um intervalo de tempo reduzido (Cappa *et al.*, 2019; Begum *et al.*, 2015). Com os avanços tecnológicos nos programas de melhoramento e nas técnicas de biologia molecular, tornou-se possível integrar informações genéticas de relacionamento entre indivíduos, provenientes tanto de dados de pedigree quanto de marcadores moleculares, aprimorando a precisão na estimativa dos valores genéticos e, conseqüentemente, a eficácia da seleção (Ismael *et al.*, 2022; Laverdière *et al.*, 2022).

A matriz A, baseada no pedigree, representa as relações de parentesco entre os indivíduos por meio da descendência comum, enquanto a matriz G, construída com base em marcadores moleculares, fornece o parentesco por estado, refletindo as similaridades genéticas entre os indivíduos. Métodos como o ABLUP (*Best Linear Unbiased Prediction*) e o GBLUP (*Genomic BLUP*) utilizam essas matrizes para estimar valores genéticos, sendo o GBLUP particularmente eficaz ao incorporar informações genômicas (Nadeu *et al.*, 2023). Diversos estudos têm aplicado essas abordagens em diferentes culturas e características, como no crescimento e qualidade de pinheiro (Ukrainetz; Mansfield, 2020), eucalipto (Thavamanikumar *et al.*, 2020), qualidade de frutas em citros (Imai *et al.*, 2019) e características de crescimento e reprodutivas em gado (Forneris *et al.*, 2021), entre outros. A escolha entre essas fontes de informação ou a integração delas representa um desafio recorrente na prática dos programas de melhoramento.

A avaliação genética em programas de melhoramento genético inclui simultaneamente indivíduos fenotipados e genotipados, aqueles apenas fenotipados e aqueles exclusivamente genotipados. Embora as tecnologias de genotipagem tenham avançado significativamente, o custo elevado desse processo continua sendo um obstáculo para muitos programas (Sood *et al.*, 2020). Nesse cenário, Misztal, Legarra e Aguilar (2009), Christensen e Lund (2010) e Legarra *et al.* (2011) propuseram a matriz H como uma solução eficiente. Essa matriz combina dados genealógicos dos indivíduos apenas fenotipados, informações genotípicas dos indivíduos somente genotipados e, quando disponíveis, ambas as fontes de informação nos indivíduos com dados completos. Essa integração permite uma avaliação genética mais precisa, maximizando a eficiência do processo de seleção.

Além da integração das informações genotípicas e de pedigree na matriz H, é possível estabelecer um índice de seleção quando ambas as fontes de dados estão disponíveis (Resende; Silva; Azevedo, 2014). O índice é definido como uma combinação linear dos valores genéticos preditos a partir das matrizes G e A, em que os coeficientes de ponderação são atribuídos de acordo com as confiabilidades da predição dos valores genéticos gerada por cada método. Apesar de seu potencial para melhorar a precisão e a eficiência na seleção de indivíduos geneticamente superiores, essa abordagem ainda é pouco explorada nos programas de melhoramento genético.

A simulação é uma ferramenta amplamente empregada na tomada de decisão em programas de melhoramento genético, pois permite a avaliação prévia de diferentes estratégias e cenários antes de sua aplicação prática. Ela possibilita a análise de fatores essenciais, como a eficiência das ferramentas utilizadas, o impacto de abordagens distintas e os custos associados, proporcionando uma visão abrangente dos resultados esperados (Bancic *et al.*, 2025). Essa análise orienta os pesquisadores a tomar decisões e a otimizar recursos de forma eficaz. Embora não substitua a

implementação prática, a simulação oferece uma base sólida para identificar as estratégias mais promissoras, minimizando riscos e aumentando a probabilidade de sucesso no processo de melhoramento. A aplicação da simulação tem se mostrado relevante na agricultura, conforme demonstrado em estudos como os de Azevedo *et al.* (2024), Muleta, Pressoir e Morris (2019) e Lima *et al.* (2022).

Diante disso, o objetivo deste estudo foi avaliar o impacto da utilização isolada e combinada das matrizes de parentesco A e G nos parâmetros genéticos e na seleção de indivíduos geneticamente superiores em programas de seleção recorrente. Em particular, o estudo visa comparar o desempenho das seguintes abordagens: (i) uso das matrizes A e G isoladamente; (ii) combinação das matrizes A e G na matriz H; e (iii) uso da combinação de A e G por meio de um índice de seleção que utiliza os pesos baseados nas confiabilidades dos modelos de predição genômicos baseados em modelos mistos.

Adicionalmente, avaliou-se o impacto da porcentagem de indivíduos genotipados na matriz H, visando à otimização dos custos dos programas de melhoramento. As abordagens foram comparadas em termos de ganho genético, diversidade genética e endogamia, utilizando a simulação de um programa de seleção recorrente ao longo de 15 ciclos de melhoramento.

## Materiais e Métodos

### Dados

O conjunto de dados deste estudo originou-se da simulação de um programa de melhoramento genético com seleção recorrente, realizada por meio do software R (R Core Team, 2021) utilizando o pacote AlphaSimR (Faux *et al.*, 2016). A simulação consistiu em 15 ciclos iniciais de *burn-in* seguidos por 15 ciclos de melhoramento, repetidos 50 vezes. Inicialmente, foram simulados 50 genitores, dos quais foram realizados 150 cruzamentos. Cada cruzamento originou 20 indivíduos, resultando em uma população F1 composta por 3000 indivíduos. Por meio da seleção massal, na qual os indivíduos são escolhidos exclusivamente com base no fenótipo, 1000 indivíduos foram selecionados para fenotipagem em cinco características distintas.

O genoma dos indivíduos consistia em 12 cromossomos, com 12 loci de características quantitativas (QTLs - *Quantitative Trait Loci*) cada e 200 polimorfismos de nucleotídeo único (SNPs - *Single Nucleotide Polymorphisms*) por cromossomo. As características fenotípicas foram modeladas considerando diferentes níveis de herdabilidade (baixa, moderada e alta), com valores definidos em 0,10, 0,25, 0,30, 0,60 e 0,80, aplicados em todos os cenários. Os graus de correlação entre as características foram classificados como fraco (0 a 0,10), moderado (0,40 a 0,60) e alto (0,60 a 1), todos com sinal positivo.

Utilizando os dados fenotípicos, genotípicos e de genealogias simulados, os valores preditos genéticos foram calculados utilizando os métodos ABLUP, GBLUP e HBLUP. Adicionalmente, foi calculado um índice que combina os valores preditos pelos métodos ABLUP e GBLUP. A comparação entre as abordagens foi realizada da seguinte forma: i) Ganho genético: Avaliado pela média paramétrica do valor genético dos pais a cada ciclo; ii) Diversidade Genética: Medida pela variância genética paramétrica dos pais a cada ciclo; iii) Endogamia: Calculada pela média da diagonal da matriz de parentesco entre os pais, subtraindo 1 (Resende; Silva; Azevedo, 2014).

A simulação foi desenvolvida de forma a mimetizar um programa de melhoramento com seleção recorrente, garantindo análises consistentes e comparáveis entre os cenários avaliados. Ademais, a proporção de indivíduos genotipados incluídos na matriz H foi avaliada em diferentes cenários, iniciando com 10% e recebendo incrementos graduais de 10% até atingir 100%.

**GBLUP e ABLUP**

A predição dos valores genéticos dos indivíduos por meio do ABLUP ou GBLUP é baseada no modelo linear misto, conforme descrito por Resende (2008) e Resende *et al.*(2012):

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

$$y = Xb + Zg + e \quad (2)$$

em que  $y$  é o vetor de observações fenotípicas,  $X$  e  $Z$  são as matrizes de incidência associadas aos efeitos fixos ( $b$ ), representando, neste caso, a média da característica, aos efeitos genéticos aleatórios ( $a$ ) e aos efeitos genéticos genômicos aleatórios ( $g$ ) dos indivíduos, respectivamente, com  $a \sim N(0, A\sigma_a^2)$  para o ABLUP e  $g \sim N(0, G\sigma_g^2)$  para o GBLUP, onde  $\sigma_a^2$  e  $\sigma_g^2$  são as variâncias genéticas aditivas.  $e$  é o vetor de erros residuais, assumindo  $e \sim N(0, I\sigma_e^2)$ , onde  $I$  é a matriz identidade e  $\sigma_e^2$  a variância residual. As equações do modelo misto para a predição dos valores genéticos dos indivíduos, utilizando as matrizes de parentesco  $A$  (pedigree) e  $G$  (genômica) de forma isolada, são expressas da seguinte maneira:

Para a matriz  $A$ :

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + A^{-1} \frac{\sigma_e^2}{\sigma_g^2} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{g} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix}$$

Para a matriz  $G$ :

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + G^{-1} \frac{\sigma_e^2}{\sigma_g^2} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{g} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix}$$

em que os componentes de variância ( $\sigma_e^2$ ,  $\sigma_a^2$  e  $\sigma_g^2$ ) são estimados via REML (*Restricted Maximum Likelihood*);

As confiabilidades do GBLUP ( $r_{g\hat{g}}^2$ ) são definidas como sendo:

$$r_{g\hat{g}}^2 = 1 - \frac{PEV}{\hat{\sigma}_g^2}$$

em que  $PEV = C^{22} \hat{\sigma}_e^2$  e  $C^{22} = \left( Z'Z + G^{-1} \frac{\sigma_e^2}{\sigma_g^2} \right)^{-1}$ . A confiabilidade do ABLUP ( $r_{a\hat{a}}^2$ ) é definida de modo similar.

A matriz  $H$  representa uma combinação das informações genéticas provenientes do pedigree ( $A$ ) e dos marcadores moleculares ( $G$ ), integrando as relações de parentesco dos indivíduos genotipados e não genotipados. Essa matriz é definida como:

$$H = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & G \end{bmatrix},$$

em que os subscritos 1 e 2 correspondem, respectivamente, os indivíduos genotipados e não genotipados (Resende; Silva; Azevedo, 2014). De acordo com Misztal, Legarra e Aguilar (2009), as equações de modelos mistos para a matriz H baseadas de forma similar aos modelos (1) e (2), utilizadas para o método HBLUP, são dadas por:

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + H^{-1} \frac{\sigma_e^2}{\sigma_g^2} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{g} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix}$$

### Índice de seleção

Dispondo dos valores preditos com base em dados de pedigree e marcadores moleculares, é possível estabelecer um índice de seleção que integre essas duas fontes de informação (Resende; Silva; Azevedo, 2014). Esse índice é estimado como  $I_j = b_1 \hat{g} + b_2 \hat{a}$ , onde  $b_1$  e  $b_2$  representam os coeficientes de ponderação atribuídos aos valores genéticos genômicos ( $\hat{g}$ ) e tradicionais ( $\hat{a}$ ), respectivamente. Os pesos ( $b_1$  e  $b_2$ ) são determinados com base nas confiabilidades dos modelos de predição (1) e (2), de acordo com as seguintes expressões:

$$b_1 = \frac{1 - r_{\hat{a}\hat{a}}^2}{1 - r_{\hat{g}\hat{g}}^2 r_{\hat{a}\hat{a}}^2}$$

e

$$b_2 = \frac{1 - r_{\hat{g}\hat{g}}^2}{1 - r_{\hat{g}\hat{g}}^2 r_{\hat{a}\hat{a}}^2}$$

em que  $r_{\hat{g}\hat{g}}^2$  é a confiabilidade da seleção genômica e  $r_{\hat{a}\hat{a}}^2$  é a confiabilidade da predição dos valores genéticos baseada em pedigree.

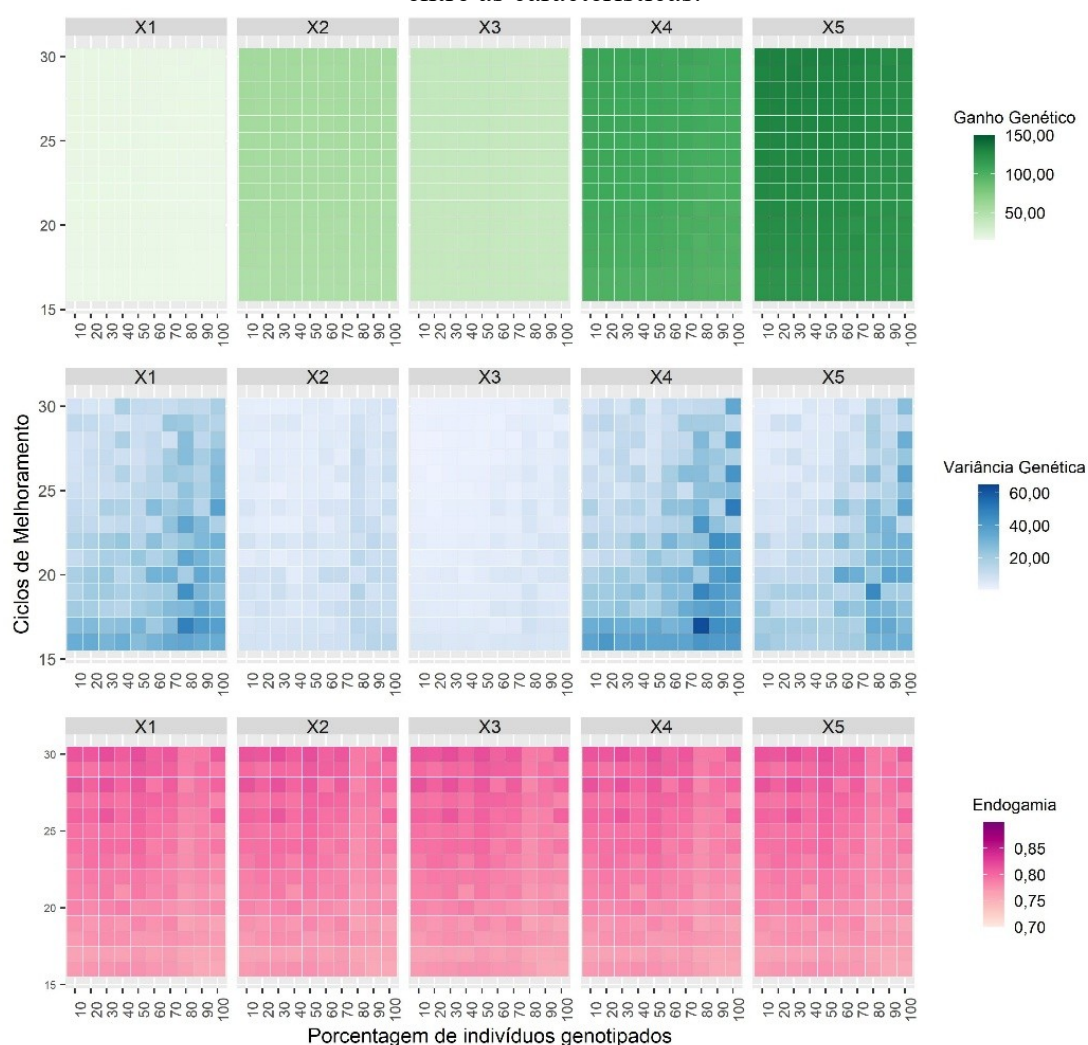
Os *heatmaps* foram construídos a partir da média das 50 repetições para cada medida comparativa analisada: ganho genético, variância genética e endogamia. Para assegurar uma escala uniforme entre os diferentes cenários, os valores mínimo e máximo foram definidos com base nos extremos observados em todas as condições avaliadas. Além disso, a porcentagem de indivíduos genotipados com melhor desempenho foi incluída no *heatmap* de comparação entre as estratégias de seleção. Nos *heatmaps* que comparam as diferentes abordagens de seleção, a seleção baseada exclusivamente na matriz A foi representada por "A", na matriz G por "G", na matriz H por "H", e a abordagem baseada no índice, que integra múltiplas informações, foi indicada como "AG".

## Resultados e Discussão

Os resultados relativos à proporção de indivíduos genotipados na matriz H são apresentados nas Figuras 1, 2 e 3, diferenciando-se pelos cenários de correlação fraca, moderada e alta, respectivamente. Em termos de ganho genético e diversidade genética, foi observado maior ganho genético nos cenários com correlação alta entre as características, acompanhado de menores valores de diversidade genética. Por outro lado, nos cenários com correlação baixa, o ganho genético foi reduzido, enquanto a diversidade genética aumentou. Ao analisar as herdabilidades das

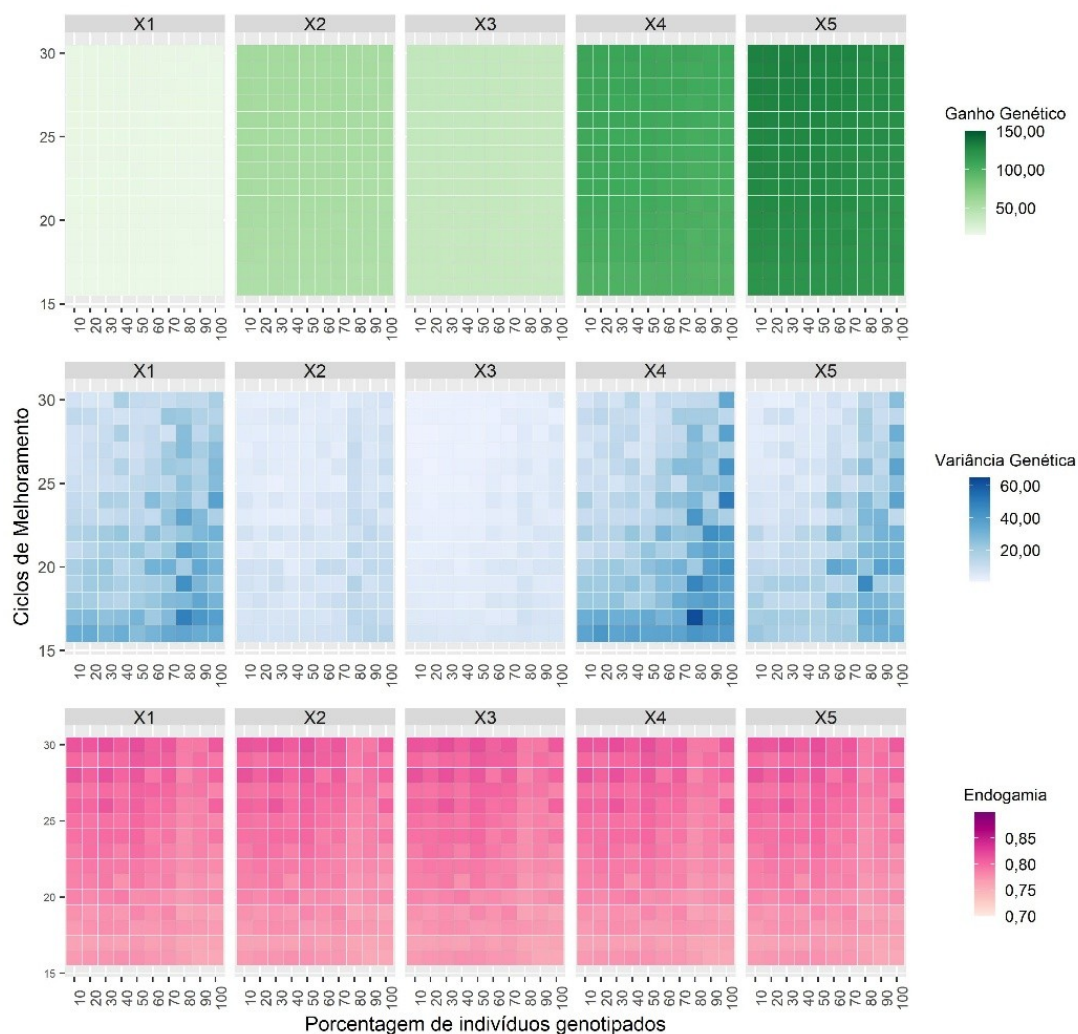
características, conclui-se que aquelas com maior herdabilidade geraram maior ganho genético, independentemente da intensidade da correlação. Adicionalmente, as características 1, 4 e 5 apresentaram maior diversidade genética nos cenários de correlação baixa e moderada, com valores de herdabilidade de 0,10, 0,60 e 0,80, respectivamente. A maior diversidade genética em características com baixa herdabilidade ocorre porque a variação fenotípica é amplamente influenciada por fatores ambientais, reduzindo a eficiência da seleção em fixar alelos favoráveis, indicando a possibilidade de melhoramento genético para tais características (Evangelista *et al.*, 2020). Por outro lado, características com alta herdabilidade permitem maior eficiência na seleção de alelos desejáveis, o que inicialmente pode aumentar a diversidade, mas tende a reduzi-la ao longo dos ciclos devido à pressão seletiva.

Figura 1: *Heatmaps* do ganho genético (verde), diversidade genética (azul) e coeficiente de endogamia (rosa) ao longo dos ciclos de melhoramento, utilizando a matriz H na seleção. As análises consideram diferentes porcentagens de indivíduos genotipados em cinco características fenotípicas, ordenadas de acordo com os níveis de herdabilidade, com correlações positivas fracas entre as características.



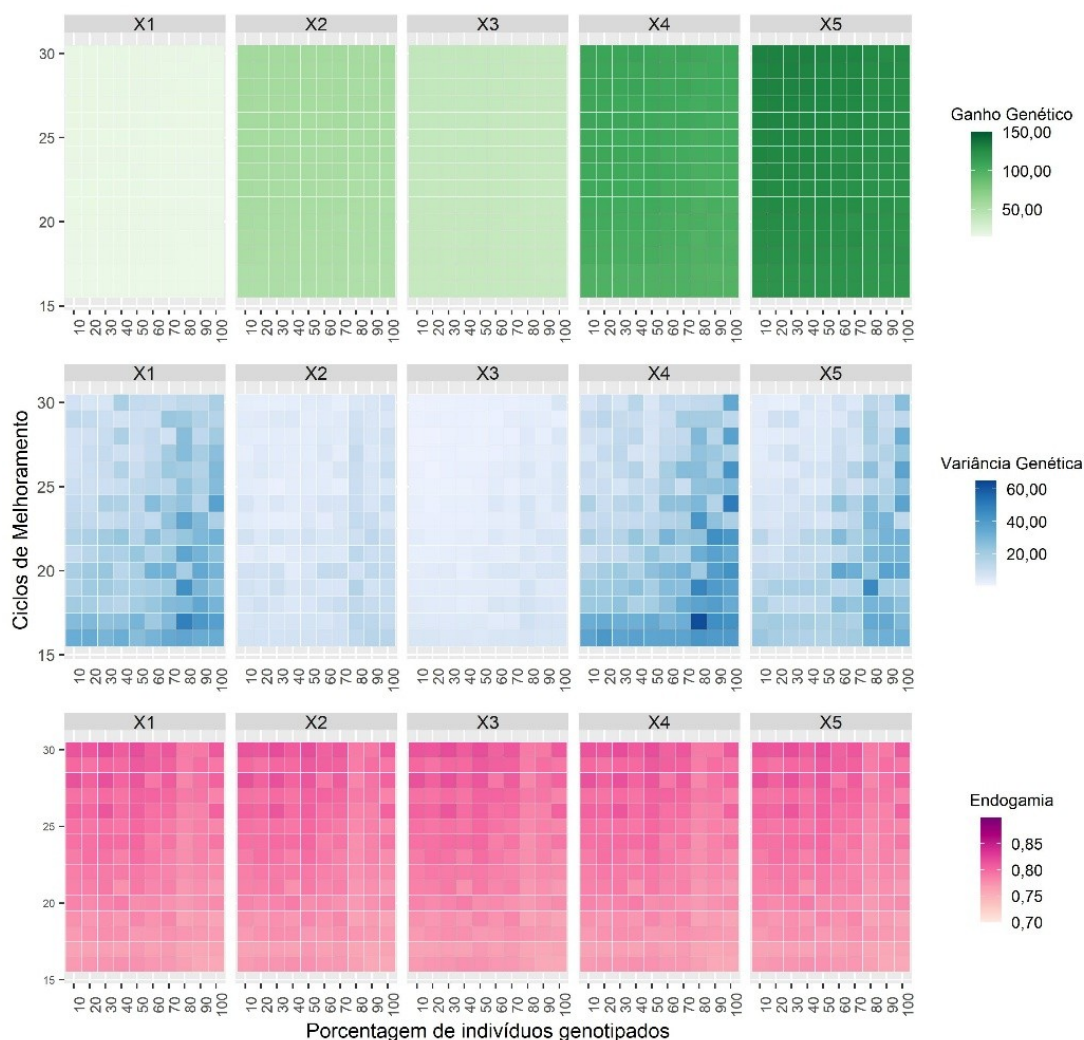
Fonte: dos autores (2025).

Figura 2: *Heatmaps* do ganho genético (verde), diversidade genética (azul) e coeficiente de endogamia (rosa) ao longo dos ciclos de melhoramento, utilizando a matriz H na seleção. As análises consideram diferentes porcentagens de indivíduos genotipados em cinco características fenotípicas, ordenadas de acordo com os níveis de herdabilidade, com correlações positivas moderadas entre as características.



Fonte: dos autores (2025).

Figura 3: *Heatmaps* do ganho genético (verde), diversidade genética (azul) e coeficiente de endogamia (rosa) ao longo dos ciclos de melhoramento, utilizando a matriz H na seleção. As análises consideram diferentes porcentagens de indivíduos genotipados em cinco características fenotípicas, ordenadas de acordo com os níveis de herdabilidade, com correlações positivas altas entre as características.

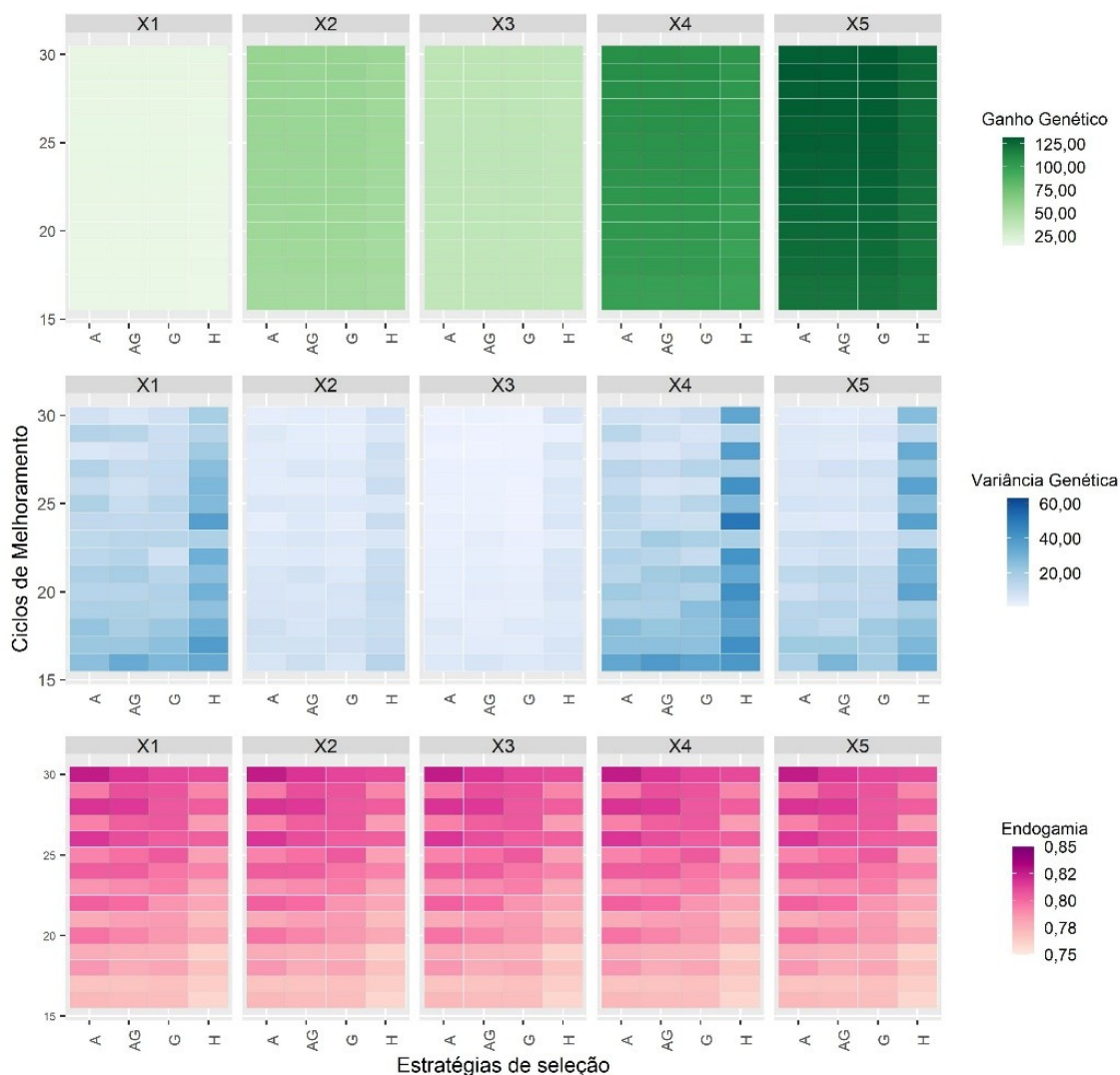


Fonte: dos autores (2025).

O aumento da porcentagem de indivíduos genotipados da matriz H apresentou um padrão de comportamento apenas na endogamia, que diminuiu nos ciclos finais de melhoramento, especialmente nos cenários de alta correlação. Por outro lado, não foi possível identificar um padrão semelhante para o ganho e variância genética. Esse efeito sugere que a maior precisão nas estimativas genéticas, proporcionada pela inclusão de dados genotípicos, resulta em uma seleção mais eficiente, o que contribui para a redução da endogamia ao longo do tempo. Esses resultados corroboram o trabalho de Ukrainetz e Mansfield (2020), que demonstraram que a inclusão de dados genotípicos na matriz H aumenta a precisão das previsões genéticas. O estudo reportou que as correlações de classificação, uma medida da concordância entre as classificações dos indivíduos, aumentaram de 0,08 para 0,25 no HBLUP quando 40% da população de treinamento foi genotipada. Assim, o investimento em genotipagem de um número maior de indivíduos se mostra compensatório, promovendo maior precisão na seleção e controle da endogamia, especialmente em cenários de alta correlação e programas de melhoramento de longo prazo. Portanto, a matriz H com 100% dos indivíduos genotipados apresentou melhor desempenho, sendo utilizada na comparação das estratégias de seleção.

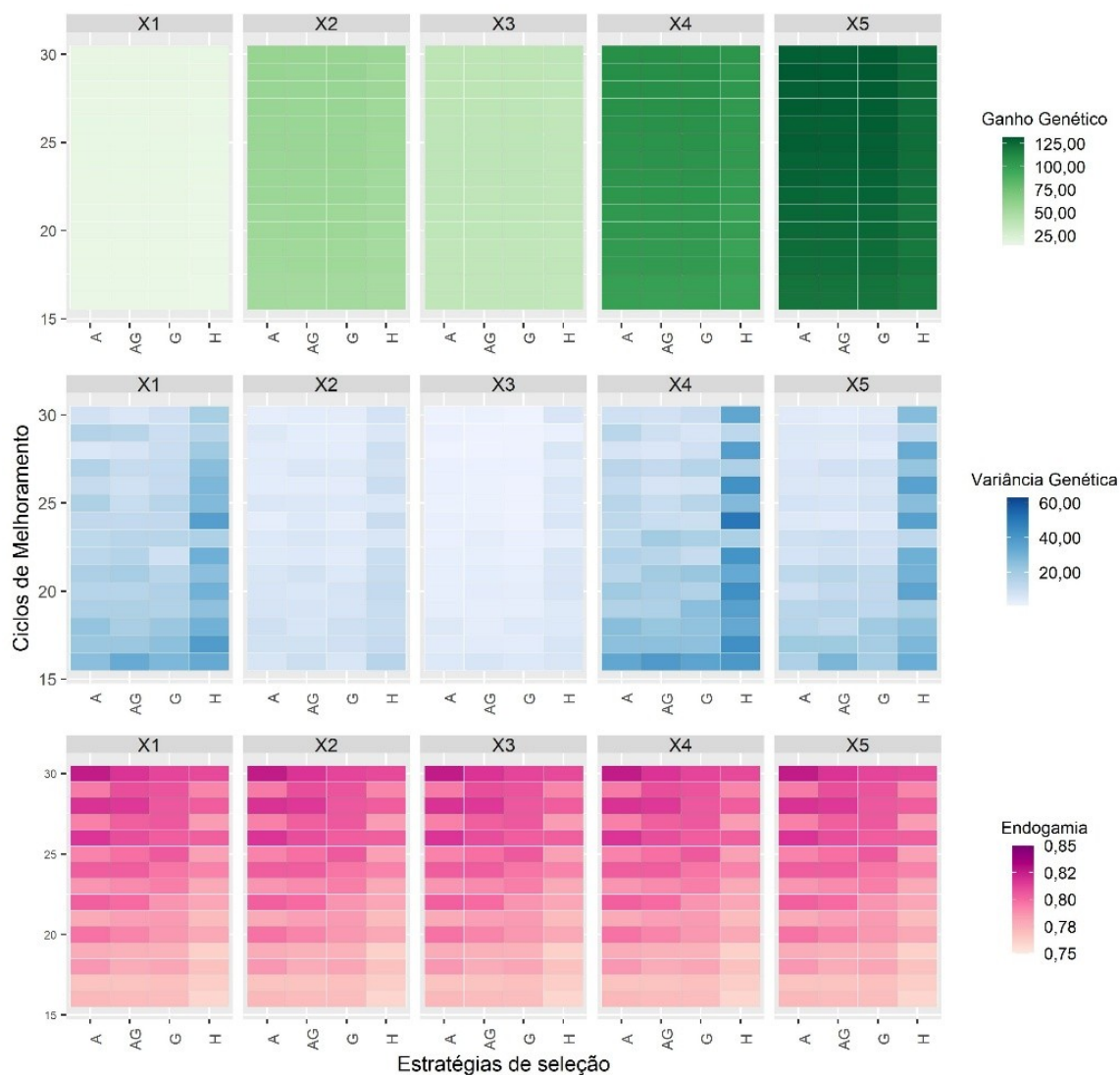
Ao analisar o impacto das diferentes abordagens de seleção nos cenários de correlação baixa, moderada e alta (Figuras 4, 5 e 6, respectivamente), observou-se que o índice baseado nos valores genéticos estimados pelas matrizes A e G resulta em um aumento da endogamia ao longo dos ciclos de melhoramento. Esse efeito foi mais pronunciado nos cenários com correlações fracas entre as características. Em termos de ganho genético, as diferentes abordagens não apresentaram diferenças, indicando que o uso das matrizes A e G proporciona ganhos semelhantes em comparação com outras estratégias. Embora o índice tenha se mostrado tão eficiente quanto as demais abordagens em termos de ganho genético e diversidade genética, o aumento progressivo da endogamia ao longo dos ciclos destaca a importância de implementar estratégias complementares que mitiguem possíveis impactos negativos a longo prazo. Os resultados também indicaram que a herdabilidade das características não influenciou o desempenho das diferentes abordagens analisadas, mantendo-se consistente entre elas. De modo similar, as correlações entre as características não alteraram o padrão dos parâmetros genéticos avaliados, que se manteve consistente independentemente da estratégia adotada.

Figura 4: *Heatmaps* do ganho genético (verde), diversidade genética (azul) e coeficiente de endogamia (rosa) ao longo dos ciclos de melhoramento, considerando as matrizes de parentesco e o índice: A, G, H e AG. As análises consideram cinco características fenotípicas, ordenadas de acordo com os níveis de herdabilidade e correlações positivas fracas.



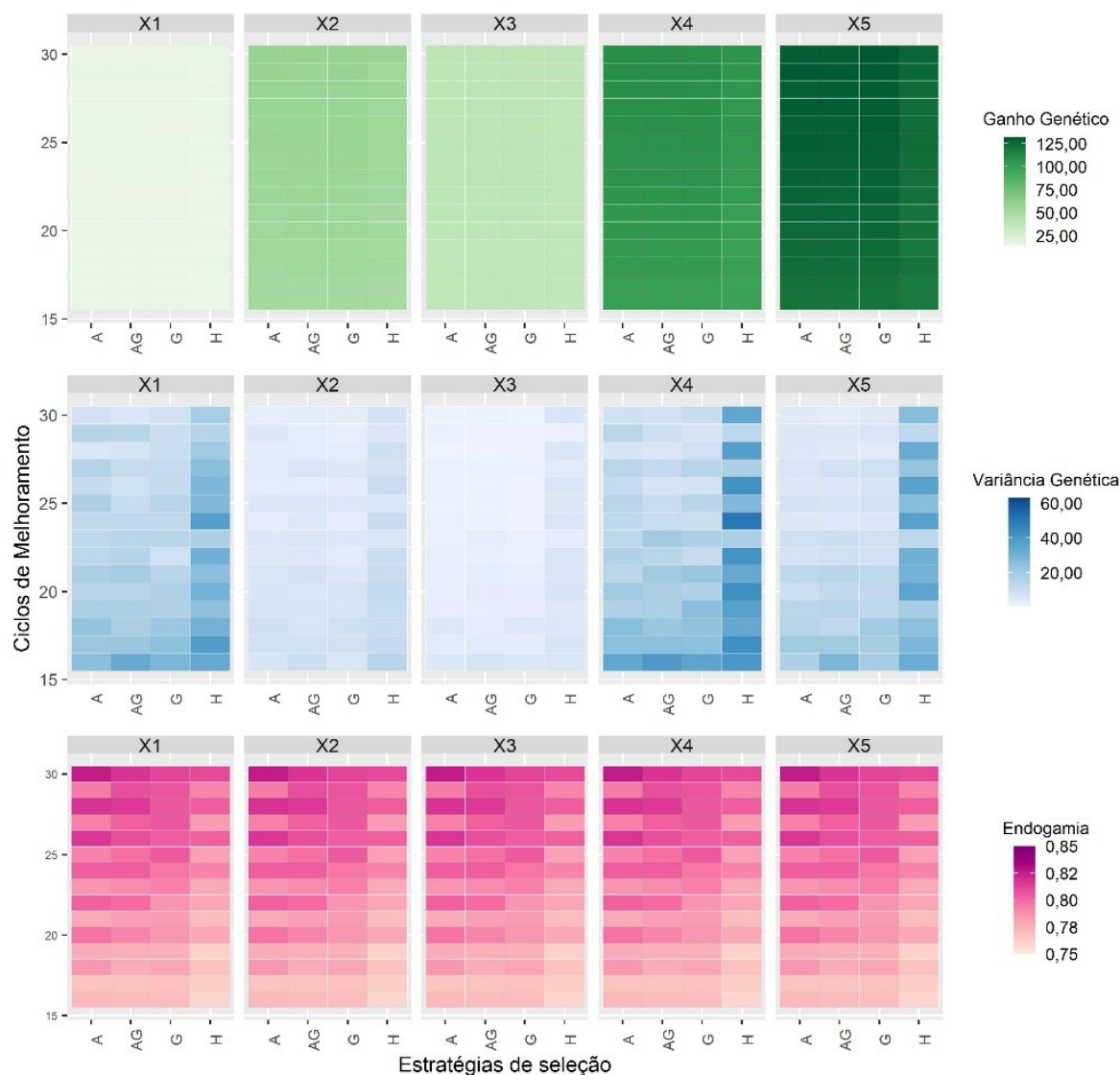
Fonte: dos autores (2025).

Figura 5: *Heatmaps* do ganho genético (verde), diversidade genética (azul) e coeficiente de endogamia (rosa) ao longo dos ciclos de melhoramento, considerando as matrizes de parentesco e o índice: A, G, H e AG. As análises consideram cinco características fenotípicas, ordenadas de acordo com os níveis de herdabilidade e correlações positivas moderadas.



Fonte: dos autores (2025).

Figura 6: *Heatmaps* do ganho genético (verde), diversidade genética (azul) e coeficiente de endogamia (rosa) ao longo dos ciclos de melhoramento, considerando as matrizes de parentesco e o índice: A, G, H e AG. As análises consideram cinco características fenotípicas, ordenadas de acordo com os níveis de herdabilidade e correlações positivas altas.



Fonte: dos autores (2025).

## Conclusão

Destaca-se como melhor desempenho em termos de preservação da diversidade genética e controle da endogamia a seleção utilizando a matriz H, demonstrando ser uma alternativa eficaz para a combinação das matrizes A e G. Em relação ao ganho genético, não foram observadas diferenças significativas entre as matrizes avaliadas, indicando que o desempenho no progresso genético não foi afetado pela escolha da abordagem utilizada. No entanto, o uso do índice contribui para o aumento da endogamia ao longo dos ciclos de melhoramento. O aumento na proporção de indivíduos genotipados na matriz H apresenta potencial para contribuir na redução da endogamia.

## Agradecimentos

Os autores agradecem ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo apoio financeiro.

## Referências

- AZEVEDO, C. F. *et al.* Using visual scores for genomic prediction of complex traits in breeding programs. **Theoretical and Applied Genetics**, v. 137, n. 1, 2024.
- BANČIČ, J. *et al.* Plant breeding simulations with AlphaSimR. **Crop Science**, v. 65, n. 1, p. e21312, 2025.
- BEGUM, H. *et al.* Genome-wide association mapping for yield and other agronomic traits in an elite breeding population of tropical rice (*Oryza sativa*). **PloS One**, v. 10, n. 3, e0119873, 2015.
- CAPPA, E. P. *et al.* Improving genomic prediction of growth and wood traits in Eucalyptus using phenotypes from non-genotyped trees by single-step GBLUP. **Plant Science**, v. 284, p. 9-15, 2019.
- CHRISTENSEN, O. F.; LUND, M. S. Predição genômica quando alguns animais não são genotipados. **Genet Sel Evol**, v. 42, n. 2, 2010.
- CROUÉ, I.; DUCROCQ, V. Genomic and single-step evaluations of carcass traits of young bulls in dual-purpose cattle. **Journal of Animal Breeding and Genetics**, v. 134, n. 4, p. 300-307, 2017.
- EVANGELISTA, A. F. *et al.* Estimação de parâmetros genéticos para características de crescimento em bovinos Nelore Mocho da Região Norte do Brasil. 2020.
- FAUX, A. *et al.* AlphaSim: Software for Breeding Program Simulation. **The Plant Genome**, 9(3), 2016.
- FORNERIS, N. S. *et al.* Estimating inbreeding depression for growth and reproductive traits using pedigree and genomic methods in Argentinean Brangus cattle. **Journal of Animal Science**, v. 99, n. 11, skab289, 2021.
- IMAI, A. *et al.* Single-step genomic prediction of fruit-quality traits using phenotypic records of non-genotyped relatives in citrus. **PloS One**, v. 14, n. 8, e0221880, 2019.
- ISMAEL, A. *et al.* Genetic variation in drought-tolerance traits and their relationships to growth in *Pinus radiata* D. Don under water stress. **Frontiers in Plant Science**, v. 12, p. 766803, 2022.
- LAVERDIÈRE, J. P. *et al.* Breeding for adaptation to climate change: Genomic selection for drought response in a white spruce multi-site polycross test. **Evolutionary Applications**, v. 15, n. 3, p. 383-402, 2022.
- LEGARRA, F. *et al.* Use of a reduced set of single nucleotide polymorphisms for genetic evaluation of resistance to Salmonella carrier state in laying hens. **Poultry Science**, v. 90, n. 4, p. 731-736, 2011.

LIMA, L. P. *et al.* Evaluation of Bayesian methods of genomic association via chromosomic regions using simulated data. **Scientia Agricola**, v. 79, n. 3, 2022.

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

MULETA, K. T.; PRESSOIR, G.; MORRIS, G. P. Optimizing genomic selection for a sorghum breeding program in Haiti: A simulation study. **G3: Genes, Genomes, Genetics**, v. 9, n. 2, p. 391-401, 2019.

NADEAU, S. *et al.* Increasing genomic prediction accuracy for unphenotyped full-sib families by modeling additive and dominance effects with large datasets in white spruce. **Frontiers in Plant Science**, v. 14, p. 1137834, 2023.

PÉRTILE, S. F. N. *et al.* Seleção e associação genômica ampla para o melhoramento genético animal com uso do método ssGBLUP. **Pesquisa Agropecuária Brasileira**, v. 51, p. 1729-1736, 2016.

RESENDE, M. D. V. **Genômica quantitativa e seleção no melhoramento de perenes e animais**. Colombo: Embrapa Florestas, p. 330, 2008.

RESENDE, M. D. V. *et al.* **Seleção genômica ampla (GWS) via modelos mistos (REML/BLUP), inferência bayesiana (MCMC), regressão aleatória multivariada e estatística espacial**. Viçosa: Ed. UFV, 2012. 7 p.

RESENDE, M. D. V.; SILVA, F. F.; AZEVEDO, C. F. (2014). **Estatística matemática, biométrica e computacional: Modelos mistos, multivariados, categóricos e generalizados (REML/BLUP), inferência bayesiana, regressão aleatória, seleção 57 genômica, QTL-GWAS, estatística espacial e temporal, competição, sobrevivência**. Viçosa: UFV.

SOOD, S. *et al.* Making the most of all data: combining non-genotyped and genotyped potato individuals with HBLUP. **The Plant Genome**, v. 13, n. 3, e20056, 2020.

THAVAMANIKUMAR, S. *et al.* Genomic studies reveal substantial dominant effects and improved genomic predictions in an open-pollinated breeding population of *Eucalyptus pellita*. **G3: Genes, Genomes, Genetics**, v. 10, n. 10, p. 3751-3763, 2020.