

GUILHERME FERREIRA MELCHERT

**ESTUDOS GENÉTICOS EM PROGÊNIES AUTOFECDUNDADAS DE *Eucalyptus*
spp.**

Botucatu

2025

GUILHERME FERREIRA MELCHERT

**ESTUDOS GENÉTICOS EM PROGÊNIES AUTOFECONDADAS DE *Eucalyptus*
spp.**

Dissertação apresentada à Faculdade de
Ciências Agronômicas da Unesp Câmpus
de Botucatu, para obtenção do título de
Mestre em Ciência Florestal.

Orientador: Evandro Vagner Tambarussi

Botucatu

2025

M518e

Melchert, Guilherme

Estudos genéticos em progênies autofecundadas de
Eucalyptus spp. / Guilherme Melchert. -- Botucatu, 2025
81 p.

Dissertação (mestrado) - Universidade Estadual Paulista
(UNESP), Faculdade de Ciências Agronômicas, Botucatu
Orientador: Evandro Vagner Tambarussi

1. genética-quantitativa. 2. linhagens-endogâmicas. 3.
seleção-gênômica. I. Título.

CERTIFICADO DE APROVAÇÃO

TÍTULO DA DISSERTAÇÃO: ESTUDOS GENÉTICOS EM PROGÊNIES ENDOGÂMICAS DE *Eucalyptus* spp.

AUTOR: GUILHERME FERREIRA MELCHERT

ORIENTADOR: EVANDRO VAGNER TAMBARUSSI

COORIENTADOR: MARCIO JOSÉ DE ARAUJO

Aprovado como parte das exigências para obtenção do Título de Mestre em Ciência Florestal, pela Comissão Examinadora:

Documento assinado digitalmente
 **EVANDRO VAGNER TAMBARUSSI**
Data: 09/03/2025 19:46:46-0300
Verifique em <https://validar.itl.gov.br>

Prof. Dr. EVANDRO VAGNER TAMBARUSSI (Participação Virtual)
Departamento de Produção Vegetal / Faculdade de Ciências Agrônomicas - Unesp Botucatu

Pesquisadora Dr.^a FABIANA REZENDE MUNIZ (Participação Virtual)
Sustentabilidade, Pesquisa e Inovação / Suzano Papel e Celulose S/A

Documento assinado digitalmente
 **BRUNO CESAR ROSSINI**
Data: 07/03/2025 15:10:05-0300
Verifique em <https://validar.itl.gov.br>

Documento assinado digitalmente
 **FABIANA REZENDE MUNIZ**
Data: 07/03/2025 16:00:32-0300
Verifique em <https://validar.itl.gov.br>

Dr. BRUNO CESAR ROSSINI (Participação Virtual)
Departamento de Genética / Instituto de Biociências de Botucatu

Botucatu, 07 de março de 2025

AGRADECIMENTOS

Gostaria, primeiramente, de expressar minha profunda gratidão aos meus pais, Jacicleide e William, pelo apoio incondicional, tanto emocional quanto financeiro, ao longo de todos os meus anos de estudo.

Agradeço também à equipe do Laboratório de Melhoramento de Plantas Agrícolas e Florestais (LAMPAPF), em especial ao meu orientador, Evandro Vagner Tambarussi, por me proporcionar a oportunidade de desenvolver esta pesquisa e por todas as portas que abriu para mim desde a graduação. Minha sincera gratidão à Dandara, por toda a assistência na pesquisa, pelas conversas enriquecedoras e pela troca de experiências, que contribuíram significativamente para o meu crescimento acadêmico e profissional.

Aos meus amigos e colegas, agradeço o apoio e incentivo ao longo desses dois anos de mestrado. Um agradecimento especial à minha namorada, Mayara Martins Bueno, por sempre acreditar em mim, mesmo nos momentos em que eu próprio duvidava.

À Suzano S.A., minha imensa gratidão por disponibilizar os dados essenciais para esta pesquisa, além da valiosa colaboração, tanto financeira quanto intelectual, que permitiu aprimorar ainda mais o estudo.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - 88887.86646/2023-00.

RESUMO

O melhoramento genético por meio do cruzamento de indivíduos heterozigotos, utilizando a seleção recorrente recíproca, é comumente aplicado no gênero *Eucalyptus*. Indivíduos com características de interesse são clonados para explorar os efeitos aditivos e não-aditivos. Atualmente, o desafio do melhoramento genético desse gênero é aumentar a resposta à seleção, o que torna fundamental a busca por novos métodos de melhoramento. Assim, o uso de linhagens é uma alternativa para explorar a complementaridade entre genótipos contrastantes. No entanto, para implementar essa estratégia, é necessário compreender os padrões de endogamia da população autofecundada. Dessa forma, a pesquisa apresentada teve como objetivo estudar esses padrões em uma população $S_{0:1}$. Vinte clones superiores de diferentes espécies de *Eucalyptus* spp. (população S_0) foram autofecundados, e mudas dessas autofecundações foram produzidas. Aproximadamente 30 mudas de cada genitor foram implantadas em um experimento de campo com 30 blocos e uma planta por parcela, enquanto o restante dos indivíduos (62 no total) foram implantados em um pomar de hibridação. Tanto os indivíduos quanto seus genitores (682 indivíduos) foram genotipados usando o painel customizado da empresa, com 10.132 SNPs. Adicionalmente, as mudas foram avaliadas quanto ao Diâmetro na Altura do Peito (DAP, em centímetros) aos 36 meses de idade. No primeiro capítulo, os coeficientes de endogamia foram estimados por meio das *Runs of Homozygosity* (ROHs). A depressão por endogamia foi estimada aos 36 meses, utilizando uma regressão entre os fenótipos dos indivíduos e seus respectivos coeficientes de endogamia. Também foram definidos blocos haplotípicos para investigar o impacto da endogamia sobre essas estruturas genômicas. Já no segundo capítulo, os valores genéticos genômicos dos indivíduos foram estimados, e uma análise de associação genômica ampla (GWAS) foi realizada para identificar possíveis regiões do genoma com maior influência nas características avaliadas. Foram testados quatro modelos frequentistas (ABLUP, GBLUP, GBLUP-AD e HBLUP) e cinco modelos bayesianos (BRR, Bayes A, Bayes B, Bayes C e Bayes LASSO) de predição genômica, com o objetivo de prever o desempenho de indivíduos não fenotipados no pomar de hibridação. No primeiro capítulo, não foi observada depressão endogâmica até os 36 meses de idade. Os coeficientes de endogamia individuais, estimados pelas ROHs, variaram de 0 a 0,41, com os cromossomos mais endogâmicos sendo os 4 (0,27), 9 (0,39) e 10 (0,28). Para os blocos haplotípicos, os cromossomos 6 (56 blocos) e 8 (51 blocos) apresentaram o maior número de blocos, sendo que os blocos de maior extensão estavam associados aos cromossomos mais endogâmicos. No segundo capítulo, a análise GWAS não identificou nenhuma região do genoma com maior influência no fenótipo. Foram determinados os parâmetros de seleção genômica para todos os modelos avaliados. Dentre os modelos de seleção genômica testados, o GBLUP se destacou como superior em todos os parâmetros estimados. Os resultados obtidos indicam que o GBLUP tem grande potencial para uso no gênero *Eucalyptus*, devido à sua adaptação à homozigose no genoma. Além disso, esses resultados reforçam a importância de compreender os padrões de endogamia para selecionar de forma criteriosa os indivíduos que avançarão para as próximas gerações de autofecundação, com o objetivo de obter híbridos superiores e específicos.

Palavras-chave: genética-quantitativa; linhagens-endogâmicas; seleção-genômica.

ABSTRACT

Genetic breeding through the crossing of heterozygous individuals, using reciprocal recurrent selection, is commonly applied to the *Eucalyptus* genus. Individuals with traits of interest are cloned to exploit additive and non-additive effects. Currently, the challenge in the genetic breeding of this genus is to improve the response to selection, which makes the search for new breeding methods fundamental. Thus, the use of inbred lines is an alternative to explore complementarity between contrasting genotypes. However, to implement this strategy, it is necessary to understand the inbreeding patterns of self-fertilized populations. Therefore, the present research aimed to study these patterns in an S0.1 population. Twenty superior clones from different *Eucalyptus* spp. (S0 population) were self-fertilized, and seedlings from these self-fertilizations were produced. Approximately 30 seedlings from each parent were planted in a field experiment with 30 blocks and one plant per plot, while the remaining individuals (62 in total) were planted in a hybridization orchard. Both the individuals and their parents (682 individuals) were genotyped using the company's custom panel, with 10,132 SNPs. Additionally, the seedlings were evaluated for diameter at breast height (DBH, in centimeters) at 36 months of age. In the first chapter, inbreeding coefficients were estimated using *Runs of Homozygosity* (ROHs). Inbreeding depression was estimated at 36 months, using a regression between an individual's phenotypes and their respective inbreeding coefficients. Haplotype blocks were also defined to investigate the impact of inbreeding on these genomic structures. In the second chapter, the genomic estimated breeding values of the individuals were assessed, and a genome-wide association study (GWAS) was performed to identify potential regions of the genome with greater influence on the evaluated traits. Four frequentist models (ABLUP, GBLUP, GBLUP-AD, and HBLUP) and five Bayesian models (BRR, Bayes A, Bayes B, Bayes C, and Bayes LASSO) of genomic prediction were tested to predict the performance of non-phenotyped individuals in the hybridization orchard. In the first chapter, no inbreeding depression was observed up to 36 months of age. Individual inbreeding coefficients, estimated by ROHs, ranged from 0 to 0.41, with chromosomes 4 (0.27), 9 (0.39), and 10 (0.28) being the most inbred. For the haplotype blocks, chromosomes 6 (56 blocks) and 8 (51 blocks) showed the largest number of blocks, with the longest blocks being associated with the most inbred chromosomes. In the second chapter, the GWAS analysis did not identify any region of the genome with greater influence on the phenotype. Genomic selection parameters were determined for all evaluated models. Among those tested, GBLUP stood out as superior for all estimated parameters. The results obtained indicate that GBLUP has significant potential for use with the *Eucalyptus* genus, due to its adaptability to homozygosity in the genome. Furthermore, these results reinforce the importance of understanding inbreeding patterns to judiciously select individuals that will advance to the next generations of self-fertilization, with the aim of obtaining superior and specific hybrids.

Keywords: quantitative-genetics; inbred-lines; genomic-selection.

SÚMARIO

INTRODUÇÃO GERAL.....	13
CAPÍTULO 1 - EFEITOS DA AUTOFECUNDAÇÃO NO GENOMA DE <i>Eucalyptus</i> spp.	23
1.1 INTRODUÇÃO.....	25
1.2 MATERIAL E MÉTODOS.....	27
1.2.1 População estudada.....	27
1.2.2 Controle e qualidade dos SNPs.....	27
1.2.3 Análise de Parentesco.....	28
1.2.4 <i>Runs of Homozygosity</i> (ROH).....	28
1.2.5 Depressão por Endogamia.....	29
1.2.6 Decaimento do Desequilíbrio de Ligação (DL).....	30
1.2.7 Blocos Haplotípicos.....	30
1.3 RESULTADOS.....	30
1.3.1 Taxas de autofecundação e de cruzamento do experimento.....	30
1.3.2 <i>Runs of Homozygosity</i> (ROHs).....	31
1.3.3 <i>Regressão e correlação entre valores fenotípicos e genotípicos e análise de depressão endogâmica</i>	37
1.3.4 <i>Decaimento do desequilíbrio de ligação</i>	38
1.3.5 <i>Blocos Haplotípicos</i>	39
1.4 DISCUSSÃO.....	40
1.4.1 Efeitos da autofecundação no genoma do eucalipto.....	40
1.4.2 Ausência de depressão por endogamia dentro da $S_{0:1}$ de <i>Eucalyptus</i>	42
1.5 CONCLUSÕES.....	43
REFERÊNCIAS	45
CAPÍTULO 2 – SELEÇÃO GENÔMICA APLICADA A UMA POPULAÇÃO AUTOFECUNDADA DE <i>Eucalyptus</i> spp.	50
2.1 INTRODUÇÃO.....	52
2.2 MATERIAL E MÉTODOS.....	54
2.2.1 População estudada.....	54

2.2.2	Associação genômica ampla (Genome-wide analysis -GWAS)	54
		
2.2.3	Controle de qualidade dos SNPs.....	55
2.2.4	Seleção e predição genômica.....	55
2.3	RESULTADOS.....	61
2.3.1	Associação genômica ampla (GWAS).....	61
2.3.2	Parâmetros da predição genômica	63
2.4	DISCUSSÃO.....	65
2.4.1	Associação genômica ampla (GWAS).....	65
2.4.2	Seleção e predição genômica.....	66
2.5	CONCLUSÕES.....	68
	REFERÊNCIAS.....	70
	CONSIDERAÇÕES FINAIS.....	75
	REFERÊNCIAS.....	77

INTRODUÇÃO GERAL

A seleção intrapopulacional consiste na escolha dos melhores indivíduos dentro de uma mesma população, visando aumentar a frequência de alelos favoráveis (Borém; Miranda, 2013), são os métodos amplamente utilizados para o melhoramento de espécies do gênero *Eucalyptus* L'Hér. (Ferreira; Lima, 2018). Seu uso resulta em ganhos genéticos significativos (Assis, 2014). Quando os indivíduos provenientes desses cruzamentos apresentam características de interesse econômico, eles são clonados com o objetivo de aproveitar todos os efeitos presentes na variância genética total (Assis, 1986).

À medida que os plantios comerciais de *Eucalyptus* se concentram cada vez mais no uso de materiais genéticos clonados, a identificação de genótipos superiores torna-se essencial para a manutenção dos ganhos de produtividade (Tambarussi *et al.*, 2022). No entanto, os avanços alcançados pelos métodos tradicionais de melhoramento estão enfrentando um platô, limitando o potencial de ganho adicional (Grattapaglia *et al.*, 2018).

Este platô ocorre em decorrência de diversos fatores como, populações de *Eucalyptus*, apresentam uma ampla base genética, resultando alta variabilidade. Isso é associado a experimentos com baixos tamanhos amostrais, gera *breeding values* pouco acurados, limitando a plena compreensão da variabilidade genética presente nos cruzamentos (Andanov *et al.*, 2017). Portanto, a seleção de indivíduos somente com base na única representação do mesmo, associado a baixa compreensão dos padrões de variabilidade resultam em testes de progênies e seus testes clonais com baixas correlações (Reis *et al.*, 2011, Shalizi; Isik, 2019). Além disso, com o aumento das áreas plantadas e evolução de pragas e doenças que geram perda de produtividade de genótipos, limita o aumento dos ganhos, e criando desafios para atender à crescente demanda de mercado (FAO, 2022).

Portanto, é essencial explorar novos métodos de melhoramento para o gênero *Eucalyptus*, aliando essas abordagens às ferramentas biotecnológicas mais avançadas para a seleção de genótipos superiores, com o objetivo de superar o platô de ganhos genéticos, promovendo maior eficiência na identificação e desenvolvimento de genótipos superiores.

Uma opção viável é a utilização de linhagens endogâmicas, prática amplamente adotada em culturas agrônomicas alógamas, como o milho, desde 1908, a fim de desenvolver híbridos mais produtivos (East, 1908; Shull, 1908; Shull, 1911; Jones, 1918). Este método eficaz explora os efeitos positivos da heterose e a complementariedade genética entre indivíduos contrastantes (Shull, 1909). Contudo, até o momento, essa abordagem não foi aplicada às populações pertencentes ao gênero *Eucalyptus*. A implementação de um programa de melhoramento desse tipo demanda uma compreensão aprofundada dos efeitos da depressão endogâmica e da heterozigosidade nas populações em processo de melhoramento do *Eucalyptus*, aspectos que permanecem sub-explorados e pouco conhecidos pelos melhoristas (Ramalho *et al.*, 2022).

O uso dos marcadores moleculares é uma ferramenta crucial nos programas de melhoramento, uma vez que está mais acessível em termos de custo, facilitando sua aplicação (Grattapaglia, 2022). Dentre os diversos tipos de marcadores disponíveis, os marcadores do tipo SNP (*Single Nucleotide Polymorphism*/ Polimorfismos de Nucleotídeo Único) estão entre os atualmente mais utilizados. Esses marcadores são baseados em alterações de bases únicas na cadeia de bases nitrogenadas (Appleby *et al.*, 2009). Os SNPs são marcadores bialélicos, encontrados em todo o genoma, influenciando caracteres fenotípicos (Grattapaglia, 2017).

O uso destes marcadores em programas de melhoramento do *Eucalyptus* permite a compreensão dos padrões genômicos associados à endogamia, incluindo a depressão induzida por esse processo, a heterozigosidade do genoma, bem como a complementariedade entre genótipos, chamada heterose (Caballero *et al.*, 2021). Portanto, torna-se crucial a utilização para estabelecer novos critérios de seleção e conduzir cruzamentos de eucaliptos de maneira mais eficaz, proporcionando ganhos genéticos mais significativos (Grattapaglia; Kirst, 2008).

Assim, a presente pesquisa teve como objetivo estudar os efeitos da endogamia em uma população autofecundada de *Eucalyptus* ($S_{0.1}$) e com o uso de informações moleculares, identificar e selecionar indivíduos para possíveis cruzamentos visando explorar heterose, bem como para avançar as gerações de autofecundação para a obtenção de linhagens.

REVISÃO BIBLIOGRAFICA

Gênero *Eucalyptus* L'Hér. no Brasil e seu sucesso em programas de melhoramento

O gênero *Eucalyptus* pertence à Família das mirtáceas com mais de 700 espécies catalogadas (Nicolle *et al.*, 2023) e ocorrência natural na Austrália, Timor e Indonésia. No Brasil, o gênero contempla as espécies florestais exóticas de maior sucesso em plantios comerciais com cerca de 7,8 milhões de hectares plantados, com uma produtividade média de $33,7 \text{ m}^3.\text{ha}^{-1}.\text{ano}^{-1}$ (Ibá, 2024), distribuído em diversos locais do Brasil.

Os primeiros plantios de *Eucalyptus* com o objetivo de testar espécies potenciais para plantios em larga escala foram registrados no início dos anos 1900 por Edmundo Navarro de Andrade (Ferreira, 1989), testando espécies consideradas adaptadas para condições edafoclimáticas do Brasil com grande destaque as espécies *Eucalyptus grandis* (Maiden) Hill e o *Eucalyptus saligna* Smith (Pinto Júnior e Silveira, 2021).

Em 1941, Carlos Arnaldo Krug foi convidado para elaborar um programa de melhoramento para o gênero no qual os principais objetivos eram maior uniformidade de plantios, melhorar o fator de forma de indivíduos, maior produção por hectare plantado e reduzir o número de falhas dentro dos plantios (De Assis *et al.*, 2016). Para atingir sucesso nos objetivos propostos foi realizada a seleção de árvores superiores, definição de locais para produção de sementes e hibridação intraespecíficas (Krug;Alves, 1949; Andrade, 1961).

O grande salto nas pesquisas com o gênero ocorreu na década de 1960, visto que devido as leis de incentivos fiscais, diversas cooperativas de pesquisa foram criadas com o objetivo de aumentar a produtividade e a viabilidade dos plantios florestais no Brasil (Gonçalves *et al.*, 2013). Atualmente, o sucesso do gênero se deve a evolução das técnicas silviculturais e dos programas de melhoramento genético, que visam aumento da produtividade, por meio do acúmulo de características de interesse em genótipos superiores (Assis *et al.*, 2015).

Técnicas de melhoramento utilizadas em programas de melhoramento para o gênero de *Eucalyptus* são a Seleção, Hibridação e a Clonagem (Eldridge *et al.*, 1994).

A seleção é fundamental para programas de melhoramento, uma vez que selecionar genótipos superiores e recombiná-los por meio da hibridação, aumenta a frequência de alelos favoráveis e exclusão de alelos desfavoráveis ou deletérios dentro das populações de melhoramento (Falconer; Mackay, 1996).

A seleção recorrente recíproca é o método de seleção mais utilizada dentro dos programas de melhoramento de *Eucalyptus*, na qual vários ciclos de seleção e recombinação são feitos de forma recíproca, com validação dos resultados através dos testes de progênies (Ferreira; Lima, 2018).

O maior exemplo de hibridação dentro do gênero *Eucalyptus* é o híbrido de *Eucalyptus grandis* x *Eucalyptus urophylla* S.T. Blake, o primeiro apresenta rápido crescimento volumétrico, bom fator de forma e alta densidade da madeira, enquanto o segundo apresenta maior tolerância a condições ambientais adversas, boa adaptação a solos mais pobres e maior rusticidade, portanto, a combinação permite concentrar as boas características de ambos em um único genótipo superior (Alfenas *et al.*, 2009).

Com a obtenção de híbridos com características favoráveis, a clonagem se torna uma ferramenta útil e eficiente, onde por meio da propagação vegetativa estes genótipos superiores são replicados, mantendo a produtividade dentro de plantios (Boejarn, 2005).

Como o gênero *Eucalyptus* possui espécies alógamas e com sistema misto de reprodução, populações com estruturas complexas de parentescos, efeitos de endogamia e heterozigosidade são existentes (Griffin; Cotterill, 1988). Com isso, entender o impacto destes efeitos dentro das populações se torna fundamental para programas de melhoramento, visto que são muitas vezes negligenciados, gerando erros de estimativa e acarretando uma diminuição nos ganhos via melhoramento (Resende, 2002; Tambarussi *et al.*, 2018).

Uso da endogamia no melhoramento genético de plantas

O uso da endogamia no melhoramento genético de plantas é um método amplamente utilizado em plantas alógamas (Borém; Miranda, 2013) e culturas de ciclo curto (Crnokrak; Barrett, 2002). Um exemplo de grande sucesso é observado na cultura do milho, na qual, desde 1908, as populações de milho foram submetidas a

sucessivos processos de autofecundação, resultando em um genoma com altas porcentagens de homozigose (Duvick, 2001). Desta forma, o objetivo é a produção de híbridos mais produtivos (East, 1908; Shull, 1908; Shull, 1911; Jones, 1918), por meio do cruzamento de indivíduos contrastantes (Shull, 1909).

No entanto, para a produção de linhagens vigorosas e úteis, a escolha da população base desempenha um papel fundamental (Hallauer *et al.*, 2010). Esta população deve possuir altas frequências de alelos de interesse e baixas frequências de alelos deletérios, a fim de reduzir a depressão por endogamia durante a autofecundação (Charlesworth; Willis, 2009).

As linhagens endogâmicas são obtidas por meio da seleção criteriosa de genitores de alto potencial, que são submetidos a sucessivas autofecundações (Hallauer *et al.*, 2010). Esse processo tem como finalidade promover o aumento da homozigosidade no genoma, com a expectativa de que, após a sexta geração, o nível de homozigose atinja mais de 96% (Jones, 1918).

As autofecundações repetidas permitem a expressão e subsequente eliminação dos alelos deletérios, anteriormente mascarados em estado de heterozigose, resultando na purificação genética da população (Hayward, 2012). Dessa forma, torna-se viável o cruzamento de linhagens contrastantes, otimizando o vigor híbrido decorrente das interações genéticas entre linhagens (Barret; Harder, 1996).

Este método de melhoramento é eficaz, principalmente porque explora a heterose em híbridos simples, nos quais os indivíduos resultantes desses cruzamentos exibem produtividade superior aos seus genitores (Allard, 1971).

A principal dificuldade do uso desse método no melhoramento do gênero *Eucalyptus* spp. é o grande tempo para completar um ciclo da cultura e a baixa produtividade das plantas autofecundadas (Ramalho *et al.*, 2022). No entanto, atualmente, com o avanço dos marcadores moleculares para seleção precoce e o uso de técnicas para florescimento precoce, como o *top grafting* (Castro *et al.*, 2021), esse método se tornou viável dentro da cultura.

A condução de programas de melhoramento que utilizam este método depende de coeficientes de endogamia estimados com precisão, desequilíbrios de ligação e

análise depressão por endogamia. Tornando-se cruciais para a seleção de indivíduos destinados a cruzamentos e para a continuidade de gerações sucessivas de autofecundação (Crute; Pink, 1996; Do Canto *et al.*, 2016).

O desequilíbrio de ligação (DL) desempenha um papel importante, pois se refere a regiões cromossômicas altamente interligadas, sendo fortemente correlacionado à endogamia e ao processo de autofecundação (Norborg, 2000). À medida que os indivíduos se autofecundam, o genoma tende a se fixar, formando grandes blocos de desequilíbrio de ligação, o que restringe a variabilidade entre os loci (Hendrick, 2009). Essa limitação de variabilidade é especialmente importante em linhagens endogâmicas, pois classes genotípicas distintas tornam-se mais evidentes, permitindo o cruzamento entre linhagens contrastantes para explorar a complementariedade genética.

Outro aspecto relevante do desequilíbrio de ligação é seu papel na identificação de blocos haplotípicos (Gabriel *et al.*, 2002). À medida que o desequilíbrio de ligação aumenta no genoma, grandes regiões altamente ligadas (blocos haplotípicos) podem ser detectadas. Essa redução na variabilidade dentro dessas regiões aumenta a eficiência dos modelos de seleção genômica, facilitando a previsão e a escolha de genitores com características desejáveis, o que aprimora a eficácia na melhoria de características de interesse.

A escolha da metodologia mais adequada para estimar a endogamia é essencial, devendo ser considerada a população em estudo e os marcadores moleculares empregados (Caballero, 2021). Diversas abordagens estão disponíveis, sendo fundamental selecionar aquela que melhor reflete as características da população e que maximize a acurácia dos coeficientes de endogamia estimados (Keller *et al.*, 2002).

Dentre as metodologias utilizadas, as *Runs of Homozygosity*, ou corridas de homozigossidade, merecem destaque. A metodologia consiste em localizar segmentos contínuos de homozigose no genoma (Gibson *et al.*, 2006). Embora sejam amplamente utilizados no melhoramento animal, especialmente para estimativas de autozigossidade (Ferenčaković *et al.*, 2011), ou seja, alelos ou regiões endogâmicas que são endogâmicas, mas não decorrente da autofecundação. Essas análises ainda são pouco comuns em plantas. No entanto, elas representam uma ferramenta valiosa

para o melhoramento genético, auxiliando na escolha de indivíduos com altos valores de endogamia, para possíveis cruzamentos visando explorar heterose, bem como auxiliar na descoberta de grupos heteróticos (Purfield *et al.*, 2012).

A análise de ROH pode identificar os níveis de endogamia e os níveis de parentesco entre os indivíduos, especialmente quando se utiliza um elevado número de marcadores codominantes. Isso permite a identificação de alelos compartilhados por descendência (Gibson *et al.*, 2006; Lencz *et al.*, 2007), uma prática comum no melhoramento de gado para orientar os cruzamentos.

Karimi (2013) introduziu o conceito de "Ilhas de ROH" para descrever locais no genoma onde vários indivíduos compartilham a mesma corrida de homozigossidade. Essa abordagem é particularmente útil no melhoramento de plantas, especialmente quando se trabalha com linhagens endogâmicas. Visto que permite a identificação de grupos de indivíduos com regiões homozigotas, pouco afetados pelos efeitos da depressão por endogamia, tornando-os candidatos ideais para serem utilizados em futuros cruzamentos (Barragan *et al.*, 2024).

Uso de marcadores no melhoramento

O fato do gênero *Eucalyptus* ser composto por plantas perenes, ou seja, que possuem ciclos de vida longos e estruturas complexas de parentesco dentro das populações, torna o uso de marcadores moleculares essencial dentro de programas de melhoramento do gênero, acelerando o ciclo de melhoramento e aumentando a eficiência de estimativas (Grattapaglia; Kirst, 2008).

São diversos os tipos de marcadores utilizados, cada um com sua especificidade e uso, dentre os marcadores mais utilizados no melhoramento de espécies florestais são os Microssatélites SSR (*Single Repeat Sequences*/ Sequência de Repetição Simples) e os marcadores do tipo SNP (*Single Nucleotide Polymorphism*/ Polimorfismos de Nucleotídeo Único).

Os Microssatélites SSR, são caracterizados por localizadas grandes sequências repetidas de bases nitrogenadas por todo genoma (Viera *et al.*, 2016). Seus usos são diversos. Permitindo caracterizar indivíduos, apresentar variabilidade de uma população, análises de parentesco, diversidade genética e análises de

associação genômica ampla (genome wide association (GWAS), para associações entre fenótipos e regiões genômicas.

Os marcadores do tipo SNP, cuja principal característica é a alteração de uma única base nitrogenada em um fragmento de DNA homólogo, com ocorrência maior que 1% dentro de uma população (Ganal *et al.*, 2009; VIGNAL *et al.*, 2002), têm uma importância significativa devido aos efeitos que esses marcadores apresentam, para identificar qual é sua influência nas características fenotípicas (Grattapaglia, 2017).

O uso desses marcadores no melhoramento do *Eucalyptus* é diversificado, incluindo a quantificação da variabilidade de populações, a identificação de clones potenciais e a análise de parentesco entre indivíduos (Grattapaglia, 2022), desempenhando um papel crucial nas tomadas de decisão e possibilitando acelerar o ciclo de melhoramento (Collard; Mackill, 2008).

Os marcadores moleculares são empregados em estudos de GWAS, uma ferramenta de extrema importância no melhoramento de plantas. Isso ocorre devido a possibilidade de estabelecer associações entre fenótipos e genótipos, permitindo uma compreensão mais aprofundada dos loci genômicos que exercem um maior efeito sobre as características de interesse (Yang *et al.*, 2013).

Seleção genômica em programas de melhoramento

Sabe-se que os ganhos genéticos são amplamente atribuídos à obtenção de valores genéticos precisos (Hayward, 2012). O resultado do uso bons delineamentos experimentais, da aplicação adequada de métodos de melhoramento, do conhecimento detalhado da genealogia do material estudado, além do uso de marcadores moleculares (Bernardo, 2010). A seleção genômica surge como uma ferramenta poderosa e importante para programas de melhoramento, possibilitando, através de informações fenotípicas e de marcadores moleculares, obter valores genéticos acurados e com maior confiabilidade para as tomadas de decisão do melhorista (Meuwissen *et al.*, 2001).

A seleção assistida por marcadores (SAM) que é usada em programas de melhoramento observa apenas poucos genes de grande efeito, dificultando a associação entre gene e fenótipo de um caráter quantitativo, que é controlado por muitos genes de pequeno efeito (Dekkers; Hospital, 2002; Heffner *et al.*, 2009).

Portanto, a seleção genômica preenche essa lacuna, uma vez que usando uma genotipagem de alta densidade, é possível observar genes de pequeno efeito por todo o genoma, facilitando a associação entre genoma e fenótipo (Heffner *et al.*, 2009).

Assim, por meio da seleção genômica cria-se a possibilidade de prever valores genéticos genômicos (VGG) de indivíduos antes mesmo de testá-los a campo (Peixoto *et al.*, 2024). Por meio de uma população previamente avaliada, com dados fenotípicos e genômicos, conhecida como população de treinamento, os efeitos dos marcadores são testados e conhecidos, possibilitando prever valores genéticos de indivíduos que foram somente genotipados (Meuwissen *et al.*, 2001).

Diversos são os fatores que influenciam a metodologia da seleção genômica. Assim, quantidade de marcadores, controle genético do caráter, número de QTLs ligado ao caráter, desequilíbrio de ligação entre os marcadores e os QTLs, número de indivíduos na população de treinamento, tamanho efetivo, qualidade da genotipagem e parentesco dos indivíduos (Grattapaglia; Resende, 2011).

Os modelos de seleção genômica são diversos, cada um com diferentes inferências e premissas. Entre os modelos frequentistas, os mais utilizados são o RR-BLUP (Meuwissen, 2001) e o GBLUP (Hayes *et al.*, 2009), que se baseiam em modelos mistos. Na inferência Bayesiana, destacam-se os modelos BayesA, BayesB (Meuwissen, 2001), BayesC (Jannick *et al.*, 2009) e o LASSO Bayesiano (De Los Campos *et al.*, 2009). Além disso, modelos de redes neurais também podem ser aplicados na seleção genômica (Haykin, 2001).

Para que um modelo de seleção genômica seja considerado eficiente na predição de valores genéticos de indivíduos não fenotipados, é necessária a estimativa de parâmetros (Gianola; Van Kaam, 2008). Dentre os parâmetros disponíveis, a capacidade preditiva é fundamental, pois consiste na correlação entre os valores preditos e os valores observados, evidenciando a eficiência do modelo em prever os valores fenotípicos de indivíduos não fenotipados com base nos efeitos dos marcadores (Pérez-Rodríguez *et al.*, 2020).

Outro parâmetro relevante é o Erro Quadrático Médio (EQM), que quantifica o desvio médio das predições do modelo em relação aos valores genéticos reais dos indivíduos. Esse indicador é fundamental para avaliar a precisão do modelo e

identificar os erros de predição (Gianola *et al.*, 2021). O *Slope*, que se refere ao desvio dos valores preditos em relação aos observados, ou seja, o quanto os valores diferem dos valores observados, através de uma regressão (Goddard; Hayes, 2009).

Em programas de melhoramento de espécies do gênero *Eucalyptus*, a seleção genômica se bem conduzida, ou seja, com uma população de treinamento de um alto tamanho amostral e uma genotipagem de qualidade, se torna uma ferramenta que traz muitos benefícios para as tomadas de decisão do melhorista (Peixoto, *et al.*, 2024).

Assim, o uso da seleção genômica possibilita a obtenção de valores genéticos mais acurados, maximizando os ganhos de seleção (Denis; Bouvet, 2013). A possibilidade de realizar uma seleção precoce de indivíduos (Resende, 2012), predição de valores genéticos de indivíduos antes de ir a campo e a capacidade de recomendar cruzamento com base nos valores genéticos de indivíduos com maior acurácia também são vantagens que incentivam o uso desta tecnologia (Assis, 2011).

CAPÍTULO 1

EFEITOS DA AUTOFECUNDAÇÃO NO GENOMA DE *Eucalyptus* spp.

RESUMO

Nos programas de melhoramento do gênero *Eucalyptus*, os efeitos aditivos e de dominância são explorados para a seleção de genitores e indivíduos para clonagem, respectivamente. No entanto, é essencial buscar alternativas que aproveitem de maneira mais eficiente os efeitos aditivos. O objetivo da presente pesquisa foi analisar os padrões de endogamia em uma população autofecundada de diferentes espécies de *Eucalyptus* spp. Foram selecionados 20 clones comerciais. A partir da autofecundação desses clones, foram obtidas mudas, que foram divididas em dois grupos, um grupo de indivíduos em um plantio a campo e outro em um pomar de hibridação. Cerca de 30 mudas de cada genitor foram implantadas em um experimento de campo com 30 blocos, com uma planta por parcela. O restante dos indivíduos, totalizando 62, foram plantados em um pomar de hibridação. Tanto os indivíduos quanto seus genitores (682 indivíduos) foram genotipados utilizando o painel customizado de 10.132 SNPs. Adicionalmente, os indivíduos foram avaliados para Diâmetro na Altura do Peito (DAP) aos 36 meses de idade. Para determinar quais indivíduos eram efetivamente frutos de autofecundação, foi realizada uma análise de parentesco. Os coeficientes de endogamia foram estimados com base nas *Runs of Homozygosity* (ROH), tanto para a população como para cada cromossomo, além de se estimar o número de corridas por cromossomo. A depressão por endogamia foi calculada por meio de um modelo misto, utilizando informações fenotípicas e coeficientes de endogamia. O decaimento do desequilíbrio de ligação foi estimado no software TASSEL5, e, por fim, foram identificados os blocos haplotípicos para cada cromossomo na população. Os resultados mostraram que apenas 62,10% dos indivíduos foram comprovadamente autofecundados. Os coeficientes de endogamia (F) variaram de 0 a 0,41, sendo que os cromossomos mais endogâmicos foram o 4 (0,27), 9 (0,39) e 10 (0,28). Não foi observada depressão por endogamia na população, tanto para os indivíduos autofecundados quanto para os cruzados. No entanto, foi identificada uma taxa mais elevada de desequilíbrio de ligação entre os marcadores, decorrente da autofecundação. Os cromossomos 6 (56 blocos) e 8 (51 blocos) apresentaram o maior número de blocos haplotípicos. Os resultados obtidos permitiram compreender melhor os padrões de endogamia no gênero *Eucalyptus*, evidenciando que ele possui resistência à autofecundação e à endogamia, já que mesmo indivíduos autofecundados não apresentam altas taxas de endogamia. Esses resultados serão úteis para o aprimoramento de programas de melhoramento que utilizam linhagens endogâmicas, permitindo selecionar indivíduos para as próximas gerações de autofecundação.

Palavras-chave: endogamia; genética-molecular; marcadores-moleculares.

ABSTRACT

In Eucalyptus breeding programs, additive and dominance effects are exploited for the selection of parents and individuals for cloning, respectively. However, it is essential to seek alternatives that more efficiently leverage additive effects. The objective of the present research was to analyze inbreeding patterns in a self-pollinated population of different Eucalyptus spp. Twenty commercial clones were selected. From the self-pollination of these clones, seedlings were obtained and divided into two groups: one group of individuals in a field experiment and another in a hybridization orchard. Approximately 30 seedlings from each parent were planted in a field experiment with 30 blocks, with one plant per plot. The remaining individuals, totaling 62, were planted in a hybridization orchard. Both the individuals and their parents (682 individuals) were genotyped using a custom panel of 10,132 SNPs. Additionally, the individuals were evaluated for Diameter at Breast Height (DBH) at 36 months of age. To determine which individuals were effectively the result of self-pollination, a parentage analysis was performed. Inbreeding coefficients were estimated based on Runs of Homozygosity (ROH), both for the population and for each chromosome, in addition to estimating the number of runs per chromosome. Inbreeding depression was calculated using a mixed model, utilizing phenotypic information and inbreeding coefficients. Linkage disequilibrium decay was estimated using TASSEL5 software, and finally, haplotype blocks were identified for each chromosome in the population. The results showed that only 62.10% of the individuals were confirmed to be self-pollinated. Inbreeding coefficients (F) ranged from 0 to 0.41, with chromosomes 4 (0.27), 9 (0.39), and 10 (0.28) being the most inbred. No inbreeding depression was observed in the population, neither for self-pollinated nor for cross-pollinated individuals. However, a higher rate of linkage disequilibrium was identified between markers, resulting from self-pollination. Chromosomes 6 (56 blocks) and 8 (51 blocks) presented the highest number of haplotype blocks. The results obtained allowed for a better understanding of inbreeding patterns in the genus Eucalyptus, demonstrating that it exhibits resistance to self-pollination and inbreeding, since even self-pollinated individuals do not show high inbreeding rates. These results will be useful for the improvement of breeding programs that utilize inbred lines, allowing for the selection of individuals for the next generations of self-pollination.

Keywords: inbreeding; molecular-genetics; molecular-markers.

1.1 INTRODUÇÃO

Em programas de melhoramento de plantas, é de fundamental importância compreender os efeitos aditivos, que correspondem à contribuição individual de cada alelo no genoma, independentemente de sua magnitude, e os efeitos de dominância, que refletem na interação entre alelos de um mesmo gene, ou seja, o efeito de um alelo influencia ou mascara o do outro (Falconer; Mackey, 1996). A análise desses efeitos permite selecionar parentais e indivíduos que apresentam elevados valores genéticos, possibilitando a condução de programas de melhoramento que preservam a variabilidade, promovendo ganhos consistentes com o avanço das gerações (Resende, 2002).

No gênero *Eucalyptus*, os efeitos aditivos são explorados na seleção de parentais para cruzamentos, visando acumular alelos de interesse em grupos de indivíduos ou em um único indivíduo, promovendo avanços das gerações de melhoramento (Resende *et al.*, 2017). Por outro lado, quando indivíduos apresentam elevados efeitos de dominância, que resultam em performance superior, eles são preferencialmente selecionados para uso como clones comerciais (Resende; Higa, 2013).

Os efeitos de dominância estão associados principalmente ao estado de heterozigosidade dos locos no genoma, onde a interação entre alelos distintos resulta em uma performance superior em comparação ao estado de homozigosidade, caracterizando o fenômeno da heterose (Lynch; Walsh, 1998). Em contrapartida, os efeitos aditivos podem ser expressos tanto em heterozigotos quanto em homozigotos, uma vez que sua contribuição à característica fenotípica é cumulativa e independente da interação entre os alelos de um loco (Falconer; Mackay, 1996). Contudo, a magnitude dos efeitos aditivos é influenciada pela frequência dos alelos favoráveis, sendo mais evidente em populações que apresentam maior homozigosidade para esses alelos, enquanto os efeitos de dominância se manifestam exclusivamente em locos heterozigotos (Snustad; Simmons, 2015).

Assim, compreender os efeitos da heterozigosidade e da endogamia associados aos sistemas reprodutivos das plantas é essencial para o desenvolvimento de métodos de melhoramento específicos para cada tipo de sistema reprodutivo (Acquah, 2012). No caso de plantas que realizam cruzamentos, como as do gênero *Eucalyptus*, o método que explora os efeitos de endogamia é o método de linhagem

endogâmicas, método desenvolvido por Shull (1909), para a produção de linhagens de milho. Esse método consiste na seleção genitores heterozigotos e realização de sucessivas autofecundações, até que altas taxas de endogamia sejam alcançadas no genoma (Durick, 2001). Após isso, linhagens contrastantes são cruzadas visando explorar a complementariedade genética e os efeitos de heterozigosidade, ou seja, o vigor híbrido decorrente da dominância presente nesses cruzamentos (Hallauer *et al.*, 2010).

O método de linhagens endogâmicas ainda não foi amplamente explorado para o gênero *Eucalyptus* (Ramalho *et al.*, 2022). As populações presentes nos programas de melhoramento do gênero geralmente apresentam uma ampla base genética composta por indivíduos heterozigotos, o que resulta em uma elevada carga genética (Santos *et al.*, 2011). Assim, estes indivíduos apresentam uma presença de alelos deletérios, que quando expressos, podem acarretar menor adaptação e diminuição do desempenho da população (Charlesworth; Willis, 2009). Nesse contexto, o método surge como uma alternativa para reduzir a carga genética por meio da autofecundação, permitindo maximizar os efeitos aditivos positivos desses alelos e, posteriormente os efeitos dominância (Pérez-Pereira *et al.*, 2021). Durante a autofecundação, alelos deletérios são expressos, ocasionando depressão por endogamia, o que possibilita a seleção de indivíduos ou linhagens livres de carga genética (Crnokrak; Barrett, 2002). No momento do cruzamento de linhagens contrastantes, híbridos simples produtivos podem ser gerados, atendendo à demanda de mercado por uma produção florestal (Peréz-López *et al.*, 2014).

Para a condução de programas de melhoramento que utilizam linhagens endogâmicas, ferramentas biotecnológicas, como os marcadores moleculares, são de grande utilidade (Peixoto *et al.*, 2022). Além de auxiliar na seleção de indivíduos, esses marcadores permitem identificar quais indivíduos são comprovadamente autofecundados, estimar coeficientes de endogamia e realizar análises mais precisas da depressão por endogamia (Caballero *et al.*, 2021; Polak *et al.*, 2021). Esses avanços resultam na obtenção de linhagens vigorosas e livres de carga genética (LABROO *et al.*, 2021). No caso do gênero *Eucalyptus*, os marcadores moleculares também podem acelerar as gerações de autofecundações. Onde, por meio do uso destes marcadores permitem realizar a seleção precoce de indivíduos endogâmicos, com boas performances, para induzir ao processo de autofecundação de uma maneira

mais acelerada, superando umas das principais limitações do método para esse gênero (Liu *et al.*, 2004).

Portanto, no presente capítulo objetivou-se realizar análises de parentesco, coeficiente de endogamia e depressão por endogamia, em uma população autofecundada de *Eucalyptus* spp., utilizando marcadores moleculares do tipo SNP. Visa-se com esse estudo observar os efeitos da endogamia na população como um todo e nos indivíduos, resultantes do processo de autofecundação.

1.2 MATERIAL E MÉTODOS

1.2.1 População estudada

A pesquisa foi realizada em parceria com a Suzano S.A., onde foram selecionadas um total de 20 genitores superiores para o estudo, sendo 11 provenientes do cruzamento entre *Eucalyptus urophylla* x *Eucalyptus grandis* (“urograndis”), cinco de *Eucalyptus urophylla*, um de *Eucalyptus grandis* e três de *Eucalyptus* spp.. Esses genitores foram submetidos à autofecundação, para obtenção das sementes. Com essas sementes foram produzidas mudas. Aproximadamente 30 mudas de cada uma das progênies foram utilizadas, totalizando 600 indivíduos para serem conduzidas para um experimento a campo e o restante das mudas foram conduzidas para um pomar de hibridação, totalizando 62 indivíduos. No experimento a campo o delineamento experimental utilizado foi o de blocos ao acaso com uma planta por parcela, foram implantados oito genitores, dos 20 selecionados, de diferentes espécies, e cinco clones comerciais como testemunhas do experimento.

Para os indivíduos presentes a campo foi feita uma fenotipagem para o diâmetro na altura do peito (DAP) aos 36 meses de idade. Já para a genotipagem foi utilizando o 60K SNP chip (Silva-Junior; Faria; Grattapaglia, 2015) tanto para os indivíduos presentes a campo como no pomar de hibridação e seus respectivos genitores, totalizando 682 indivíduos.

1.2.2 Controle de qualidade dos SNPs

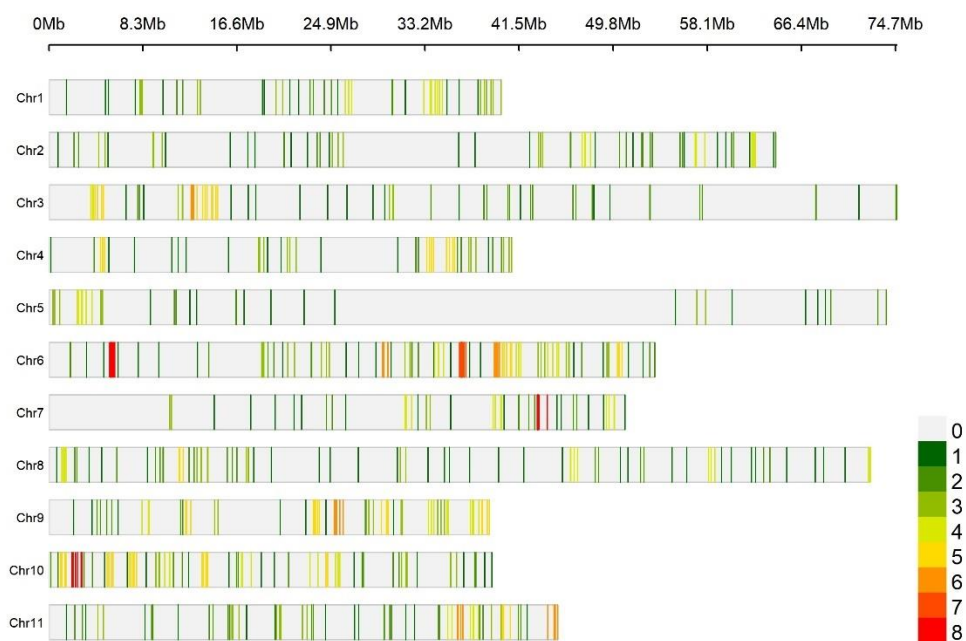
Para construção das matrizes de seleção genômica e para a estimativa dos valores de endogamia, alguns marcadores foram excluídos com base nos critérios de qualidade: MAF (*Minor allele Frequency*) de 0.05 e uma *Call Rate* de 0.95 (Granato *et al.*, 2018). Dessa forma, marcadores com frequências menores que 5% dentro da

população e indivíduos com mais de 5 % de marcadores faltantes foram removidos para evitar viés nas estimativas.

1.2.3 Análise de Parentesco

A análise de paternidade foi realizada no software Cervus 3.0.7 (Marshall *et al.*, 1998), com base na estatística Δ , definida pela diferença de Lod Score entre o primeiro e segundo candidatos a pai de cada genótipo testado. Simulações com 10.000 repetições foram realizadas com o objetivo de determinar o grau de significância da estatística Δ , aceitando-se níveis de confiança de 80%. Foram utilizados 723 marcadores para análise (Figura 1), sendo os marcadores mais polimórficos do genoma, com taxa de erro de genotipagem aceito de 5%. Onde faixas mais vermelhas indicam maior quantidade de marcadores selecionados na região. Esses marcadores foram selecionados por sua alta polimorfia na população, reduzindo o risco de associações falsas entre genitores e indivíduos autofecundados ou cruzados.

Figura 1 - Posição no genoma dos 723 marcadores polimórficos utilizados na análise de parentesco



1.2.4 Runs of homozygosity (ROH)

O efeito da endogamia foi avaliado pela identificação das *Runs of Homozygosity* (ROH) no genoma (Curik; Ferenčaković; Sölkner, 2014), utilizando o pacote “detectRUNS” (Biscarini *et al.*, 2019) no software R 4.3.2 (R Core Team, 2023). As

ROH foram identificadas por meio do método de janela deslizante (BJELLAND *et al.*, 2013).

Dos 682 indivíduos que estavam presentes no experimento, apenas 461 foram genotipados. Portanto, para mitigar o viés da análise, foram selecionados exclusivamente os 350 indivíduos autofecundados. Além disso, foram utilizados todos os 10.132 marcadores moleculares disponíveis para a análise. Foram estimados os valores de endogamia para cada indivíduo pelo método da ROH (F_{ROH}), dado por:

$$F_{ROH} = \frac{\sum L_{ROH}}{L_{genoma}} \quad (1)$$

Onde L_{ROH} é a distância dos seguimentos de homoziguidade e L_{genoma} é a distância total do genoma.

Foi estimado o coeficiente de endogamia e sua variação para cada cromossomo, com o objetivo de identificar qual cromossomo era mais endogâmico e quais os possíveis indivíduos que se desviassem da média de endogamia.

Um Manhattan plot foi desenvolvido com o objetivo de observar a frequência de indivíduos da população que apresentaram o determinado marcador em estado de homoziguidade para cada cromossomo

A partir dos valores de endogamia individual (F_{ROH}) foi realizada uma regressão entre coeficientes de endogamia e valores de DAP, para avaliar a presença ou ausência de depressão endogâmica na população.

1.2.5 Depressão por endogamia

Por meio da metodologia de Runs of Homozygosity (ROH) e dos valores de F_{ROH} , foi realizada a regressão entre valores de DAP e os valores de F_{ROH} para analisar os valores de correlação de Pearson (ρ), indicativo da presença ou ausência de depressão endogâmica, utilizando o seguinte modelo misto:

$$Y_{ij} = \mu + \alpha_i + r_j + bF_{ijk} + \varepsilon_{ijk} \quad (2)$$

Onde Y_{ij} é a variável dependente; μ é a média geral; α_i é o efeito aleatório da progênie i ; r_j é o efeito fixo do bloco j , b é o coeficiente de regressão no coeficiente de endogamia F_{ijk} e ε_{ijk} é o erro aleatório. O efeito de progênie α_i foi assumido com distribuição normal $N(0, G\sigma_s^2)$, onde G é a matriz de relação genômica realizada, calculada segundo VanRaden (2008), sendo σ_s^2 a variância da progênie.

1.2.6 Decaimento do desequilíbrio de ligação (DL)

Os desequilíbrios de ligação foram estimados com base na diferença (D) entre a frequência observada e a esperada de dois gametas. Quanto maior a diferença, mais ligada estão os alelos (Mueller, 2004). O desequilíbrio de ligação (r^2) entre pares de SNPs, para todos os 10132 marcadores da população de estudos foram obtidos no software TASSEL 5 (Bradbury *et al.*, 2007). O gráfico de decaimento do desequilíbrio de ligação em função da distância em pares de base (bp) foi elaborado usando os pacotes base do software R versão 4.3.2 (R Core Team, 2023).

1.2.7 Blocos Haplotípicos

Os blocos haplotípicos, para todos os 10132 marcadores, foram definidos com base no algoritmo de intervalo de confiança de Gabriel *et al.* (2002), utilizando o software Haploview v. 4.1 (Barret *et al.*, 2005). Os pares de SNPs foram considerados em forte desequilíbrio de ligação (DL) se o valor do limite superior do intervalo de confiança for maior que 0,98 do coeficiente de desequilíbrio normalizado (D') e se o limite inferior possuir um mínimo de 0,70. O D' entre os locos A e B será estimado como $D'_{AB} = D/D_{Max}$, onde D é calculado como $D = p_{A1B1}p_{A2B2} - p_{A1B2}p_{A2B1}$ e D_{Max} :

$$D_{Max} \begin{cases} -\min\{p_{A1}p_{B1}, p_{A2}p_{B2}\}, & \text{quando } D < 0 \\ -\min\{p_{A1}p_{B2}, p_{A2}p_{B1}\}, & \text{quando } D \geq 0 \end{cases} \quad (3)$$

As posições físicas de cada SNP foram determinadas de acordo com o mapa de consenso do genoma de *Eucalyptus grandis*, versão 2.0. A extensão dos blocos também foi estimada como o quadrado da correlação de frequência alélica (r^2). O valor crítico de r^2 foi calculado seguindo o método descrito por Breseghello e Sorrells (2006).

1.3 RESULTADOS

1.3.1 Taxas de autofecundação e de cruzamentos do experimento.

Com base na análise de paternidade, 62,10% dos indivíduos foram identificados como autofecundados, enquanto 37,90% resultaram de cruzamentos (Tabela 1). A menor quantidade de sementes obtida dos genitores, que deveriam idealmente produzir 30 sementes cada, decorre tanto de uma elevada taxa de mortalidade quanto da impossibilidade de extração de DNA. Essa diferença dos níveis de autofecundação e cruzamentos pode ser atribuída a uma possível contaminação ocorrida durante o processo de autofecundação, ou seja, a entrada de pólen externo não desejado no momento da autofecundação. Mas, isso pode ser uma característica do genótipo entre ter uma resistência durante o processo de autofecundação.

Tabela 1 - Taxas de autofecundação e cruzamentos nas 20 progênes de *Eucalyptus* spp..

Genitor	Autofecundados	Cruzados	Genitor	Autofecundados	Cruzados
1	19	-	11	29	4
2	11	2	12	32	-
3	28	3	13	8	12
4	6	21	14	17	-
5	1	32	15	13	18
6	9	24	16	8	21
7	26	3	17	14	-
8	15	2	18	13	20
9	10	9	19	15	1
10	27	3	20	12	16
Autofecundados (%)			62.10		
Cruzados (%)			37.90		

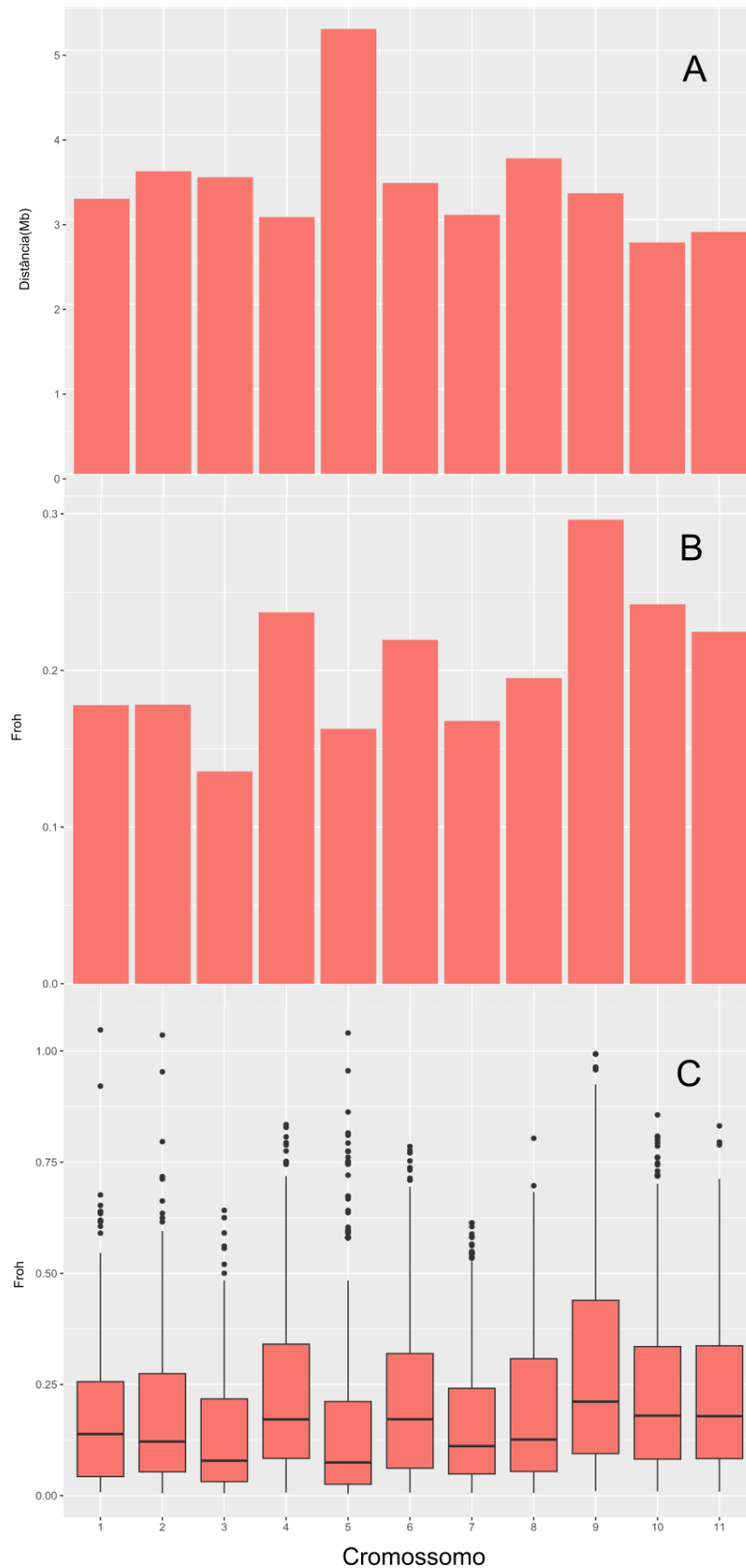
1.3.2 *Runs of Homozygosity* (ROHs)

A partir dos valores de *Runs of homozygosity* (ROHs), foi possível observar que, entre os 11 cromossomos avaliados, o cromossomo cinco apresentou a maior distância média de ROHs (5,3 Mb) no genoma da população autofecundada de *Eucalyptus* spp. (Figura 2-A).

Esses resultados indicam que os indivíduos da população não apresentam um alto nível de endogamia. Em populações com níveis mais elevados de endogamia, seria esperado um maior número de ROHs, bem como ROHs mais longas, que abrangeriam mais alelos e cobririam uma porção maior do genoma em homozigose. Os valores de F_{ROH} variaram de 0,16 a 0,39 entre os 11 cromossomos, destacando-se os cromossomos 4 (0,27), 9 (0,39) e 10 (0,28), que apresentaram os maiores níveis de homozigose (Figura 2-B).

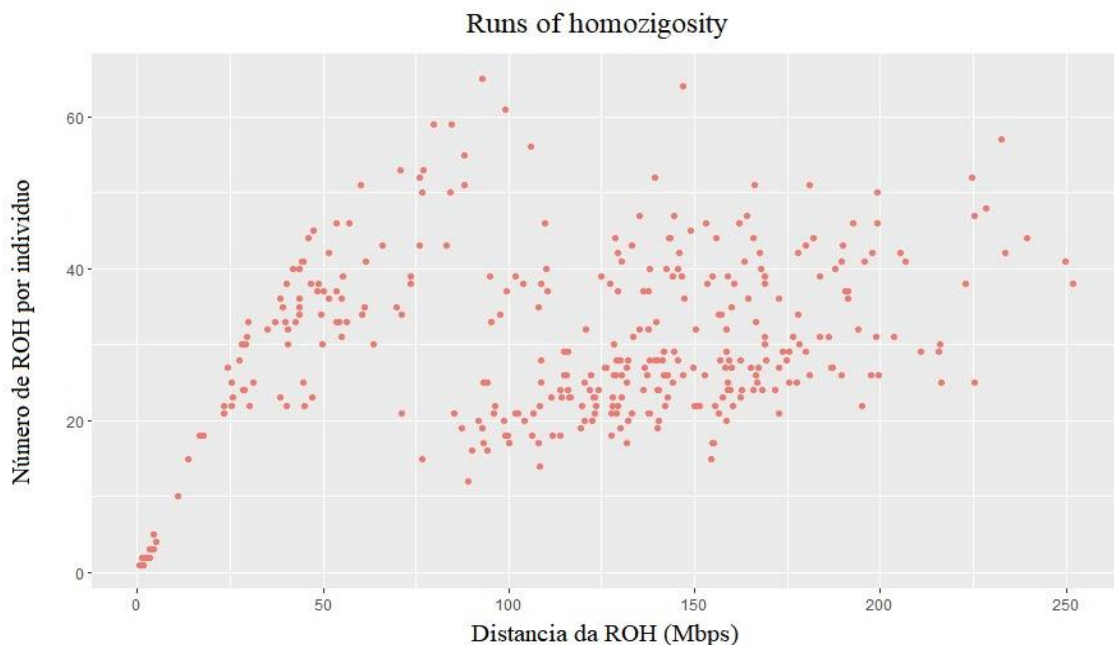
Foi possível observar alta taxa de variação para os coeficientes de endogamia para cada cromossomo permitindo identificar, de forma mais precisa, indivíduos com taxas mais elevadas de homozigosidade, mesmo em cromossomo mais heterozigotos (Figura 2-C).

Figura 2 - Tamanho médio das *Runs of Homozigosity* (ROH) em Megabases (Mb) (A), coeficientes de endogamia (F_{ROH}) (B) e sua variação (C), na população autofecundada de *Eucalyptus* spp.



Os indivíduos da população apresentaram, em média, de 20 a 40 ROHs, com distâncias que variaram principalmente entre 100 e 150 Mega bases (Mbps) (Figura 3), indicando indivíduos com comportamento nos segmentos de homozigosidade de forma variada. É possível observar indivíduos autofecundados com poucas corridas e com menores distância, bem como com muitos segmentos e maiores distancias, sendo um indicativo de alta variabilidade no coeficiente de endogamia dentro da população.

Figura 3 - Distribuição do número e tamanho das Runs of Homozygosity (ROHs) por indivíduo em uma população autofecundada de *Eucalyptus* spp.



Os cromossomos 8, 11 e 6, com 1312, 1136 e 1019 segmentos respectivamente, apresentaram os maiores números de ROHs (Tabela 2). No entanto, embora o cromossomo 9 tenha exibido um número inferior de ROHs (973), ele apresentou o maior valor de F_{ROH} . Isso ocorre porque, apesar do menor número de ROHs, o tamanho médio dessas regiões no cromossomo 9 é mais elevado em comparação aos demais, resultando em valores mais altos de F_{ROH} (Tabela 2).

Tabela 2 - Número de *Runs of Homozygosity* (ROHs) por cromossomo em *Eucalyptus* spp..

Cromossomo	Nº de ROHs
1	594
2	1012
3	965
4	995
5	649
6	1019
7	902
8	1312
9	973
10	996
11	1136

Os indivíduos com os maiores valores de F_{ROH} foram IND063 (0,416), IND049 (0,413) e IND342 (0,396) (Tabela 3). Os valores de F_{ROH} variaram de 0 a 0,41, sendo que valores iguais a zero indicam indivíduos altamente heterozigotos, nos quais nenhuma região do genoma atende aos critérios para definição de uma ROH.

Tabela 3 - Os 20 indivíduos com os maiores coeficientes de endogamia estimados pelo método de *Runs of Homozygosity* (F_{ROH}) em *Eucalyptus* spp..

Indivíduo	F_{ROH}	Genitor
IND063	0.417	GEN11
IND049	0.413	GEN28
IND342	0.396	GEN5
IND272	0.387	GEN25
IND371	0.385	GEN22
IND435	0.378	GEN20
IND191	0.373	GEN26
IND294	0.373	GEN19
IND207	0.372	GEN8
IND447	0.369	GEN20
IND040	0.358	GEN28
IND367	0.358	GEN22
IND142	0.357	GEN13
IND092	0.349	GEN12
IND115	0.342	GEN13
IND345	0.340	GEN5
INDP048	0.337	GEN25
IND123	0.330	GEN13
IND219	0.330	GEN8
IND116	0.330	GEN13

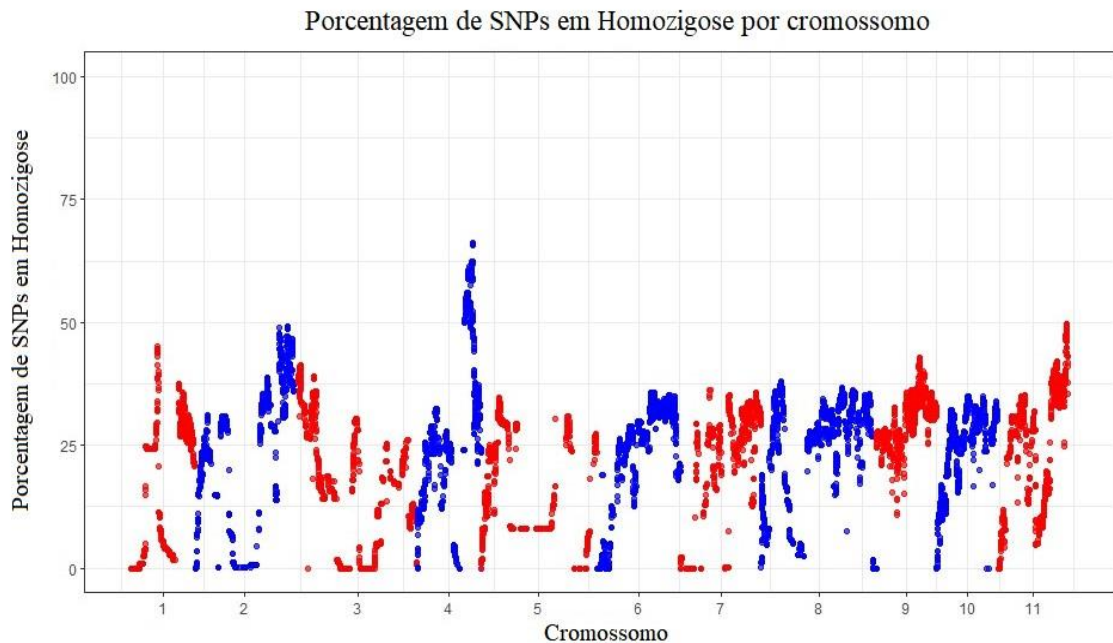
Os valores individuais de F_{ROH} dos genitores (Tabela 4) variaram de 0,0007 a 0,09, indicando que esses indivíduos apresentam alta heterozigosidade por todo seu genoma. Esse resultado era esperado, considerando que os genitores são provenientes de cruzamentos.

Tabela 4 - Coeficientes de endogamia estimados pelo método de *Runs of Homozygosity* (F_{ROH}) dos genitores avaliados em uma população autofecundada de *Eucalyptus* spp..

Genitores	F_{ROH}
GEN6	0.094
GEN19	0.052
GEN26	0.047
GEN9	0.042
GEN23	0.042
GEN20	0.018
GEN28	0.008
GEN16	0.007
GEN18	0.007
GEN15	0.005
GEN5	0.005
GEN24	0.005
GEN27	0.005
GEN13	0.004
GEN10	0.003
GEN17	0.002
GEN11	0.002
GEN22	0.001
GEN14	0.001
GEN2	0.001
GEN4	0.001

Observou-se uma variação na frequência de SNPs em homozigose ao longo dos cromossomos, com destaque para o cromossomo 9, que apresentou uma maior abundância de SNPs em estado de homozigose em quase todo o seu comprimento, com uma taxa de variação entre 14% a 45%. No caso do cromossomo 4, foi identificado que os SNPs eram majoritariamente homozigotos na população, com alguns marcadores atingindo até 70% de homozigosidade. Entretanto, dentro do mesmo cromossomo, foram observados SNPs frequentemente heterozigotos, indicando uma possível resistência à endogamia nessas regiões (Figura 4).

Figura 4 - Porcentagem de homozigose por SNP na população autofecundada de *Eucalyptus* spp

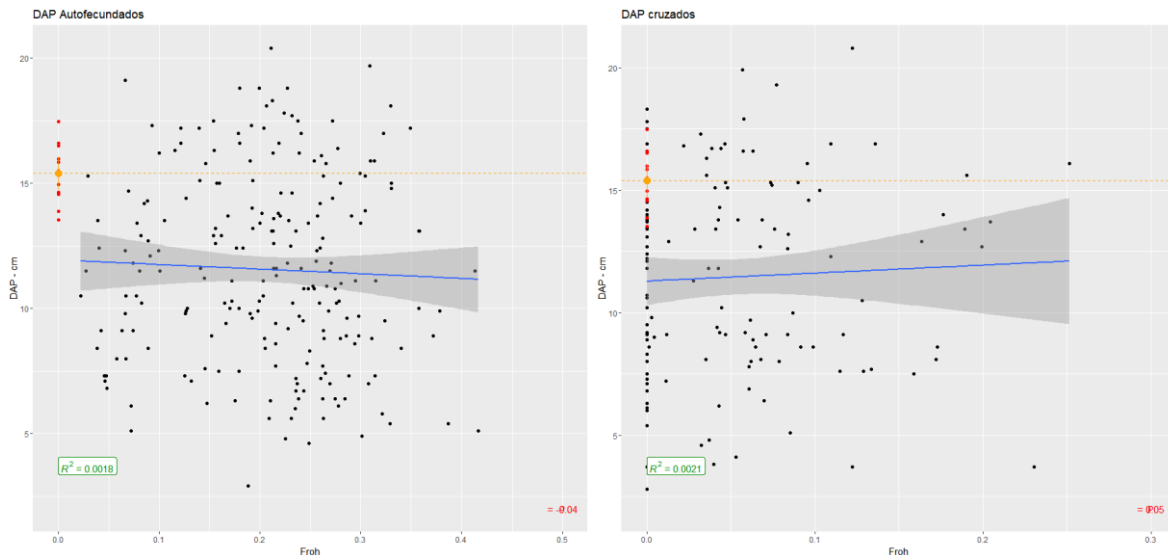


1.3.3 Regressão e correlação entre valores fenotípicos e genotípicos e análise de depressão endogâmica

Não foi identificada uma depressão por endogamia evidente entre os indivíduos autofecundados e cruzados na população avaliada para o DAP aos 36 meses (Figura 5). Isso se deve à baixa correlação entre valores de endogamia e a performance fenotípica, com coeficientes próximas de zero tanto para indivíduos autofecundados (-0,04) quanto para indivíduos cruzados (0,05).

A nível de população a depressão por endogamia não foi aparente, mas, é possível observar, que há a presença de indivíduos com boas e más performances a medida que as taxas de endogamia se elevam.

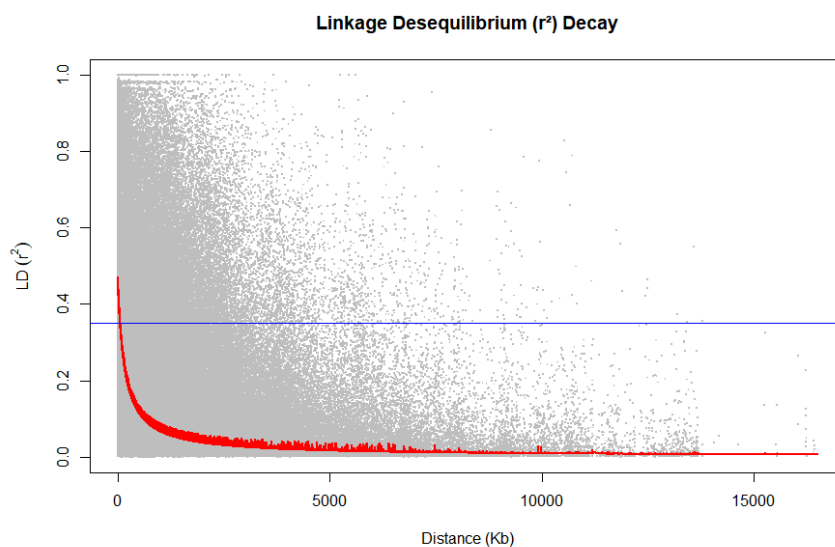
Figura 5 - Regressão entre os valores fenótipos e coeficientes de endogamia (F_{ROH}) para o diâmetro à altura do peito (DAP) aos 36 meses de idade em uma população autofecundada de *Eucalyptus* spp..



1.3.4 Decaimento do desequilíbrio de ligação

O decaimento do desequilíbrio de ligação foi observado abaixo de 0,35 a uma distância de 300 kilobases (Kb), indicando que o aumento da endogamia está associado a um maior desequilíbrio de ligação na população (Figura 6).

Figura 6 - Decaimento do desequilíbrio de ligação (r^2) em uma população autofecundada de *Eucalyptus* spp..



1.3.5 Blocos haplotípicos

Entre os 11 cromossomos avaliados, os cromossomos 6 e 8 apresentaram a maior quantidade de blocos haplotípicos, sendo esses 56 e 51, respectivamente. No entanto, os blocos de maior extensão em pares de bases (acima de 100 Kb) foram identificados em regiões mais endogâmicas do genoma.

Nos cromossomos 9 e 4 apresentam os valores mais elevados dos coeficientes de endogamia (Figura 2), mas isso não fez com que estas regiões apresentassem número de blocos elevados, portanto, as taxas elevadas de endogamia nesses cromossomos estão associadas a blocos de maiores dimensões como no cromossomo 4 que apresentou uma distância de bloco de 499 kb (Tabela 5).

Tabela 5 - Número de blocos haplotípicos para cada cromossomo para indivíduos de autofecundação de *Eucalyptus* spp..

Cromossomo	Blocos	Distância		Nº de SNPs		F_{ROH}
		Min (Kb)	Max (Kb)	Max	Min	
1	41	0	494	6	2	0,192
2	36	0	492	8	2	0,215
3	35	0	425	5	2	0,164
4	24	0	499	7	2	0,272
5	27	0	100	4	2	0,181
6	56	0	417	14	2	0,263
7	36	0	498	13	2	0,194
8	51	0	442	6	2	0,235
9	48	0	380	9	2	0,353
10	42	0	478	16	2	0,282
11	38	0	472	6	2	0,253

F_{ROH} = coeficiente de endogamia.

1.4 DISCUSSÃO

1.4.1 Efeitos da autofecundação no genoma do eucalipto

A autofecundação gerou efeitos significativos no genoma do *Eucalyptus*, promovendo o aumento na homoziguidade, o que resultou em um incremento nos desequilíbrios de ligação entre os marcadores e na formação de blocos haplotípicos (Hamrick; Godt, 1996; Wright *et al.*, 2008). As taxas de endogamia individual (0 a 0,41) (Tabela 3), estimadas por meio das ROHs (Froh) indicam que, em média, o genoma do *Eucalyptus* não atinge, em média, taxas de homoziguidade superiores a 0,50 já na primeira geração de autofecundação (Ingvarsson, 2002).

Hedrick *et al.* (2015) avaliaram progênies autofecundadas de *Eucalyptus grandis*, onde as taxas médias de heteroziguidade foram de 65,5%, indicando que a homoziguidade foi menor do que o esperado para a primeira geração de autofecundação. Esses resultados são semelhantes aos encontrados na presente pesquisa. Como o gênero *Eucalyptus* apresenta naturalmente altas taxas de cruzamentos, a espécie possui mecanismos evolutivos, tanto pré quanto pós-zigóticos, que dificultam a autofecundação e, conseqüentemente, a expressão de homoziguidade no genoma (Myburg *et al.*, 2014).

Por meio da análise de ROH, foi possível observar que algumas regiões cromossômicas apresentam um maior número de SNPs em estado de homoziguidade, como no cromossomo 4, onde quase 70% dos indivíduos autofecundados possuem SNPs homozigotos nessa região (Figura 6) considerando tanto indivíduos autofecundados como seus genitores. Hendrick *et al.* (2015) observaram que, no gênero *Eucalyptus*, a autofecundação exerce uma forte pressão de seleção em locos homozigotos. Por exemplo considerando um genitor com o genótipo A_1A_2 , durante a autofecundação há uma alta predominância dos genótipos A_1A_1 e A_1A_2 , indicando que o alelo A_2 seja deletério e, portanto, não contribui para a formação de zigotos devido à forte depressão por endogamia. Dessa forma, a região do cromossomo 4 provavelmente passou por uma alta pressão de seleção, resultando em uma elevada quantidade de marcadores endogâmicos nessa área.

O cromossomo 9 apresentou os valores mais elevados de endogamia quando comparado aos demais (Figura 4), possivelmente devido à presença de marcadores altamente ligados nesse cromossomo, o que facilitou a fixação de alelos e resultou em maiores taxas de homoziguidade (Charlesworth; Charlesworth, 1999). As ROHs

identificadas variaram de 20 a 40 segmentos por indivíduo, com tamanhos entre 100 e 200 Mbps (Figura 3). Com o avanço das gerações de autofecundação, espera-se que mais alelos sejam fixados em estado de homozigose, levando ao aumento do tamanho das ROHs e do número de segmentos por indivíduo (Haudry *et al.*, 2008).

Altas taxas de endogamia, associadas a elevados desequilíbrios de ligação, são essenciais para o uso eficiente das linhagens endogâmicas, pois permitem uma diferenciação mais clara entre genótipos contrastantes, facilitando a formação de grupos heteróticos (Slotte *et al.*, 2010). Além disso, a autofecundação contribui para o aumento dos níveis de desequilíbrio de ligação no genoma, uma vez que, durante a formação dos gametas, os alelos tendem a recombinar entre si e se tornar altamente ligados devido a fatores como recombinação e pressões de seleção, levando ao desequilíbrio de ligação (Hartl; Clark, 1997, Gordo; Charlesworth, 2001).

A autofecundação também resulta em uma redução do tamanho efetivo da população (N_e), especialmente em espécies de plantas com alta taxa de cruzamento (Kageyama, 1981). Essa redução pode acelerar o fenômeno de deriva genética, favorecendo a fixação de alelos que, sob condições normais de cruzamento, não seriam fixados, levando à formação de blocos de desequilíbrio de ligação maiores (Wright *et al.*, 2005). Esses alelos fixados contribuem para a criação de grandes blocos haplotípicos no genoma, anteriormente ausentes, devido ao aumento da homozigosidade promovida pela autofecundação (Escobar *et al.*, 2010).

Os blocos haplotípicos observados apresentaram distâncias que variaram de 0 a 494 kb. Estudos anteriores, como os realizados por Melchinger (1999) e Slotte *et al.* (2010), mostraram distâncias semelhantes para blocos haplotípicos em espécies do gênero *Eucalyptus*, confirmando altos níveis de desequilíbrio de ligação entre os marcadores na população estudada.

Portanto, as estimativas de endogamia, combinadas com os desequilíbrios de ligação e os blocos haplotípicos, fornecem uma compreensão detalhada dos efeitos da endogamia no genoma, bem como identificar indivíduos com os maiores níveis de homozigosidade. Esse conhecimento, aliado à seleção genômica, permite selecionar os melhores candidatos para avançar nas próximas gerações de autofecundação, otimizando o processo de melhoramento genético (Wang *et al.*, 2018).

1.4.2 Ausência de depressão por endogamia dentro da $S_{0:1}$ de *Eucalyptus*

Ausência de depressão por endogamia na primeira geração de autofecundação também foram relatados por Pope e Coop (2020), Demontis e Vuilleumier (2021), e Abu Awad e Roze (2020), indicando que, em alguns casos, os efeitos negativos da endogamia não são imediatamente evidentes nessa fase inicial. Vale ressaltar que a depressão por endogamia ocorre diferentemente para cada caráter, população e idade do caráter avaliado (Tambarussi *et al.*, 2017).

Como os genitores eram clones, ou seja, indivíduos com alta heterozigosidade, a primeira geração de autofecundação ainda apresenta uma taxa relativamente elevada de heterozigosidade (Byers; Waller, 1999). No caso do gênero *Eucalyptus*, o genoma na primeira geração de autofecundação apresenta uma taxa de homozigosidade relativamente baixa, não alcançando os 50% esperados de genoma homozigoto (Hendrick *et al.*, 2015). Nesse contexto, muitos dos alelos deletérios presentes em locos heterozigotos permanecem mascarados pelos efeitos da dominância (Charlesworth; Willis, 2009). Portanto, como não há um conjunto de alelos deletérios suficientemente expressos dentro da população da presente pesquisa, esta acaba por não apresentar uma depressão por endogamia (Lande *et al.*, 1985).

Além dos baixos coeficientes de endogamia, uma das causas da ausência de depressão por endogamia na população seria a baixa carga genética dos genitores selecionados. Mas, mesmo com a ausência dentro da população é possível observar que existem indivíduos que apresentam alta depressão por endogamia, enquanto outros não, o que pode ser atribuído à menor carga genética (Maldonado; Miranda Filho, 2002). A redução na presença de alelos deletérios nesses indivíduos permite que, mesmo com taxas mais elevadas de endogamia, eles apresentam boas performances fenotípicas (García-Dorado, 2015).

Com o aumento do número de gerações de autofecundação, o genoma dos indivíduos apresentará taxas mais elevadas de homozigosidade, o que permitirá a expressão de alelos antes 'mascarados', resultando em depressão por endogamia (Crnokrak; Barrett, 2002). Hendrick *et al.* (2015), observou que após o ciclo de desenvolvimento aos sete anos, apenas 18 de 120 indivíduos sobreviveram. A ausência de depressão por endogamia apresenta um aspecto positivo, onde é possível selecionar indivíduos para dar sequência às próximas gerações de autofecundação, em que umas das dificuldades do método de linhagens endogâmicas, é encontrar linhagens vigorosas (Charlesworth; Charlesworth, 1987).

Como as taxas de endogamia dos indivíduos variaram de 0 a 0,41 (Tabela 3), associado a ausência da depressão por endogamia, além de selecionar indivíduos para futuras autofecundações, cria-se a possibilidade de selecionar indivíduos contrastantes para realizar cruzamentos, onde permite identificar genitores com bons alelos para cruzamentos subsequentes, visando explorar a heterose, e identificar as melhores combinações entre genótipos para as futuras gerações de autofecundação (Falconer; Mackay, 1996; Hallauer *et al.*, 2010).

1.5 CONCLUSÕES

Com a autofecundação na população estudada, por meio dos coeficientes de endogamia e da análise de depressão por endogamia, foi possível observar e compreender os padrões de endogamia e heterozigosidade que o gênero *Eucalyptus* possui, que apesar de uma resistência à endogamia, aspectos como ausência de depressão endogâmica na população e indivíduos com boas performances mesmo endogâmicos, portanto livres de cargas genéticas, são positivos para a condução de programas de melhoramento do gênero que utilizem os método de linhagens endogâmicas.

REFERÊNCIAS

- ABU AWAD, D.; ROZE, D. Epistasis, inbreeding depression, and the evolution of self-fertilization. **Evolution**, v. 74, n. 7, p. 1301-1320, 2020.
- BARRETT, J. C.; FRY, B.; MALLER, J.; DALY, M. J. Haploview: analysis and visualization of LD and haplotype maps. **Bioinformatics**, 21(2), 263–265, 2005.
- BISCARINI, F.; COZZI, P.; GASPA, G.; MARRAS, G. detectRUNS: an R package to detect runs of homozygosity and autozygosity in diploid genomes. **BMC Bioinformatics**, v. 20, n. 1, p. 1–13, 2019.
- BJELLAND, D.W.; WEIGAL, K.A.; VUKASINOVIC, N.; NKRUMAH, J.D. Evaluation of inbreeding depression in Holstein cattle using whole-genome SNP markers and alternative measures of genomic inbreeding. **Journal of dairy science**, v. 96, n. 7, p. 4697-4706, 2013.
- BRADBURY, P. J.; ZHANG, Z.; KROON, D. E.; CASSTEVENS, T. M.; RAMDOSS, Y.; BUCKLER, E. S. TASSEL: Software for association mapping of complex traits in diverse samples. **Bioinformatics**, v. 23, n. 19, p. 2633–2635, 2007.
- BRESEGHELLO, F.; SORRELLS, M. E. Association Mapping of Kernel Size and Milling Quality in Wheat (*Triticum aestivum* L.) Cultivars. **Genetics**, v. 172, p. 1165-1177, 2006.
- BYERS, D. L.; WALLER, D. M. Do plant populations purge their genetic load? Effects of population size and mating history on inbreeding depression. **Annual Review of Ecology and Systematics**, v. 30, p. 479-513, 1999.
- CABALLERO, A.; VILLANUEVA, B.; DRUET, T. On the estimation of inbreeding depression using different measures of inbreeding from molecular markers. **Evolutionary Applications**, v. 14, n. 2, p. 416-428, 2021.
- CHARLESWORTH, D.; CHARLESWORTH, B. Inbreeding depression and its evolutionary consequences. **Annual Review of Ecology and Systematics**, 18, 237–268, 1987.
- CHARLESWORTH, D.; CHARLESWORTH, B. Genetic linkage and molecular evolution. **Current Biology**, v. 11, p. R684-R686, 1999.
- CHARLESWORTH, D.; WILLIS, J. H. The genetics of inbreeding depression. **Nature Reviews Genetics**, v. 10, p. 783-796, 2009.
- CRNOKRAK, P.; BARRETT, S. C. H. Perspective: Purging the genetic load: a review of the experimental evidence. **Evolution**, v. 56, n. 12, p. 2347-2358, 2002.
- CURIK, I.; FERENČAKOVIĆ, M.; SÖLKNER, J. Inbreeding and runs of homozygosity: A possible solution to general problems in animal genetics. **Animal Genetics**, Oxford, v. 45, n. 1, p. 10-19, 2014.
- DEMONTIS, C. G. B.; VUILLEUMIER, A. Inbreeding depression under mixed outcrossing, self-fertilization, and sib-mating. **BMC Ecology and Evolution**, 21, 51, 2021.

DUVICK, D. N. Biotechnology in the 1930s: the development of hybrid maize. **Nature Reviews Genetics**, 2(1), 69-74, 2001.

ESCOBAR, J.S. An integrative test of the dead-end hypothesis of selfing evolution in Triticeae (Poaceae). **Evolution**, v. 64, p. 2855-2872, 2010.

FALCONER, D. S.; MACKAY, T. F. C. **Introduction to Quantitative Genetics (4th ed.)**. Longman Group Ltd, 1996.

GABRIEL, S. B.; SCHAFFNER, S. F.; NGUYEN, H.; MOORE, J. M.; ROY, J.; BLUMENSTIEL, B.; DEFELICE, M.; LOCHNER, A.; FAGAN, T.; LICUL, C.; HUBER, W.; LUNETTA, K.; DIETZCH, E.; ROHLAND, N.; RIEGER, L.; LAIRD, N.; CHUEH, H.; RAYMOND, D.; ARDLEIGH, P.; FELDMAN, H.; TAJIMA, R.; TANYA, T.; ZIAUGRA, L.; SHERMAN, L.; ADLER, D.; KRUGLYAK, L.; VARILLY, P.; PATIL, N.; KASLOW, D. C.; RIJMEN, F.; GABRIEL, C.; WARD, R.; DALY, M. J.; ALTSHULER, D. The structure of haplotype blocks in the human genome. **Science**, v. 296, n. 5576, p. 2225–2229, 2002.

GARCÍA-DORADO, A. On the consequences of ignoring purging on genetic recommendations for minimum viable population rules. **Heredity**, 115(3), 185-193, 2015.

GORDO, I.; CHARLESWORTH, B. Genetic linkage and molecular evolution. **Current Biology**, 11(17), R684-R686, 2001.

GRANATO, I. S.; GALLI, G.; DE OLIVEIRA COUTO, E. G.; SOUZA, M. B.;

HALLAUER, A. R., CARENA, M. J.; MIRANDA FILHO, J. B. Quantitative Genetics in Maize Breeding. **Springer**, 2010.

HAMRICK, J. L.; GODT, M. J. W. Effects of life history traits on genetic diversity in plant species. **Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences**, 351(1345), 1291–1298, 1996.

HARTL, D.L.; CLARK, A.G. **Principles of Population Genetics (3rd ed.)**. Sinauer Associates, 1997.

HAUDRY, A. Mating system and recombination affect molecular evolution in four Triticeae species. **Genetic Research**, v. 90, p. 97-109, 2008

HEDRICK, P. W.; HELLSTEN, U.; GRATTAPAGLIA, D. Examining the cause of high inbreeding depression: analysis of whole-genome sequence data in 28 selfed progeny of *Eucalyptus grandis*. **New Phytologist**, 206(2), 613-624, 2015.

INGVARSSON, P.K. A metapopulation perspective on genetic diversity and differentiation in partially self-fertilizing plants. **Evolution**, v. 56, p. 2368-2373, 2002.

KAGEYAMA, P. Y. Endogamia em espécies florestais. **Série Técnica IPEF**, n. 8, p. 1–12, 1981.

LABROO, M.R.; STUDER, A.J.; RUTKOSKI, J.E. Heterosis and hybrid crop breeding: a multidisciplinary review. **Frontiers in genetics**, v. 12, p. 643761, 2021.

LANDE, R.; SCHEMSKE, D. W. The evolution of self-fertilization and inbreeding depression in plants. I. Genetic models. **Evolution**, 39(1), 24–40, 1985.

LIU, P.; ZHU, J.; LU, Y. Marker-assisted selection in segregating generations of self-fertilizing crops. **Theoretical and Applied Genetics**, v. 109, p. 370–376, 2004.

LYNCH, M. *et al.* **Genetics and analysis of quantitative traits**. Sunderland, MA: Sinauer, 1998.

MALDONADO, F. A. A.; MIRANDA FILHO, J. B. Inbreeding depression in maize populations of reduced size. **Scientia Agricola**, v. 59, n. 2, p. 335–340, 2002. [

MARSHALL, T. C.; SLATE, J.; KRUIK, L. E. B.; PEMBERTON, J. M. Statistical confidence for likelihood-based paternity inference in natural populations. **Molecular Ecology**, 7(5), 639–655, 1998.

MELCHINGER, A. E. Genetic diversity and heterosis. **Genetics and exploitation of heterosis in crops**, p. 99–118, 1999.

MENDONÇA, L. F.; FRITSCHÉ-NETO, R. snpReady: a tool to assist breeders in genomic analysis. **Molecular Breeding**, 38, 1–7, 2018.

MUELLER, J. C. Linkage disequilibrium for different scales and applications. **Briefings in Bioinformatics**, 5(4), 355–364, 2004.

MYBURG, A. A.; GRATTAPAGLIA, D.; TUSKAN, G. A.; HELLSTEN, U.; HAYES, R. D.; GRIMWOOD, J.; JENKINS, J.; LINDQUIST, E.; TICE, H.; BAUER, D.; CADLWELL, J.; SPATAFORA, J. W.; PAIVA, J. A.; MASTERS, C. L.; JANSEN, R. K.; JOLY, S.; SCHATZ, M. C.; GOODSTEIN, D. M.; SHU, S.; Yang, X.; CHENG, H.; MITROS, T.; PAMPANHINHA, J. D.; ECKER, J. R.; RALPH, S. G.; DOUGLAS, C. J.; SCHATZ, M. C.; LANGER, G. D.; MACHADO, M. R.; STERCK, L.; VAN DE PEER, Y.; RITLAND, K.; SILVA, J. C.; FOSTER, G. S.; DOUGLAS, C. J.; DAVIDSON, R. M. The genome of *Eucalyptus grandis*. **Nature**, v. 510, p. 356–362, 2014.

PACHECO, C. A. P.; GAMA, E. E. G.; MIRANDA FILHO, J. B. Depressão endogâmica e heterose de híbridos de populações de milho. **Bragantia**, 61(1), 19–29, 2002.

PEIXOTO, L.A.; BHERING, L.L.; CRUZ, C.D. **Seleção Genômica Aplicada ao Melhoramento Genético**. Viçosa: UFV, 2022.

PÉREZ-LÓPEZ, F. J.; SÁNCHEZ-CASTRO, S.; GUTIÉRREZ-RAMOS, G.; MARTÍNEZ-MORALES, E.; LÓPEZ-SALINAS, E. High yielding inbred lines of maize as progenitors of single cross hybrids. **Agrociencia**, v. 48, n. 4, p. 439–450, 2014.

PÉREZ-PEREIRA, N.; POUSO, R.; RUS, A.; VILAS, A.; LÓPEZ-CORTEGANO, E.; GARCÍA-DORADO, A.; QUESADA, H.; CABALLERO, A. Long-term exhaustion of the inbreeding load in *Drosophila melanogaster*. **Heredity**, v. 127, n. 4, p. 373–383, 2021.

POLAK, G.; SAS-JAWORSKA, M.; CHMURSKA-GRĄDZKA, M.; WALEWSKI, T.;

POPE, N. B.; COOP, G. Purging due to self-fertilization does not prevent accumulation of expansion load. **PLOS Genetics**, 16(10), e1009370, 2020.

PRUSAK, B.; ŁUKASZEWICZ, M.; ZIETEK, J.; ROGOWSKA, E.; PRUSAK, A. Suitability of pedigree information and genomic methods for analyzing inbreeding of Polish cold-blooded horses covered by conservation programs. **Genes**, v. 12, n. 3, p. 429, 2021.

R CORE TEAM. **R: A Language and Environment for Statistical Computing**. R Foundation for Statistical Computing, Vienna, Austria, 2023.

RAMALHO, M. A. P.; SANTOS, H. G.; SOUZA, T. S. *Eucalyptus* breeding programs: a proposal for the use of inbred progenies. **Cerne**, v. 28, p. e103049, 2022.

RESENDE, M. D. V. Genética Quantitativa e Seleção em Plantas Perenes: Aplicações ao Melhoramento do Eucalipto. **Embrapa Florestas**, 2002.

RESENDE, R. T. *et al.* Assessing the expected response to genomic selection of individuals and families in *Eucalyptus* breeding with an additive-dominant model. **Heredity**, v. 119, n. 4, p. 245-255, 2017.

RESENDE, R. T.; RESENDE, M. D. V.; SILVA, F. F.; AZEVEDO, C. F.; TAKAHASHI, E. K.; GRATTAPAGLIA, D. Genetic improvement of *Eucalyptus*: strategies, advances, and future directions. **Scientia Forestalis**, 41(98), 173-189, 2013.

SANTOS, L. *et al.* Evaluation of the genetic diversity of a set of parents of *Eucalyptus* spp. by using microsatellite markers to direct matings. In: **BMC Proceedings**. BioMed Central. p.1-2, 2011.

SHULL, G. H. Hybridization methods in plant breeding. **Proceedings of the American Society for Horticultural Science**, Alexandria, v. 6, p. 79-94, 1909.

SILVA-JUNIOR, O. B.; FARIA, D. A.; GRATTAPAGLIA, D. Genome-wide microsatellite characterization in *Eucalyptus*: abundance, distribution and potential usefulness for studies of gene flow and hybridization. **Genetics and Molecular Biology**, São Paulo, v. 38, n. 3, p. 348-357, 2015.

SLOTTE, T.; FOXE, J. P.; HAZZOURI, K. M.; WRIGHT, S. I. Genome-wide evidence for efficient positive and purifying selection in *Capsella grandiflora*. **Molecular Biology and Evolution**, v. 27, p. 1813-1821, 2010.

SNUSTAD, D. P.; SIMMONS, M. J. **Principles of genetics**. John Wiley; Sons, 2015.

TAMBARUSSI, E. V.; BOSHIER, D.; VENCOSKY, R.; FREITAS, M. L. M. F.; SEBENN, A. M. Inbreeding depression from selfing and mating between relatives in the Neotropical tree *Cariniana legalis* Mart. Kuntze. **Conservation Genetics**, v. 18, p. 225-234, 2017.

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

WANG, X.; XU, Y.; HU, Z.; XU C. Genomic selection methods for crop improvement: Current status and prospects. **The Crop Journal**, v. 6, n. 4, p. 330-340, 2018.

WRIGHT, S. I.; GAUT, B. S. Molecular population genetics and the search for adaptive evolution in plants. **Molecular Ecology**, 14(1), 21-41, 2005.

WRIGHT, L. I.; TREGENZA, T.; HOSKEN, D. J. Inbreeding, inbreeding depression and extinction. **Conservation Genetics**, v. 9, p. 833-843, 2008.

CAPÍTULO 2

SELEÇÃO GENÔMICA APLICADA A UMA POPULAÇÃO AUTOFECONDADA DE *Eucalyptus* spp.¹

RESUMO

A seleção genômica (SG) no gênero *Eucalyptus* visa identificar genótipos superiores e reduzir os ciclos de melhoramento, mas enfrenta desafios devido à ampla base genética, o que limita sua eficiência. Como alternativa, linhagens endogâmicas surgem como uma solução para aumentar essa eficiência e gerar clones produtivos. O objetivo deste estudo foi aplicar a seleção genômica em uma população autofecundada de diferentes espécies de *Eucalyptus* spp. e prever os valores genéticos genômicos (GEBVs) de indivíduos sem fenótipo conhecido, com foco na produção de linhagens endogâmicas. A população estudada foi composta por 662 indivíduos, dos quais 600 foram fenotipados para o DAP (diâmetro à altura do peito) aos 36 meses em um experimento a campo, e 62 indivíduos estavam localizados em um pomar de hibridação, sem fenótipo conhecido. Todos os indivíduos, incluindo progênies e parentais, foram genotipados para 10.132 marcadores SNP. A análise incluiu uma associação genômica ampla (GWAS) utilizando um modelo misto e a aplicação de seleção genômica com quatro modelos frequentistas (GBLUP, GBLUP aditivo dominante, HBLUP e ABLUP) e cinco bayesianos (BRR, BayesA, BayesB, BayesC e Bayes LASSO), com validação cruzada k-fold. Os parâmetros avaliados incluíram capacidade preditiva, erro quadrático médio (EQM) e coeficiente de determinação (R^2). Os resultados da GWAS não identificaram uma região genômica ou marcador com grande efeito sobre o DAP, confirmando que essa característica é de natureza quantitativa. Entre os modelos de SG, o GBLUP teve o melhor desempenho geral, com capacidade preditiva de 0,48 e coeficiente de determinação de 0,21. O erro quadrático médio variou de 3,72 a 15,50, com o Bayes LASSO apresentando o menor erro. No entanto, o GBLUP destacou-se por integrar todos os parâmetros de forma mais eficiente. Os resultados mostram o potencial da seleção genômica no melhoramento genético de *Eucalyptus* com linhagens endogâmicas, permitindo a identificação de indivíduos promissores e a redução dos ciclos de melhoramento, um dos principais desafios do gênero.

Palavras-chave: seleção-genômica; modelos-frequentistas; modelos-bayesianos

¹ Capítulo redigido de acordo com as normas do periódico **Plants: Plant Genetics, Genomics and Biotechnology**

ABSTRAT

Genomic selection (GS) in the genus *Eucalyptus* aims to identify superior genotypes and reduce breeding cycles but faces challenges due to the broad genetic base, which limits its efficiency. As an alternative, inbred lines emerge as a solution to increase this efficiency and generate productive clones. The objective of this study was to apply genomic selection in a self-pollinated population of different *Eucalyptus* spp. and predict the genomic estimated breeding values (GEBVs) of individuals without known phenotypes, with a focus on the production of inbred lines. The studied population consisted of 662 individuals, of which 600 were phenotyped for DBH (diameter at breast height) at 36 months in a field experiment, and 62 individuals were located in a hybridization orchard without known phenotypes. All individuals, including progenies and parents, were genotyped for 10,132 SNP markers. The analysis included a genome-wide association study (GWAS) using a mixed model and the application of genomic selection with four frequentist models (GBLUP, additive-dominant GBLUP, HBLUP, and ABLUP) and five Bayesian models (BRR, BayesA, BayesB, BayesC, and Bayes LASSO), with k-fold cross-validation. The evaluated parameters included predictive ability, mean squared error (MSE), and coefficient of determination (R^2). The GWAS results did not identify a genomic region or marker with a large effect on DBH, confirming that this trait is quantitative in nature. Among the GS models, GBLUP had the best overall performance, with a predictive ability of 0.48 and a coefficient of determination of 0.21. The mean squared error ranged from 3.72 to 15.50, with Bayes LASSO showing the lowest error. However, GBLUP stood out for integrating all parameters more efficiently. The results demonstrate the potential of genomic selection in the genetic improvement of *Eucalyptus* with inbred lines, allowing for the identification of promising individuals and the reduction of breeding cycles, one of the main challenges of the genus.

Keywords: genomic-selection; frequentist-models; bayesian-models.

2.1 INTRODUÇÃO

O uso de marcadores moleculares para a seleção de indivíduos foi inicialmente proposto por Lebowitz e Soller (1987), por meio da seleção assistida por marcadores (SAM). No entanto, o método apresentava limitações, pois utilizava marcadores associados a características de grandes efeitos (QTL), o que dificultava sua aplicação, principalmente no gênero *Eucalyptus* (Grattapaglia; Resende, 2011), pois a maioria dos caracteres relacionados com a produtividade é de natureza quantitativa, sendo controlados por muito genes de pequenos efeitos, o que reduz a eficiência dos modelos de SAM (Grattapaglia; Kirst, 2008).

Assim, visando superar essas limitações, Meuwissen (2001) desenvolveu a metodologia da seleção genômica, que permite mensurar de forma eficiente os efeitos de características quantitativas. Essa abordagem tornou-se uma ferramenta poderosa e versátil para as tomadas de decisão de melhoristas em programas de melhoramento, permitindo estimar valores genéticos com maior precisão (Resende *et al.*, 2012; Faria, 2014; Denis *et al.*, 2013). Como resultado, a seleção mostrou-se mais eficiente, reduzindo o número de ciclos necessários para o melhoramento (Heffner *et al.*, 2009).

Os modelos de seleção genômica integram informações fenotípicas e de marcadores moleculares de indivíduos e/ou populações, e utilizam dessas informações para treinar os modelos com o objetivo de prever a performance dos indivíduos antes mesmo de serem implantados em testes de campo (Daetwyler *et al.*, 2008). Contudo, a eficiência dessa ferramenta depende de diversos fatores, como o tamanho amostral elevado, uma vez que um grande número de indivíduos é necessário para que os modelos compreendam os padrões de variabilidade da população (Valente, 2014). Além disso, a qualidade e a densidade da genotipagem, o grau de parentesco entre os indivíduos da população ou populações e um delineamento experimental bem planejado e conduzido são fatores cruciais (Calus *et al.*, 2014). Tais elementos impactam diretamente a capacidade dos modelos de prever fenótipos com maior precisão (López-Cruz *et al.*, 2015; Hayes *et al.*, 2009).

Nos programas de melhoramento de *Eucalyptus*, a seleção genômica é principalmente usada para a predição da performance de indivíduos destinados à clonagem, com o objetivo de reduzir o tempo necessário para os ciclos de

melhoramento (Grattappaglia; Resende, 2011). Enquanto o desenvolvimento de um clone comercial levaria em torno de 13 a 14 anos, o uso da seleção genômica permite reduzir esse período para aproximadamente sete anos (Resende *et al.*, 2012; Grattappaglia *et al.*, 2018; Grattappaglia *et al.*, 2011; Paludeto *et al.*, 2021). No entanto, devido à ampla base genética das populações de melhoramento de *Eucalyptus*, que resulta em alta variabilidade, os modelos de seleção genômica apresentam menor eficiência para prever o desempenho de indivíduos em comparação com outras culturas como milho, soja e arroz (Frisch; Melchinger, 2007; Hufford *et al.*, 2012).

Diante da alta demanda do mercado florestal por híbridos produtivos e adaptados, o método de linhagens endogâmicas surge como uma alternativa promissora para atender a essas exigências (Ramalho *et al.*, 2022). Esse método permite explorar os efeitos aditivos por meio da autofecundação de indivíduos e, posteriormente, pela realização de cruzamentos específicos, resultando em híbridos simples altamente produtivos e livres de carga genética indesejada (Sprague *et al.*, 1942).

No contexto da aplicação da seleção genômica, o uso de linhagens endogâmicas potencializa a eficiência dessa técnica no gênero *Eucalyptus* (Gaynor *et al.*, 2017). Isso ocorre porque, com as autofecundações, o genoma dos indivíduos torna-se mais homozigoto, restringindo a variabilidade genética e diminuindo o tamanho efetivo populacional (Allard, 1999). Essas condições facilitam a compreensão dos padrões de variabilidade genética pelos modelos de seleção genômica, resultando em predições de genótipos mais precisas e confiáveis (Bernardo, 2006).

Portanto, objetivou-se com o estudo através da GWAS observar regiões genômicas com maiores efeitos dentro da população de estudo para o caráter avaliado e aplicar a ferramenta da seleção genômica em uma população autofecundada composta por diferentes espécies de *Eucalyptus spp.*, assim como prever os valores genéticos genômicos (*Genomic estimated breeding value*- GEBVs) de indivíduos sem fenótipo conhecido, visando a produção de linhagens endogâmicas.

2.2 MATERIAL E MÉTODOS

2.2.1 População estudada

A pesquisa foi realizada em parceria com a Suzano S.A., onde foram selecionadas um total de 20 genitores superiores para o estudo, de diferentes espécies do gênero *Eucalyptus* spp. Esses genitores foram submetidos à autofecundação, para obtenção das sementes. Com essas sementes foram produzidas mudas. Aproximadamente 30 mudas de cada uma das progênies foram utilizadas, totalizando 600 indivíduos para serem conduzidas para um experimento a campo e o restante das mudas foram conduzidas para um pomar de hibridação, totalizando 62 indivíduos. Todos os 662 indivíduos provenientes de autofecundações de espécies de *Eucalyptus* spp. e seus respectivos parentais foram utilizados nas análises. Dos 662 indivíduos, 600 estão presente em um experimento de blocos casualizados, com uma planta por parcela, totalizando em 30 blocos. Para esses indivíduos o diâmetro à 1,30 m (DAP) foi mensurado aos 36 meses de idade, e genotipados. Os outros 62 indivíduos, que são irmãos completos de autofecundação dos indivíduos avaliados a campo, foram inseridos em um pomar de hibridação e não possuem fenótipo conhecido, sendo apenas genotipados. Todos os indivíduos e seus genitores totalizando 682 indivíduos, foram genotipados para 10.132 marcadores do tipo SNP, usando o chip customizado pertencente a Suzano S.A. para obtenção dos dados genômicos.

2.2.2 Associação Genômica Ampla (Genome-wide analysis -GWAS)

As associações entre marcadores moleculares e o DAP foram obtidas utilizando, os efeitos de todos os 10132 marcadores disponíveis. Assim esses efeitos foram estimados baseado em um modelo linear misto implementado no pacote “rrBLUP” do R:

$$\mathbf{y} = \mathbf{S}\mathbf{a} + \mathbf{Q}\mathbf{v} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon} \quad (1)$$

Sendo $\mathbf{y}$ é o vetor de observações fenotípicas ajustadas pelo efeito de bloco, $\mathbf{a}$ é o vetor dos efeitos aleatórios do marcador, $\mathbf{v}$ é o vetor do efeito de estrutura populacional (efeito fixo), $\mathbf{u}$ é o vetor de os efeitos poligênicos (efeito aleatório), e $\boldsymbol{\varepsilon}$ é o vetor dos efeitos residuais. Os termos $\mathbf{S}$, $\mathbf{Q}$ e $\mathbf{Z}$ são as matrizes de incidência relacionadas aos vetores $\mathbf{a}$, $\mathbf{v}$ e $\mathbf{u}$, respectivamente. A variância $\mathbf{u}$ é expressa por $\sigma_u^2 = 2 K \sigma_g^2$ e a $\boldsymbol{\varepsilon}$ é expressa $\sigma_e^2 = 2 R \sigma_R^2$, onde $\mathbf{K}$ é uma matriz de coeficientes de

parentesco segundo Endelman e Jannink (2012). As variâncias σ_g^2 e σ_e^2 correspondem às variâncias genéticas e residuais, respectivamente. O pacote “corrplot” foi utilizado para gerar os gráficos.

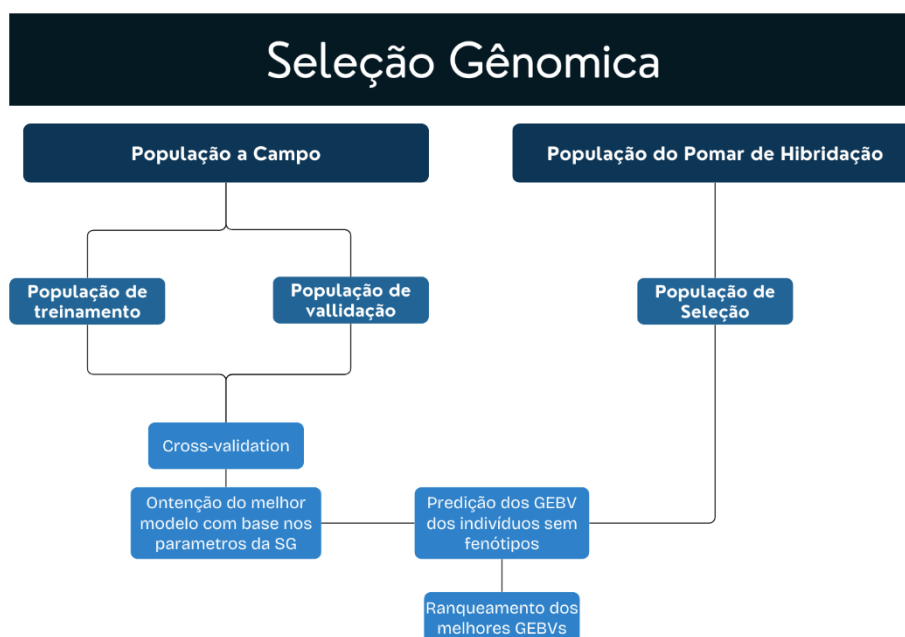
2.2.3 Controle de qualidade dos SNPs

Para construção das matrizes de seleção genômica, alguns marcadores foram excluídos com base nos critérios de qualidade: MAF (*Minor allele Frequency*) de 0.05 e uma *Call Rate* de 0.95 (Granato *et al.*, 2018). Dessa forma, marcadores com frequências menores que 5% dentro da população e indivíduos com mais de 5 % de marcadores faltantes foram removidos, como objetivo de retirar marcadores não explicativos dentro da população e para facilitar os modelos de seleção genômica de compreender padrões de variabilidade.

2.2.4 Seleção e predição genômica

As análises de seleção e predição genômica a população foram separadas em dois grupos. i) População de treinamento e validação: composta pelos indivíduos presentes no experimento a campo, com fenótipo e genótipo conhecido. Essa população foi utilizada para treinar e validar os modelos de seleção genômica testados, possibilitando a identificação do modelo mais adequado para a predição (Figura 1). ii) População de seleção: constituída por 62 indivíduos genotipados, localizados em um pomar de hibridação da Suzano S.A. Esses indivíduos são irmãos completos de autofecundação dos indivíduos avaliados a campo, mas não possuem fenótipos conhecidos.

Figura 1 - Esquema de Seleção genômica para as populações a campo e do pomar de hibridação de *Eucalyptus* spp..



Assim, seleção genômica ocorreu em duas etapas. Na primeira etapa, os modelos foram treinados e validados quanto à sua capacidade preditiva, erro quadrático médio e coeficiente de determinação (R^2). Na segunda etapa foram preditos os valores genéticos genômicos dos indivíduos presentes no pomar de hibridação.

Os modelos de predição genômica foram ajustados em um passo único (*single-step*), ou seja, dados fenotípicos, genotípicos e informações do delineamento experimental foram fornecidos simultaneamente aos modelos para ajuste.

Quatro modelos frequentistas e cinco modelos bayesianos de predição genômica foram testados. Dentre os modelos frequentistas, o primeiro modelo foi o ABLUP (Hederson, 1984), onde a matriz de parentesco fornecida para as predições é baseada no pedigree (A). O segundo foi o GBLUP (Vanraden, 2008) utilizando uma matriz de parentesco genômico (G) obtida com base nos efeitos aditivos dos marcadores. O terceiro modelo testado foi o GBLUP aditivo-dominante (Vitezica *et al.*, 2013), utilizando uma matriz de parentesco que leva em consideração os efeitos aditivos e dominantes dos marcados e, por último, o quarto modelo foi o HBLUP

(Legarra *et al.*, 2009) que faz uma combinação entre a matriz de pedigree (A) e a matriz genômica (G) para obtenção de uma matriz híbrida (H).

O modelo ABLUP (Herderson, 1984), pode ser matricialmente definido por:

$$Y_{ij} = X\beta + Z_a a + \varepsilon \quad (2)$$

Sendo, Y o vetor dos valores fenotípicos, β (px1) o vetor dos efeitos fixos (blocos e média geral), a (qx1) o vetor dos efeitos aleatórios genético aditivos, sendo que $a \sim N(0, A\sigma_a^2)$ (A é a matriz de pedigree e σ_a^2 é variância genética aditiva (qxq)), ε o vetor dos efeitos residuais, em que $\varepsilon \sim N(0, I\sigma_e^2)$ (I é a matriz identidade e σ_e^2 é variância residual), X (nxp) e Z (nxq) são as matrizes de incidência para os efeitos fixos e aleatórios, respectivamente.

O modelo GBLUP (Vanraden, 2008), pode ser matricialmente definido por:

$$Y_{ij} = X\beta + Z_g g + \varepsilon \quad (3)$$

Sendo, g (qx1) o vetor aleatório dos efeitos genéticos genômicos aditivos, com $g \sim N(0, G\sigma_g^2)$ (G é a matriz genômica e σ_g^2 é variância genética genômica aditiva (qxq)), Z_g (nxq) é a matriz de incidência para os efeitos genéticos genômicos aditivos.

O modelo GBLUP AD (Vitezica *et al.*, 2013), pode ser matricialmente definido por:

$$Y_{ij} = X\beta + Z_g g + Z_d d + \varepsilon \quad (4)$$

Sendo, d (qx1) o vetor aleatório dos efeitos genéticos genômicos de dominância, sendo que $d \sim N(0, G_d\sigma_d^2)$ (G_d é a matriz genômica de dominância e σ_d^2 é variância genética de dominância (qxq)). Z_d (nxq) é a matriz de incidência para os efeitos de dominância.

O modelo HBLUP (Legarra *et al.*, 2009), pode ser matricialmente definido por:

$$Y_{ij} = X\beta + Z_h h + \varepsilon \quad (5)$$

Sendo h (qx1) o vetor de efeitos aleatórios genéticos aditivos, com $h \sim N(0, H\sigma_h^2)$ (H é a matriz genômica híbrida e σ_h^2 é variância genética aditiva (qxq)). Z_h (nxq) é a matriz de incidência para o efeito aleatório.

Dentre os modelos bayesianos que foram testados e validados, tem-se o Bayes Ridge Regression (BRR), BayesA, BayesB, BayesC e o Bayes LASSO, onde todos os modelos seguiram a mesma estrutura matricial:

$$Y_{ij} = X\beta + \varepsilon \quad (6)$$

Sendo, Y_{ij} o vetor de valores fenotípicos observados, X é a matriz de genótipos ($n \times p$), onde n é o número de indivíduos e p o número de marcadores, β é o vetor de efeitos aleatórios dos marcadores e ε é o erro residual.

A diferença entre os modelos bayesianos testados está nas suposições a priori feitas para cada modelo, sendo estabelecidos parâmetros dos efeitos dos marcadores, para que fosse possível testar todos os efeitos presentes na população e qual das priori melhor se adequava a variabilidade dos dados. No caso do modelo BRR (Gianola *et al.*, 2009), assume-se que todos os marcadores possuem variâncias iguais, ou seja, os efeitos dos marcadores são considerados homogêneos. Esse modelo pode ser representado pela seguinte fórmula:

$$\beta_j \sim N(0, \sigma_\beta^2)$$

$$\sigma_\beta^2 \sim Inv - \chi^2(v_\beta, S_\beta)$$

$$\varepsilon \sim N(0, \sigma_\varepsilon^2), \sigma_\varepsilon^2 \sim Inv - \chi^2(v_\varepsilon, S_\varepsilon)$$

Sendo, β_j os efeitos dos marcadores seguem uma distribuição normal com média zero e variância σ_β^2 , σ_β^2 variância dos efeitos dos marcadores segue uma distribuição inversa qui-quadrada ($Inv - \chi^2$), v_β são os graus de liberdade e S_β a escala, ε é o erro experimental que segue uma distribuição normal e variância igual a zero ($N(0, \sigma_\varepsilon^2)$).

O BayesA (Meuwissen *et al.*, 2001) assume que cada marcador possui sua própria variância, resultando em efeitos heterogêneos entre os marcadores, e pode ser representado da seguinte maneira:

$$\beta_j \sim N(0, \sigma_{\beta_j}^2)$$

$\sigma_{\beta_j}^2 \sim Inv - \chi^2(v_\beta, S_\beta)$ Sendo, β_j o efeito aleatório de cada marcador (j), $\sigma_{\beta_j}^2$ é a variância específica de cada marcador segue uma distribuição inversa qui-quadrada.

O modelo BayesB (Meuwissen *et al.*, 2001) tem como priori uma correção π , que determina a probabilidade de um marcador apresentar efeito nulo. Os marcadores com efeitos não nulos possuem heterogeneidade e pode ser representado por:

$$\beta_j \begin{cases} N(0, \sigma_{\beta_j}^2), \text{ com probabilidade } (1 - \pi) \\ 0, \text{ com probabilidade } \pi \end{cases}$$

$$\sigma_{\beta_j}^2 \sim \text{Inv} - \chi^2(v_\beta, S_\beta)$$

Sendo, β_j os efeitos dos marcadores com correção π para a probabilidade de um marcador ter efeito igual a zero, $\sigma_{\beta_j}^2$ é a variância dos marcadores não nulos e segue uma distribuição inversa qui-quadrada.

O BayesC (Habier *et al.*, 2011) também utiliza a correção π como pressuposto a priori, representando a probabilidade de um marcador apresentar efeito nulo. Contudo, sua principal diferença em relação ao BayesB está nos marcadores não nulos, já que o BayesC assume que todos os marcadores com efeito possuem a mesma variância, ou seja, efeitos homogêneos. Esse modelo pode ser representado pela seguinte equação:

$$\beta_j \begin{cases} N(0, \sigma_{\beta_j}^2), \text{ com probabilidade } (1 - \pi) \\ 0, \text{ com probabilidade } \pi \end{cases}$$

$$\sigma_{\beta_j}^2 \sim \text{Inv} - \chi^2(v_\beta, S_\beta)$$

Sendo, β_j os efeitos dos marcadores com correção π para a probabilidade de um marcador ter efeito igual a zero e marcadores não nulos possuem o mesmo efeito, $\sigma_{\beta_j}^2$ é a variância dos marcadores não nulos e segue uma distribuição inversa qui-quadrada.

O modelo Bayes LASSO (Park; Cassella, 2008) assume, a priori, uma distribuição Laplace (também conhecida como distribuição exponencial dupla). Esse pressuposto introduz uma penalização nos efeitos dos marcadores, favorecendo soluções esparsadas, ou seja, estimativas em que muitos efeitos de marcadores são reduzidos a valores próximos de zero. O modelo pode ser representado por: $\beta_j \sim \text{Laplace}(0, \lambda)$

$$\lambda \sim \text{Gamma}(a, b)$$

Sendo, β_j os efeitos dos marcadores seguindo uma distribuição Laplace, λ é o parâmetro de escala da distribuição Laplace, que controla a intensidade da penalização. Segue uma distribuição Gamma para incorporar incerteza.

Para determinar a eficiência dos modelos testados, utilizou-se o método de validação cruzada *k-folding* (Berrar, 2018). Neste método, a base de dados foi dividida em 10 partes (*Folds*), e, em cada *Fold*, uma parte dos valores foi retirada. Esses valores foram preditos pelos modelos e, posteriormente, comparados com os valores reais.

Os parâmetros utilizados para avaliar a eficiência dos modelos de predição genômica testados foram: a capacidade preditiva (*CP*), que é a correlação entre os valores preditos e reais. Quanto mais próximo de 1, maior é a capacidade do modelo de prever os genótipos. A *CP* (Pearson, 1896) pode ser representada por:

$$CP = \frac{Cov(\hat{y}, y)}{\sigma_{\hat{y}} \sigma_y} \quad (7)$$

Sendo $Cov(\hat{y}, y)$ é a covariância entre os valores preditos ($\hat{y}$) e os valores reais (y); $\sigma_{\hat{y}}$ é o desvio padrão da variância dos valores preditos e σ_y é o desvio padrão da variância dos valores reais.

Outro parâmetro avaliado foi o erro quadrático médio (EQM), que indica a média da diferença entre o valor predito e o real, ou seja, o quanto as predições do modelo se distanciam dos valores reais. Portanto, quanto mais próximo de zero, melhor. O *EQM* (Kuhn e Johnson, 2013) pode ser representado por:

$$EQM = \frac{1}{n} \sum_{i=1}^n (\hat{y} - y)^2 \quad (8)$$

Onde $\hat{y}$ são os valores preditos; y são os valores observados e n é o número de observações.

O último parâmetro utilizado para avaliar eficiência do modelo de predição foi o coeficiente de determinação (R^2) do modelo, que mede a proporção da variância total explicada pelo modelo. O R^2 (Montgomery *et al.*, 2010) pode ser representado por:

$$R^2 = 1 - \frac{SS_{res}}{SS_{total}} \quad (9)$$

Onde SS_{res} é a soma de quadrados dos resíduos e SS_{total} é soma de quadrados total.

Em uma análise preliminar dos dados, foram identificados *outliers* que poderiam influenciar negativamente o desempenho dos modelos de predição. Para evitar sua exclusão, foi aplicada a técnica de winsorização (YANG *et al.*, 2024), com o objetivo de transformar os *outliers* e, assim, assegurar a normalidade dos dados observados.

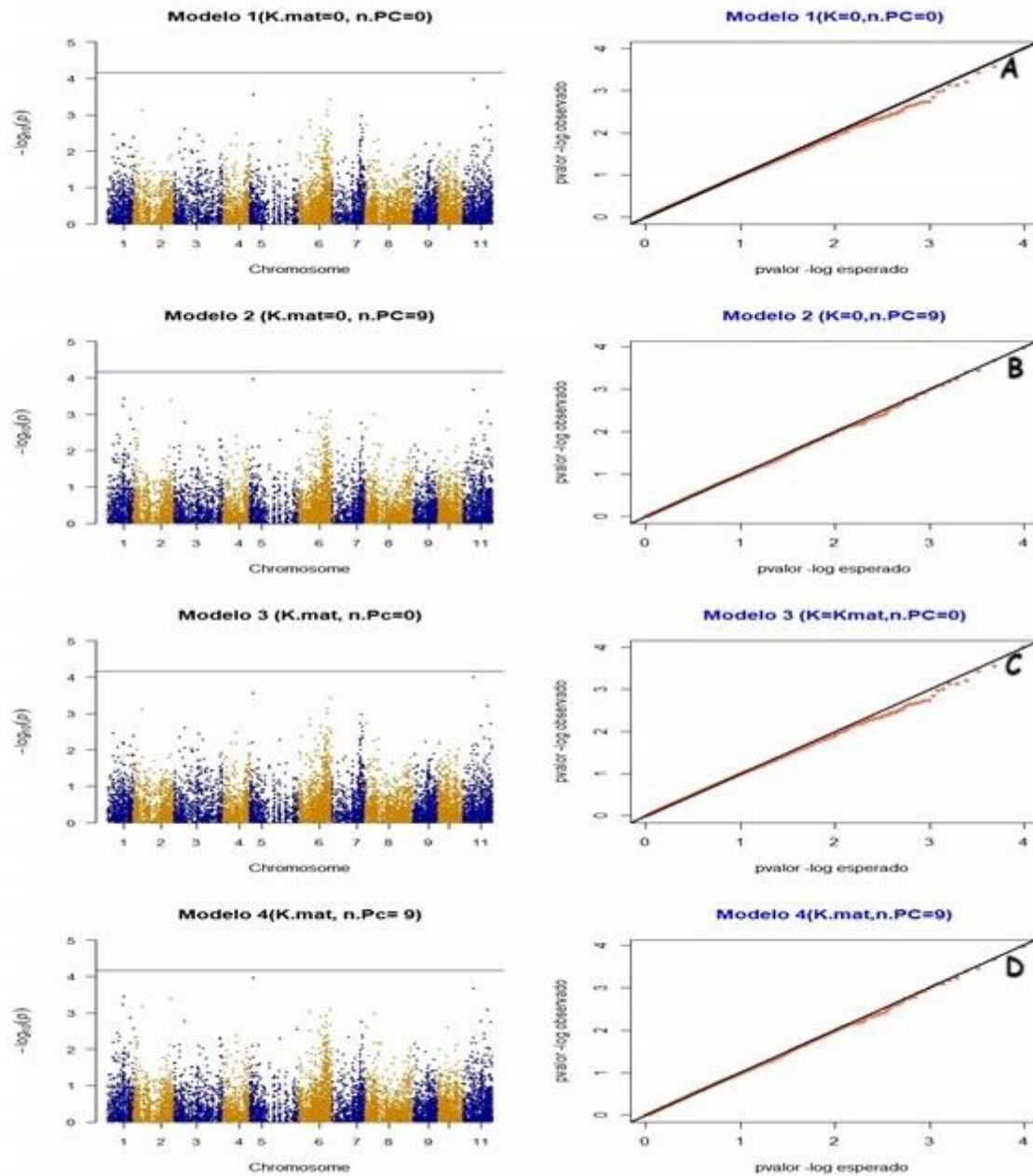
2.3 RESULTADOS

2.3.1 Associação genômica Ampla (GWAS)

Para o caráter DAP aos 36, com uma média amostral de 12,25 ($\pm 4,18$) cm (Figura 2), o modelo que considera o efeito da matriz de parentesco e dos componentes principais mostrou-se mais ajustado para indicar os efeitos dos marcadores associado ao caráter avaliado.

No modelo quatro (PCA + Matriz de parentesco), os valores de $-\log$ observados foram semelhantes aos valores esperados, indicando que o modelo está bem ajustado à base de dados. No entanto, nenhum marcador ultrapassou a linha de corte, sugerindo que não houve associações significativas detectadas com este caráter.

Figura 2 - Associação genômica ampla (GWAS) para o caráter DAP aos 36 meses, utilizando quatro diferentes modelos em progênies de uma população autofecundada de *Eucalyptus spp.*

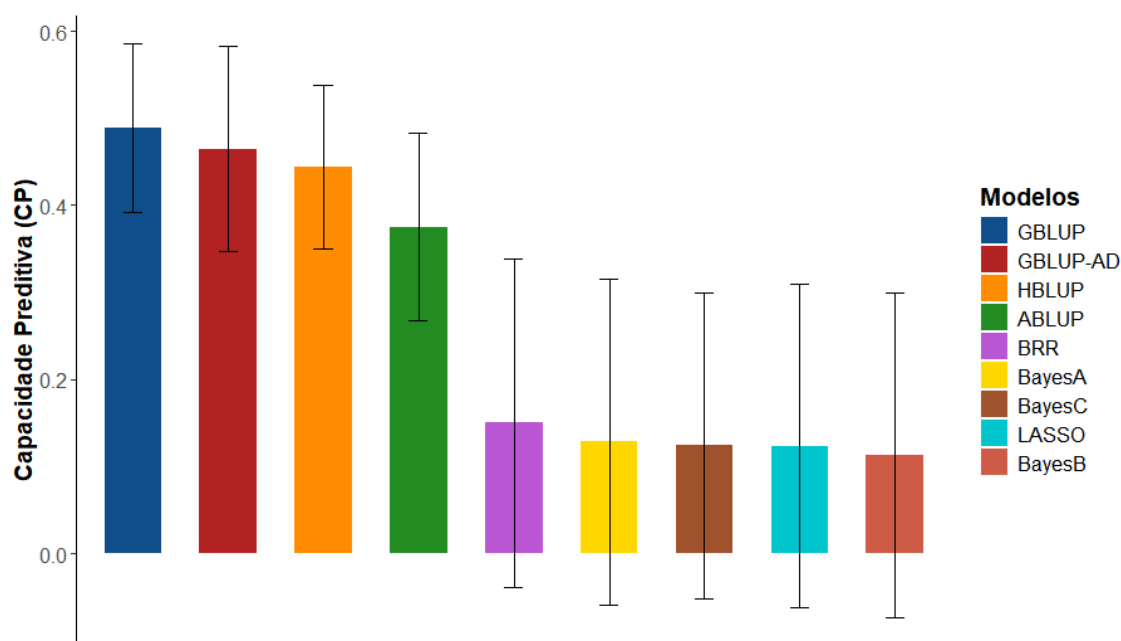


A: Modelo 1; B: Modelo 2; C: Modelo 3; D: Modelo 4

2.3.2 Parâmetros da predição genômica

Entre os modelos frequentistas, o GBLUP apresentou o melhor desempenho em termos de capacidade preditiva (0,488), pois quanto mais próximo de um, melhor é a capacidade do modelo em prever genótipos. Seguido pelo GBLUP aditivo-dominante (0,465), HBLUP (0,444) e ABLUP (0,375). Para os modelos bayesianos, as capacidades preditivas variaram de 0,114 a 0,150 (Figura 3). O modelo BRR foi o que apresentou a maior capacidade preditiva entre os modelos bayesianos, mas ainda muito inferior ao desempenho do GBLUP.

Figura 3 - Capacidade preditiva média e desvio padrão para os modelos de predição genômica, avaliados por validação cruzada, em dados de indivíduos aos 36 meses de idade em uma população autofecundada de *Eucalyptus* spp.



O modelo com menor erro foi quadrático médio (EQM) entre os modelos frequentistas foi o GBLUP (13,336), seguido do GBLUP aditivo-dominante (13,992), HBLUP (14,390) e ABLUP (15,498) (Tabela 1). Esses resultados indicam que o GBLUP apresenta menor erro associado aos valores preditos em comparação os demais modelos avaliados. Por outro lado, entre os modelos bayesianos testados, todos apresentaram valores de EQM muito semelhantes, variando de 3.719 a 3.735, com o modelo Bayes LASSO apresentando desempenho levemente superior aos demais (Tabela 1).

Apesar dos menores valores de EQM apresentados pelos modelos bayesianos, isso não refletiu em coeficientes de determinação (R^2) mais elevados. O maior valor de R^2 entre os modelos bayesianos foi obtido pelo BRR (0,055). Entre os modelos frequentistas, os maiores coeficientes de determinação (R^2) foram observados no GBLUP (0,225), seguido pelo HBLUP (0,185) e pelo GBLUP aditivo-dominante (0,185), sendo que os dois últimos apresentaram valores muito semelhantes. O GBLUP destacou-se também por apresentar o menor desvio padrão entre os modelos. O ABLUP, por sua vez, apresentou o menor valor de R^2 , com 0,124 (Tabela 1).

Tabela 1 - Erro Quadrático Médio (EQM) e coeficiente de determinação (R^2) de quatro modelos frequentistas e cinco bayesianos, para o caráter DAP aos 36 meses de idade em uma população autofecundada de *Eucalyptus* spp..

Modelos	Parâmetros	
	EQM	R^2
GBLUP	13,336 ($\pm 1,930$)	0,225 ($\pm 0,105$)
GBLUP-AD	13,992 ($\pm 2,565$)	0,187 ($\pm 0,145$)
HBLUP	14,390 ($\pm 1,408$)	0,185 ($\pm 0,077$)
ABLUP	15,498 ($\pm 1,986$)	0,124 ($\pm 0,095$)
BayesA	3,721 ($\pm 0,222$)	0,048 ($\pm 0,053$)
BayesB	3,735 ($\pm 0,219$)	0,044 ($\pm 0,053$)
BayesC	3,732 ($\pm 0,231$)	0,043 ($\pm 0,043$)
LASSO	3,719 ($\pm 0,220$)	0,046 ($\pm 0,055$)
BRR	3,723 ($\pm 0,257$)	0,055 ($\pm 0,054$)

Com base nos parâmetros avaliados para todos os modelos testados, o GBLUP mostrou-se superior aos demais, apresentando a maior capacidade preditiva e menor erro quadrático médio (EQM), destacando-se como o modelo mais adequado para as análises de predição genômica. Assim, utilizando o GBLUP, foram preditas as performances dos indivíduos do pomar de hibridação, com o objetivo de ranquear esses indivíduos para seleção (Tabela 2).

Tabela 2 - Ranqueamento dos 10 melhores valores preditos dos indivíduos do pomar de hibridação com base na predição genômica via GBLUP realizada em seus meios-irmãos aos 36 meses de idade em *Eucalyptus* spp.

Indivíduo	DAP predito	Desvio padrão	GBLUP
INDP008	16,069	2,319	3,429
INDP060	15,130	2,837	2,490
INDP061	14,047	3,290	1,407
INDP002	13,651	2,549	1,011
INDP021	13,371	3,477	0,731
INDP003	13,270	2,662	0,630
INDP048	13,075	3,838	0,435
INDP006	13,040	2,360	0,400
INDP014	13,028	2,382	0,388
INDP015	12,796	3,094	0,156

2.4 DISCUSSÃO

2.4.1 Associação genômica ampla (GWAS)

Os resultados da associação genômica ampla (GWAS) não identificaram SNPs significativos, mesmo com os efeitos da autofecundação tornando o genoma mais homozigoto (Atwell *et al.*, 2010; Zhong *et al.*, 2009). Isso pode ser explicado pelo fato de o diâmetro à altura do peito (DAP) ser uma característica quantitativa, controlada por muitos genes de pequeno efeito (Ornella *et al.*, 2012; Desta; Ortiz, 2014; Rocha *et al.*, 2022). Assim, não há regiões específicas do genoma com efeito majoritário sobre essa característica, dificultando a identificação de associações significativas.

Mesmo com o processo de autofecundação, que promove a fixação de alelos no estado homozigoto no genoma, essa fixação ainda não foi suficiente para formar grandes regiões genômicas fixadas que controlam a característica avaliada (Meuwissen *et al.*, 2001; Zargar *et al.*, 2015). Essa limitação pode ser explicada por dois fatores principais: primeiro, o caráter avaliado é quantitativo, e, até o momento, não há QTLs identificados para espécies do gênero *Eucalyptus* que expliquem uma grande proporção da variação fenotípica (Ornella *et al.*, 2012). Segundo, as taxas de

homozigossidade entre os indivíduos ainda são relativamente baixas, indicando que apenas poucos loci no genoma possuem alelos fixados (Zhong *et al.*, 2009).

Dessa forma, a seleção de indivíduos para as linhagens endogâmicas não pode ser realizada com base em um número reduzido de marcadores, uma vez que isso não capturaria de maneira eficaz a variação genética subjacente (Hayward, 2012). Para compreender a variabilidade dessa característica e aumentar a acurácia da predição genética, é necessário utilizar abordagens que considerem a contribuição conjunta de um grande número de marcadores ao longo de todo o genoma, como a seleção genômica por meio de modelos preditivos (Meuwissen, 2001).

2.4.2 Seleção e Predição Genômica

O GBLUP foi superior em relação aos demais modelos testados, isso pode ser atribuído ao uso da matriz de parentesco genômico, que possibilita uma estimativa mais precisa das relações genéticas entre os indivíduos a partir dos marcadores moleculares (Scutari, 2013). Essa estrutura permite ao GBLUP capturar de maneira mais eficiente a variabilidade genética aditiva, sendo particularmente vantajosa para características poligênicas, nas quais múltiplos loci de pequeno efeito contribuem para a expressão fenotípica (Resende *et al.*, 2012). Por outro lado, a menor performance dos modelos bayesianos testados pode estar associada à como esses modelos lidam com a distribuição dos efeitos dos marcadores, pois estes modelos assumem forma probabilística, que muitos loci possuem efeitos nulos e priorizam aqueles com maiores efeitos (Shi *et al.*, 2021), sendo menos eficientes para características amplamente poligênicas, onde os efeitos estão distribuídos ao longo de todo o genoma, onde é possível observar capacidades preditivas de variaram 0,114 a 0,150, em que o GBLUP apresentou uma capacidade preditiva de 0,48.

Outro fator que influencia esses resultados é a relação entre os modelos testados e o tamanho da base de dados utilizada. O GBLUP demonstrou maior estabilidade, sendo menos suscetível ao sobreajuste, o que o torna mais adequado para bases de dados menores. Em contrapartida, os modelos bayesianos apresentaram maior predisposição ao sobreajuste, possivelmente devido à tentativa de capturar padrões irrelevantes (Montesinos-López *et al.*, 2023; Takahashi *et al.*, 2020). Assim, o modelo GBLUP pode ser considerado o mais eficiente quando comparado aos demais.

Estudos anteriores exploraram as capacidades preditivas para diferentes espécies de *Eucalyptus*. Windhausen *et al.* (2012) e de Peixoto *et al.* (2024) relataram capacidades preditivas variando de 0,53 a 0,69 e de 0,46 a 0,61 para o milho, respectivamente. Da mesma forma, Resende *et al.* (2012) encontraram capacidades preditivas entre 0,55 e 0,56 ao avaliar populações de melhoramento de *Eucalyptus*. Por outro lado, Duarte *et al.* (2024) relataram valores preditivos mais baixos, variando de 0,24 a 0,39, em comparação com o presente estudo. Adicionalmente, Mphahlele *et al.* (2020) e Grattapaglia *et al.* (2011) demonstraram capacidades preditivas variando de 0,47 a 0,67 e de 0,54 a 0,69, respectivamente. Uma característica comum nesses estudos é o uso de amostras grandes, com populações superiores a 1000 indivíduos, para melhorar os conjuntos de treinamento e validação, além de um número considerável de marcadores significativos. Assim, essas pesquisas anteriores mostram que o uso de um maior número de indivíduos é necessário para alcançar valores mais altos de precisão preditiva, o que não é o caso no presente estudo.

Os valores inferiores obtidos neste estudo podem ser atribuídos ao baixo número de indivíduos amostrados, em comparação com as populações normalmente utilizadas na literatura. Para alcançar altos valores de capacidade preditiva, é recomendada uma população com pelo menos 1000 indivíduos, associada a uma alta intensidade de marcadores (Grattapaglia *et al.*, 2011). No entanto, apesar da quantidade reduzida de indivíduos na presente análise, os valores de capacidade preditiva foram considerados satisfatórios. Isso pode ser explicado pelo uso da técnica de winsorização, pois é uma técnica que permite limitar os impactos dos *outliers*, contribuindo para a melhoria da capacidade preditiva nos modelos testados. Portanto, valores atípicos foram evitados diminuindo as distorções das estimativas dos efeitos genéticos dos marcadores (Schafer; Graham, 2002), o que ocasionou em capacidades preditivas mais elevadas principalmente para os modelos frequentistas.

Outro fator para a eficiência dos modelos de seleção genômica foi a estrutura populacional e os efeitos da endogamia decorrentes da autofecundação (Heslot *et al.*, 2012). A redução do tamanho efetivo da população resultante desse processo favoreceu a aplicação da seleção genômica, uma vez que a menor diversidade genética pode aumentar a precisão das estimativas preditivas (Robert *et al.*, 2022). Isso ocorre, pois, populações derivadas de autofecundação compartilham uma maior proporção de alelos entre si, reduzindo a variabilidade genética e facilitando a

identificação de padrões consistentes nos efeitos genéticos (Ceballos *et al.*, 2018), tornando mais fácil que os modelos de seleção genômica captem os efeitos aditivos (Ornella *et al.*, 2012; Meuwissen *et al.*, 2001). Dessa forma, mesmo em populações com tamanho amostral reduzido, a maior homogeneidade genética favorece a capacidade preditiva dos modelos, maximizando a acurácia das estimativas genômicas (Wientjes *et al.*, 2016).

Dessa forma, os modelos de seleção genômica testados foram capazes de prever a performance dos indivíduos sem fenótipos presentes no pomar de hibridação. Esse resultado permite acelerar os ciclos de melhoramento, pois, como há a performance predita dos indivíduos, não há necessidade de levá-los a campo e testá-los para, só então, resgatá-los e cloná-los. Assim, é possível prosseguir diretamente para os pomares de hibridação, avançando no processo de formação de linhagens (Desta; Ortiz, 2014).

Visando a formação de linhagens endogâmicas, os melhores indivíduos preditos podem ser selecionados para futuros cruzamentos ou autofecundações, permitindo o avanço de mais um ciclo de melhoramento, com o objetivo de produzir linhagens endogâmicas. Além disso, esses indivíduos possibilitam a realização de cruzamentos entre si, buscando explorar a complementariedade genética entre eles, um fator crucial para o sucesso na produção de linhagens endogâmicas e no incremento do progresso genético.

2.5 CONCLUSÕES

A análise da GWAS não identificou marcadores significativamente associados ao caráter avaliado, assim não se observou a região específica do genoma que controla esse caráter. Como o DAP se trata de uma característica quantitativa este é controlado por muitos genes de pequeno efeito, indicando não se tornar possível selecionar indivíduos para as linhagens endogâmicas apenas com um pequeno número de marcadores, sendo necessário informações de todo o genoma para compreender a variabilidade do caráter.

O GBLUP destacou-se como o mais eficiente entre os modelos frequentistas. No caso dos modelos bayesianos, o BRR mostrou-se o mais eficaz, embora seu desempenho tenha sido consideravelmente inferior ao do GBLUP. Com base nos

valores preditos de DAP, foi possível identificar os melhores indivíduos. Esses resultados serão fundamentais para a seleção de indivíduos com GEBVs superiores, visando as futuras gerações de autofecundação. Esses resultados são fundamentais para as próximas etapas do programa de melhoramento baseado em linhagens endogâmicas, pois possibilitam a aceleração do ciclo de autofecundação. Os indivíduos com melhor classificação poderão ser selecionados para futuros cruzamentos e novas gerações de autofecundação, contribuindo para otimizar a eficiência e o progresso genético do programa de melhoramento.

REFERÊNCIAS

- ATWELL, S.; HUANG, Y. S.; VILHJÁLMSSON, B. J. Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana*. **Nature**, v. 465, p. 627-631, 2010.
- BERNARDO, R. Molecular markers and selection for complex traits in plants: Learning from the last 20 years. **Crop Science**, 48(5), 1649-1664, 2008.
- BERRAR, D. Cross-validation. In: **Encyclopedia of Bioinformatics and Computational Biology**. Oxford: Academic Press. v. 1, p. 542-545, 2018.
- CALUS, M. P. L.; BOUWMAN, A. C.; SCHROOTEN, P.; VEERKAMP, R. F. Genomic prediction based on data from three layer lines: a comparison between linear methods. **Genetics Selection Evolution**, 46, 57, 2014.
- CEBALLOS, F. C.; JOSHI, P. K.; CLARK, D. W.; RAMSAY, M.; WILSON, J. F. Runs of homozygosity: windows into population history and trait architecture. **Nature Reviews Genetics**, 19(4), 220–234, 2018.
- DAETWYLER, H. D.; VILLANUEVA, B.; WOOLLIAMS, J. A. Accuracy of predicting the genetic risk of disease using a genome-wide approach. **PLoS ONE**, v. 3, n. 10, p. e3395, 2008.
- DENIS, M.; BOUVET, J. M. Efficiency of genomic selection with models including dominance effect in the context of *Eucalyptus* breeding. **Tree Genetics & Genomes**, 9, 37–51, 2013.
- DESTA, Z.A.; ORTIZ, R. Genomic selection: genome-wide prediction in plant improvement. **Trends in Plant Science**, v. 19, p. 592-601, 2014.
- DUARTE, D.; JURCIC, E. J.; DUTOUR, J.; VILLALBA, P. V.; CENTURIÓN, C.; GRATTAPAGLIA, D.; CAPPA, E. P. Genomic selection in forest trees comes to life: unraveling its potential in an advanced four-generation *Eucalyptus grandis* population. **Frontiers in Plant Science**, 15, 1462285, 2024.
- ENDELMAN, J. B.; JANNINK, J. L. Shrinkage estimation of the realized relationship matrix. **G3: Genes, Genomes, Genetics**, v. 2, n. 11, p. 1405-1413, 2012.
- FARIA, D. A. **Seleção genômica ampla para volume e qualidade da madeira em eucalipto**. Tese de Doutorado, Universidade Federal de Lavras, 2014.
- FRISCH, M.; MELCHINGER, A. E. Variance of the parental genome contribution to inbred lines derived from biparental crosses. **Genetics**, 176(1), 477-488, 2007.
- GAYNOR, R. C.; GORJANC, G.; BENTLEY, A. R.; OBER, E. S.; H. P.; JACKSON, R.; MACKAY, I. J.; HICKEY, J. M. A two-part strategy for using genomic selection to develop inbred lines. **Crop Science**, v. 57, n. 5, p. 2372-2386, 2017.
- GIANOLA, D.; DE LOS CAMPOS, G.; HILL, W. G.; MANFREDI, E.; FERNANDO, R. Additive genetic variability and the Bayesian alphabet. **Genetics**, 183(1), 347–363, 2009.
- GRATTAPAGLIA, D.; RESENDE, M. D. V.; RESENDE, M.R. Genomic Selection for growth traits in *Eucalyptus*: accuracy within and across breeding populations. **BMC Proc** 5 (Suppl 7), O16, 2011.

- GRATTAPAGLIA, D.; SILVA-JUNIOR, O.; RESENDE, R.; CAPPA, E.; MÜLLER, B.; TAN, B.; ISIK, F.; RATCLIFFE, B.; EL-KASSABY, Y. Quantitative genetics and genomics converge to accelerate forest tree breeding. **Front. Plant Sci.** 9, 01693, 2018.
- GRATTAPAGLIA, D.; RESENDE, M. D. V. Genomic selection in forest tree breeding. **Tree Genetics & Genomes**, 7, 241–255m 2011
- HABIER, D.; FERNANDO, R. L.; DEKKERS, J. C. M. The impact of genetic relationship information on genome-assisted breeding values. **Genetics**, 177(4), 2389–2397, 2011.
- HAYES, B. J.; BOWMAN, P. J.; CHAMBERLAIN, A. J.; GODDARD, M. E. Genomic selection in dairy cattle: Progress and challenges. **Journal of Dairy Science**, v. 92, n. 2, p. 433-443, 2009.
- HAYWARD, M.D.; BOSEMARK, N.O.; ROMAGOSA, T. (Ed.). Plant breeding: principles and prospects. **Springer Science; Business Media**, 2012.
- HEDERSON, C. R. **Applications of linear models in animal breeding**. Guelph: University of Guelph Press. 413 p, 1984.
- HEFFNER, E. L.; SORRELLS, M. E.; JANNINK, J. L. Genomic selection for crop improvement. **Crop Science**, v. 49, n. 1, p. 1-12, 2009.
- HESLOT, N.; YANG, H.P.; SORRELLS, M.E.; JANNICK, J.L. Genomic selection in plant breeding: a comparison of models. **Crop science**, v. 52, n. 1, p. 146-160, 2012.
- HUFFORD, M. B.; XU, X.; VAN HEERWAARDEN, J.; *et al.* Comparative population genomics of maize domestication and improvement. **Nature Genetics**, v. 44, n. 7, p. 808-811, 2012.
- KUHN, M.; JOHNSON, K. **Applied Predictive Modeling**. Springer, 2013. <https://doi.org/10.1007/978-1-4614-6849-3>.
- LEBOWITZ, R. J.; SOLLER, M.; BECKMANN, J. S. Trait-based analyses for the detection of linkage between marker loci and quantitative trait loci in crosses between inbred lines. **TAG. Theoretical and applied genetics**, 73(4), 556–562, 1987.
- LEGARRA, A.; AGUILAR, I.; MISZTAL, I. A relationship matrix including full pedigree and genomic information. **Journal of Dairy Science**, 92(9), 4656–4663, 2009.
- LÓPEZ-CRUZ, M. A.; CROSSA, J.; BONNETT, D. *et al.* Increased prediction accuracy in wheat breeding trials using a marker × environment interaction genomic selection model. **G3: Genes, Genomes, Genetics**, v. 5, n. 4, p. 569-582, 2015.
- MEUWISSEN, T. H. E.; HAYES, B. J.; GODDARD, M. E. Prediction of total genetic value using genome-wide dense marker maps. **Genetics**, Bethesda, v. 157, n. 4, p. 1819-1829, 2001.
- MPHAHLELE, M.M.; ISIK, F.; MOSTERT-O'NEILL, M.M. Expected benefits of genomic selection for growth and wood quality traits in *Eucalyptus grandis*. **Tree Genetics & Genomes** 16, 49, 2020.

MONTESINOS-LÓPEZ, O. A.; MONTESINOS-LÓPEZ, A.; CROSSA, J.; GIANOLA, D. A marker weighting approach for enhancing within-family accuracy in genomic prediction. **G3: Genes, Genomes, Genetics**, 14(2), jkad278. 2023.

MONTGOMERY, D. C.; RUNGER, G. C. Applied Statistics and Probability for Engineers (5th ed.). **Wiley**, 2010.

ORNELLA, L., *et al.* Genomic prediction of genetic values for resistance to wheat rusts. **The Plant Genome**, v. 5, p. 136-148, 2012.

PALUDETO, J. G. Z. *et al.* Genomic relationship-based genetic parameters and prospects of genomic selection for growth and wood quality traits in *Eucalyptus benthamii*. **Tree Genetics & Genomes**, v. 17, p. 1–20, 2021.

PARK, T.; CASELLA, G. The Bayesian Lasso. **Journal of the American Statistical Association**, 103(482), 681–686, 2008.

PEARSON, K. Mathematical contributions to the theory of evolution. III. Regression, heredity, and panmixia. **Philosophical Transactions of the Royal Society of London**, 187, 253–318, 1896.

PEIXOTO, M. A.; LEACH, K. A.; JARQUIN, D.; FLANNERY, P.; ZYSTRO, J.; TRACY, W. F.; BHERING, L.; RESENDE, M. F. R. Utilizing genomic prediction to boost hybrid performance in a sweet corn breeding program. **Front. Plant Sci.** 15, 1293307, 2024.

RAMALHO, M. A. P.; SANTOS, H. G.; SOUZA, T. S. *Eucalyptus* breeding programs: a proposal for the use of inbred progênies. **Cerne**, v. 28, p. e103049, 2022.

RESENDE, M. D. V.; RESENDE JUNIOR, M. F. R.; SILVA, F. F. *et al.* Statistical analysis of genotype × environment interaction via reaction norms and its application to genomic selection. **PLoS ONE**, v. 7, n. 10, p. e47847, 2012.

RESENDE, M. D. V.; RESENDE JR, M. F. R.; SANSALONI, C. P.; PETROLI, C. D.; MISSIAGGIA, A. A.; AGUIAR, A. M.; GRATTAPAGLIA, D. Genomic selection for growth and wood quality in *Eucalyptus*: capturing the missing heritability and accelerating breeding for complex traits in forest trees. **New Phytologist**, 2012.

ROBERT, P., *et al.* Phenomic selection in wheat breeding: identification and optimisation of factors influencing prediction accuracy and comparison to genomic selection. **Theoretical and Applied Genetics**, v. 135, p. 895-914, 2022.

ROCHA, L. F. *et al.* Quantitative trait loci related to growth and wood quality traits in *Eucalyptus grandis* W. Hill identified through single- and multi-trait genome-wide association studies. **Tree Genetics & Genomes**, v. 18, n. 6, p. 38, 12 out. 2022.

SCHAFER, J. L.; GRAHAM, J. W. Missing data: Our view of the state of the art. **Psychological Methods**, 7(2), 147–177, 2002.

SCUTARI, M.; MACKAY, I.; BALDING, D. J. Improving the efficiency of genomic selection. **Statistical Applications in Genetics and Molecular Biology**, 12(4), 517–527, 2013.

SHI, S.; LI, X.; FANG, L.; LIU, A.; SU, G.; ZHANG, Y.; LUOBU, B.; DING, X.; ZHANG, S. Genomic Prediction Using Bayesian Regression Models With Global–Local Prior. **Frontiers in Genetics**, 12, 628205. 2021.

SHALIZI, M. N.; ISIK, F. Genetic parameter estimates and GxE interaction in a large cloned population of *Pinus taeda* L. **Tree Genetics & Genomes**, v. 15, n. 3, p. 46, 29 maio 2019.

SILVA-JUNIOR, O. B.; FARIA, D. A.; GRATTAPAGLIA, D. A flexible multi-species genome-wide 60K SNP chip developed from pooled resequencing 240 *Eucalyptus* tree genomes across 12 species. **New Phytologist**, 206(4), 1527–1540, 2015.

SIMIQUÉLI, G. F. *et al.* Realized genomic selection across generations in a reciprocal recurrent selection breeding program of *Eucalyptus* hybrids. **Frontiers in Plant Science**, v. 14, 27 out. 2023.

SPRAGUE, G. F.; TATUM, L. A. General vs. specific combining ability in single crosses of corn. **Journal of the American Society of Agronomy**, v. 34, n. 10, p. 923-932, 1942.

TAKAHASHI, Y.; UEKI, M.; TAMIYA, G.; OGISHIMA, S.; KINOSHITA, K.; HOZAWA, A.; MINEGISHI, N.; NAGAMI, F.; FUKUMOTO, K.; OTSUKA, K.; TANNO, K.; SAKATA, K.; SHIMIZU, A.; SASAKI, M.; SOBUE, K.; KURE, S.; YAMAMOTO, M.; TOMITA, H. Machine learning for effectively avoiding overfitting is a crucial strategy for the genetic prediction of polygenic psychiatric phenotypes. **Translational Psychiatry**, 10(1), 294. 2020.

TAMBARUSSI, E. V. *et al.* Are tree breeders properly predicting genetic gain? A case study involving *Corymbia* species. **Euphytica**, v. 214, n. 8, p. 150, 3 ago. 2018.

TAMBARUSSI *et al.* Escolha do genótipo. Em: **EUCALIPTO: do plantio à colheita**. [s.l.: s.n.].

VALENTE, B. D.; ROSA, G. J. M.; DE LOS CAMPOS, G.; GIANOLA, D.; SILVA, M. A. Assessment of bagging GBLUP for whole-genome prediction of broiler chicken traits. **Journal of Animal Breeding and Genetics**, 131(2), 83–94, 2014.

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

VITEZICA, Z. G., VARONA, L., LEGARRA, A. On the additive and dominant variance and covariance of individuals within the genomic selection scope. **Genetics**, 195(4), 1223–1230, 2013.

WIENTJES, Y. C. J.; BIJMA, P.; CALUS, M. P. L. An equation to predict the accuracy of genomic values by combining data from multiple populations. **Genetics Selection Evolution**, 48, 22, 2016.

WINDHAUSEN, V. S.; ATLIN, G. N.; HICKEY, J. M.; CROSSA, J.; JANNINK, J. L.; SORRELLS, M. E.; RAMAN, B.; CAIRNS, J. E.; TAREKEGNE, A.; SEMAGN, K.; BEYENE, Y.; GRUDLOYMA, P.; TECHNOW, F.; RIEDELSHEIMER, C.; MELCHINGER, A. E. Effectiveness of Genomic Prediction of Maize Hybrid

Performance in Different Breeding Populations and Environments. **G3 Genes, Genomes, Genetics** 2, 1427–1436, 2012.

YANG, L.; ZHANG, X.; CHEN, J. Winsorization greatly reduces false positives by popular differential expression methods when analyzing human population samples. **Genome Biology**, 25, 282, 2024

ZARGAR, S. M.; RAKWAL, R.; SHARMA, S.; *et al.* Recent advances in molecular marker techniques: Insight into QTL mapping, GWAS and genomic selection in plants. **Journal of Crop Science and Biotechnology**, v. 18, n. 5, p. 293-308, 2015.

ZHONG, S.; DEKKERS, J.C.M.; FERNANDO, R.L.; JANNINK, J.L. Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: a barley case study. **Genetics**, v. 182, p. 355-364, 2009.

CONSIDERAÇÕES FINAIS

A identificação precisa de indivíduos endogâmicos e a seleção genômica de genótipos superiores, pilares dos resultados apresentados nestes capítulos, representam uma transformação fundamental para as etapas subsequentes do programa de melhoramento genético da Suzano S.A. Esses achados não apenas viabilizam a implementação eficiente do método de linhagens endogâmicas em *Eucalyptus*, oferecendo o conhecimento essencial, mas também sinalizam uma estratégia promissora para atender à crescente demanda por ganhos genéticos expressivos, impulsionando o futuro do melhoramento do gênero.

REFERÊNCIAS

- ALFENAS, A. C.; ZAUZA, E. A. V.; MAFIA, R. G.; ASSIS, T. F. **Clonagem e doenças do eucalipto**. Viçosa, MG: Editora UFV, 2009.
- ALLARD, R. W. **Princípios do melhoramento genético de plantas**. São Paulo: Edgard Blucher. 381 p, 1971.
- ANDRADE, E. N. Evolução do melhoramento do gênero *Eucalyptus*. **Revista Brasileira de Silvicultura**, São Paulo, v. 2, n. 3, p. 34-40, 1961.
- APPLEBY, N.; EDWARDS, D.; BATLEY, J. New Technologies for Ultra-High Throughput Genotyping in Plants. In: **Plant Genomics, Methods in Molecular Biology**, vol 513. Humana Press, 2009.
- ASSIS, T. F. Melhoramento genético de eucalipto para produção de carvão vegetal. In: **SIMPOSIO LATINO-AMERICANO DE RECURSOS GENÉTICOS**, 1., Brasília. Anais [...]. Brasília: EMBRAPA, 1986. p. 141-150, 1986.
- ASSIS, T. F. Métodos de melhoramento do *Eucalyptus*. **Revista Brasileira de Ciências Agrárias**, Recife, v. 6, n. 3, p. 514-528, 2011.
- ASSIS, T. F. Uso de clones e híbridos no melhoramento de eucaliptos. In: **SIMPOSIO SOBRE MELHORAMENTO GENÉTICO FLORESTAL**, 2014, Viçosa. Anais [...]. Viçosa: Universidade Federal de Viçosa. p. 75-98, 2014.
- ASSIS, T. F.; ABAD, J. I.; AGUIAR, A. M. Técnicas de melhoramento genético do eucalipto para produção de madeira. **Revista Brasileira de Engenharia Florestal**, Brasília, v. 35, n. 3, p. 456-470, 2015.
- BARRAGAN, A. C.; COLLENBERG, M.; SCWAB, R.; KERSTEN, S.; KERSTENS, M. H. L.; POŽÁROVÁ, D.; BEZRUKOV, I.; BEMM, F.; KOLÁR, F.; WEIGEL, D. Deleterious phenotypes in wild *Arabidopsis arenosa* populations are common and linked to runs of homozygosity. **G3: Genes, Genomes, Genetics**, v. 14, n. 3, p. jkad290, 2024.
- BARRETT, S. C.; HARDER, L. D. Ecology and evolution of plant mating. **Trends Ecol Evol**. Feb;11(2):73-9, 1996.
- BERNARDO, R. **Breeding for Quantitative Traits in Plants (2nd ed.)**. Stemma Press, 2010.
- BOERJAN, W. Biotechnology and the domestication of forest trees. **Current Opinion in Biotechnology**, 16(2), 159-166, 2005.
- BORÉM, A.; MIRANDA, G. V. **Melhoramento de plantas**. 7. ed. Viçosa: Editora UFV. 572 p, 2009.
- BORÉM, A.; MIRANDA, G. V. **Melhoramento de plantas**. 6. ed. Viçosa: Editora UFV, 2013.
- BORRALHO, N. M. G.; POTTS, B. M.; SNIEGOWSKI, P. Genetic improvement of *Eucalyptus* for pulpwood. **Appita Journal**, 45(2), 105-113, 1992.

- CABALLERO, A.; VILLANUEVA, B.; DRUET, T. On the estimation of inbreeding depression using different measures of inbreeding from molecular markers. **Evolutionary Applications**, v. 14, n. 2, p. 416-428, 2021.
- CASTRO, G. S. A. *et al.* Avanços no florescimento precoce em eucalipto. **Revista Brasileira de Ciência Florestal**, Brasília, v. 55, n. 2, p. 225-240, 2021.
- CHARLESWORTH, D.; WILLIS, J. H. The genetics of inbreeding depression. **Nature Reviews Genetics**, v. 10, p. 783-796, 2009.
- COLLARD, B. C. Y.; MACKILL, D. J. Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. **Philosophical Transactions of the Royal Society B: Biological Sciences**, v. 363, n. 1491, p. 557-572, 2008.
- CRNOKRAK, P.; BARRETT, S. C. H. Perspective: Purging the genetic load: a review of the experimental evidence. **Evolution**, v. 56, n. 12, p. 2347-2358, 2002.
- CRUTE, I. R.; PINK, D. A. C. Genetics and utilisation of pathogen resistance in plants. **Plant Cell**, 8(11), 1747–1755, 1996.
- DE ASSIS, T. F.; ABAD, J. I. M.; AGUIAR, A. M. Melhoramento genético do eucalipto. In: **SILVICULTURA DO EUCALIPTO NO BRASIL**. p. 215, 2016.
- DEKKERS, J. C. M.; HOSPITAL, F. The use of molecular genetics in the improvement of agricultural populations. **Nature Reviews Genetics**, London, v. 3, n. 1, p. 22-32, 2002.
- DE LOS CAMPOS, G. *et al.* Predicting quantitative traits with regression models for dense molecular markers. **Genetics**, Bethesda, v. 182, n. 1, p. 375-385, 2009.
- DENIS, M.; BOUVET, J.M. Efficiency of genomic selection with models including dominance effect in the context of *Eucalyptus* breeding. **Tree Genetics & Genomes**, 9, 37–51, 2013.
- DO CANTO, J.; STUDER, B.; LUBBERSTEDT, T. Overcoming self-incompatibility in grasses: a pathway to hybrid breeding. **Theoretical and Applied Genetics**, 129(10), 1815–1829 (2016).
- DUVICK, D. N. Biotechnology in the 1930s: the development of hybrid maize. **Nature Reviews Genetics**, 2(1), 69-74, 2001.
- EAST, E. M. Inbreeding in corn. **Connecticut Agricultural Experiment Station Report**, New Haven, v. 42, p. 419-428, 1908.
- ELDRIDGE, K. G.; DAVIDSON, J.; HARWOOD, C. E.; VAN WYK, G. Eucalypt domestication and breeding. **Oxford: Clarendon Press**, 1994.
- FALCONER, D. S.; MACKAY, T. F. C. **Introduction to Quantitative Genetics (4th ed.)**. Longman Group Ltd, 1996.
- FAO. Global forest sector outlook 2050: Assessing future demand and sources of timber for a sustainable economy – Background paper for The State of the World's Forests 2022. **FAO Forestry Working Paper, No. 31**. Rome: Food and Agriculture Organization of the United Nations, 2022.

FERENČAKOVIĆ, M. *et al.* Autozygosity and the estimation of inbreeding depression in selected populations. **Animal Genetics**, Oxford, v. 42, n. 2, p. 273-283, 2011.

FERREIRA, M. Histórico do plantio de eucalipto no Brasil. **Revista Árvore**, Viçosa, v. 13, n. 4, p. 325-332, 1989.

FERREIRA, D. S.; SOUZA, A. P.; LIMA, M. M. Aplicação da seleção recorrente recíproca no melhoramento genético de *Eucalyptus* spp. **Revista Brasileira de Genética Florestal**, 22(4), 145-160, 2018.

GABRIEL, S. B.; SCHAFFNER, S. F.; REICH, D.; LAUTENBACH, A.; ALTSCHULER, D.; BOTSTEIN, D.; DALY, M. J. The structure of haplotype blocks in the human genome. **Science**, v. 296, n. 5576, p. 2225-2229, 2002.

GANAL, M. W.; ALTMANN, T.; RÖDER, M. S. SNP identification in crop plants. **Plant Biotechnology Journal**, Cambridge, v. 7, n. 3, p. 224-234, 2009.

GIANOLA, D. *et al.* Efficiency of genomic selection in crop breeding. **Nature Genetics**, New York, v. 52, n. 1, p. 15-22, 2021.

GIANOLA, D.; VAN KAAM, J. B. C. H. M. Reproducing kernel Hilbert spaces regression methods for genomic-assisted prediction of quantitative traits. **Genetics**, 178(4), 2289-2303, 2008.

GIBSON, J. *et al.* Runs of homozygosity: windows into population history and trait architecture. **Nature Reviews Genetics**, London, v. 7, n. 6, p. 384-394, 2006.

GONÇALVES, J. L. M.; ALVARES, C. A.; HIGA, A. R.; SILVA, L. D.; ALFENAS, A. C.; STAHL, J.; OLIVEIRA, J. C. Integrating genetic and silvicultural strategies to minimize abiotic and biotic constraints in Brazilian eucalypt plantations. **Forest Ecology and Management**, 301, 6-27, 2013.

GRATTAPAGLIA, D. Genomic selection in forest tree breeding. **Tree Genetics & Genomes**, Heidelberg, v. 13, n. 2, p. 1-12, 2017.

GRATTAPAGLIA, D. Aplicações de marcadores moleculares no melhoramento de eucalipto. **Revista Brasileira de Genética**, Ribeirão Preto, v. 45, n. 3, p. 267-276, 2022.

GRATTAPAGLIA, D.; KIRST, M. *Eucalyptus* applied genomics: from gene sequences to breeding tools. **New Phytologist**, 179(4), 911-929, 2008.

GRATTAPAGLIA, D.; RESENDE, M. D. V. Genomic selection in forest tree breeding. **Tree Genetics & Genomes**, Heidelberg, v. 7, n. 4, p. 763-775, 2011.

GRATTAPAGLIA, D.; RESENDE, M. D. V.; SANSALONI, C.; PETROLI, C. D.; FARIA, D.; MISSIAGGIA, A. A.; TAKAHASHI, E. K.; ZAMPROGNO, K. C. Genomic Selection for growth and wood traits in tropical *Eucalyptus*: a paradigm shift in forest tree breeding. **EMBRAPA Genetic Resources and Biotechnology**, 2018.

GRIFFIN, A. R.; COTTERILL, P. P. Genetic variation in growth of outcrossed, selfed and open-pollinated progenies of *Eucalyptus Regnans* and some implications for breeding strategy. **Silvae Genetica**, 37(4), 124-131, 1988.

- HALLAUER, A. R.; CARENA, M. J.; MIRANDA FILHO, J. B. Quantitative Genetics in Maize Breeding. **Springer**, 2010.
- HAYES, B. J. *et al.* Genomic selection for complex traits. **Nature Reviews Genetics**, London, v. 10, n. 6, p. 381-391, 2009.
- HAYKIN, S. **Kalman Filtering and Neural Networks**. Wiley-Interscience, 2001.
- HAYWARD, M.D.; BOSEMARK, N.O.; ROMAGOSA, T. (Ed.). Plant breeding: principles and prospects. **Springer Science; Business Media**, 2012.
- HEDRICK, P. W.; HELLSTEN, U.; GRATTAPAGLIA, D. The impact of genomics on conservation genetics. **Evolutionary Applications**, Hoboken, v. 9, n. 5, p. 667-683, 2016.
- HEDRICK, P.W. **Genetics of populations**. Jones; Bartlett Publishers, 2009.
- HEFFNER, E. L. *et al.* Genomic selection for crop improvement. **Crop Science**, Madison, v. 49, n. 1, p. 1-12, 2009.
- IBÁ - INDÚSTRIA BRASILEIRA DE ÁRVORES. **Relatório anual de atividades 2024**. São Paulo: IBÁ, 2024.
- JANNINK, J. L. *et al.* Bayesian models for genomic prediction. **Genetics**, Bethesda, v. 183, n. 4, p. 1167-1181, 2009.
- JONES, D. F. The effects of inbreeding and crossbreeding upon development. **Bulletin 104 of the Connecticut Agricultural Experiment Station**, New Haven, p. 5-40, 1918.
- KARIMI, Z. The role of ROH islands in plant breeding. **Journal of Plant Breeding and Genetics**, Tehran, v. 47, n. 4, p. 112-120, 2013.
- KELLER, L. F.; WALLER, D. M. Inbreeding effects in wild populations. **Trends in Ecology & Evolution**, 17(5), 230–241, 2002.
- KRUG, C. A.; ALVES, J. P. Melhoramento genético de *Eucalyptus*: Uma abordagem histórica. **Revista Brasileira de Genética**, Ribeirão Preto, v. 18, n. 2, p. 85-92, 1949.
- LENCZ, T. *et al.* Runs of homozygosity reveal highly penetrant recessive loci in *schizophrenia*. **Proceedings of the National Academy of Sciences**, Washington, v. 104, n. 50, p. 19942-19947, 2007.
- MEUWISSEN, T. H. E.; HAYES, B. J.; GODDARD, M. E. Prediction of total genetic value using genome-wide dense marker maps. **Genetics**, Bethesda, v. 157, n. 4, p. 1819-1829, 2001.
- NICOLLE, D.; RITTER, M. K.; JONES, R. C.; PHILLIPS, G. P.; FRENCH, M. E.; CUMMING, R.; BELL, S. A. J. The genus problem – *Eucalyptus* as a model system for minimising taxonomic disruption. **Taxon**, 72(1), 132-145, 2023.
- NORDBORG, M. Linkage disequilibrium, gene trees and selfing: an ancestral recombination graph with partial self-fertilization. **Genetics**, v. 154, n. 2, p. 923-929, 2000.

- PÉREZ-RODRÍGUEZ, P.; CROSSA, J.; BONNETT, D. Modelling genetic effects in genomic selection. **Theoretical and Applied Genetics**, Berlin, v. 133, n. 4, p. 1087-1099, 2020.
- PINTO JÚNIOR, E. L.; SILVEIRA, R. Adaptabilidade das espécies de eucalipto ao clima brasileiro. **Revista Brasileira de Climatologia**, Londrina, v. 19, n. 4, p. 789-798, 2021.
- PURFIELD, D. C.; BERRY, D. P.; MCPARLAND, S.; BRADLEY, D. G. Runs of homozygosity and population history in cattle. **BMC genetics**, v. 13, p. 1-11, 2012.
- RAMALHO, M. A. P.; SANTOS, H. G, SOUZA T. S. *Eucalyptus* breeding programs: a proposal for the use of inbred progenies. **Cerne**, v. 28, pp. 1-9, 2022.
- REIS, C. A. F.; GONÇALVES, F. M. A.; ROSSE, L. N.; COSTA, R. R. G. F.; RAMALHO, M. A. P. Correspondence between performance of *Eucalyptus* spp trees selected from family and clonal tests. **Genet. Mol. Res.** 10, 1172–1179, 2011.
- RESENDE, M. D. V. Genética Quantitativa e Seleção em Plantas Perenes: Aplicações ao Melhoramento do Eucalipto. **Embrapa Florestas**, 2002.
- RESENDE, M. D. V. **Melhoramento genético de árvores florestais**. 2. ed. Colombo: Embrapa Florestas. 463 p, 2012.
- SHULL, G. H. Hybridization methods in plant breeding. **Proceedings of the American Society for Horticultural Science**, Alexandria, v. 6, p. 79-94, 1909.
- SHULL, G. H. Pure-line inheritance. **Zeitschrift für induktive Abstammungs- und Vererbungslehre**, Jena, v. 5, p. 51-65, 1911.
- SHULL, G. H. The composition of a field of maize. **American Breeders Association Report**, Chicago, v. 4, p. 296-301, 1908.
- VIEIRA, M. L. C.; SANTINI, L.; DINIZ, A. L.; MUNHOZ, C. F. Microsatellite markers: what they mean and why they are so useful. **Genetics and Molecular Biology**, Ribeirão Preto, v. 39, n. 3, p. 312-328, jul./set. 2016
- VIGNAL, A.; MILAN, D.; SANCRISTOBAL, M.; EGGEN, A. A review on SNP and other types of molecular markers and their use in animal genetics. **Genetics Selection Evolution**, 34(3), 275-305, 2002.
- YANG, J. *et al.* Genome partitioning of genetic variation for complex traits using common SNPs. **Nature Genetics**, New York, v. 43, n. 6, p. 519-526, 2013.